Etude contrastive des chaînes de référence en français et en chinois : étude de cas à partir d'un corpus parallèle composé de Jean-Christophe de Romain Rolland et de sa traduction par Lei Fu : volume 1

Chang Guo

▶ **To cite this version:**

Chang Guo. Etude contrastive des chaînes de référence en français et en chinois : étude de cas à partir d'un corpus parallèle composé de Jean-Christophe de Romain Rolland et de sa traduction par Lei Fu : volume 1. Linguistique. Université de Strasbourg, 2023. Français. NNT : 2023STRAC015 . tel-04453175

HAL Id: tel-04453175
https://theses.hal.science/tel-04453175

Submitted on 12 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Strasbourg

École doctorale des Humanités (ED520) Université de Strasbourg

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES HUMANITES 520

Unité de Recherche 1339 LiLPa (Linguistique, Langues et Parole)

THÈSE présentée par :

Chang GUO

soutenue le : 6 décembre 2023

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline : Sciences du Langage

ÉTUDE CONTRASTIVE DES CHAÎNES DE RÉFÉRENCE EN FRANÇAIS ET EN CHINOIS

ÉTUDE DE CAS À PARTIR D'UN CORPUS PARALLÈLE COMPOSÉ DE *JEAN-CHRISTOPHE* DE ROMAIN ROLLAND ET DE SA TRADUCTION PAR LEI FU

Volume I

THÈSE dirigée par :

Mme SCHNEDECKER Catherine	Professeure, Université de Strasbourg
Mme WANG Xiuli	Professeure, Université des Langues et Cultures de Pékin

RAPPORTEURS :

M. LANDRAGIN Frédéric	Directeur de recherche, CNRS
Mme RICHARD Elisabeth	Professeure, Université Rennes 2

EXAMINATRICE :

Mme GARDELLE Laure	Professeure, Université Grenoble Alpes

Remerciements

Écrire une thèse est un travail à la fois individuel et collectif : individuel car nous avons passé des années seule à lire, questionner, observer, réfléchir et rédiger ; collectif puisque nous n'aurions pu terminer ce travail sans le soutien de nombreuses personnes ; d'où ces remerciements.

Je tiens tout d'abord à remercier vivement madame Catherine Schnedecker, qui, en plus d'avoir été ma codirectrice de thèse pendant huit ans, a également dirigé mon mémoire de master pendant deux années. Sur mon parcours consacré à l'étude des sciences du langage en France, elle m'a donc accompagnée pendant dix ans au total, et je n'aurais jamais pu terminer, ni mon master, ni mon doctorat, sans sa direction. Madame Schnedecker m'a, d'une part, fourni une bibliographie abondante qui inclut tous les aspects et les domaines relatifs à mon sujet de recherche ; elle m'a également, d'autre part, proposé une méthodologie scientifique stricte qui m'a encadrée et m'a permis d'avancer sur le bon chemin. C'est à elle que je dois d'avoir appris à adopter l'attitude rigoureuse requise dans la recherche scientifique.

Je tiens ensuite à remercier ma co-directrice madame Xiuli Wang, qui fut également la directrice de mon mémoire de licence. Si je m'exprime avec une certaine aisance en français aujourd'hui, c'est grâce à elle et aux quatre années d'études de langue et de culture françaises passées sous sa direction. Je n'aurais pas pu terminer ma licence et mon doctorat sans son soutien. Pendant mon doctorat, madame Wang m'a proposé de nombreuses occasions de coopérer avec des collègues chinois et de découvrir l'ambiance scientifique de différentes équipes universitaires en Chine. Ensemble, nous avons réalisé deux séjours académiques à Pékin grâce au projet CaiYuanpei du Ministère de l'Éducation de Chine.

Je voudrais également remercier les membres du jury de ma soutenance, monsieur Frédéric Landragin, madame Elisabeth Richard et madame Laure Gardelle, d'avoir accepté d'examiner mon travail. J'ai connu monsieur Landragin, rapporteur de ma thèse, à la fin de mon master par le biais du projet DEMOCRAT. En tant que directeur de ce projet, monsieur Landragin m'a introduite auprès de toute une équipe formée par les collègues du laboratoire LATTICE de Paris et du laboratoire ICAR de Lyon. Grâce à ce projet, j'ai aussi eu l'occasion de me former au maniement des outils linguistiques avec lesquels j'ai traité mon corpus. J'ai connu madame Gardelle, examinatrice de ma thèse, au tout début de mon doctorat. En tant que membre du jury de mon Comité de Suivi Individuel de thèse, elle m'a proposé un suivi annuel constructif et motivant qui m'a permis de vérifier objectivement l'avancement de ma rédaction par l'examen de ses points forts et de ses points faibles.

Je dois remercier madame Marine Le Mené, autre membre de mon Comité de Suivi de thèse. Pendant l'année où elle a effectué son post-doctorat à LILPA et comme collègue et amie, elle m'a accompagnée de façon à me permettre de progresser dans mon écriture.

Tout au long de ce travail doctoral, j'ai beaucoup apprécié les activités scientifiques organisées par l'union de recherche LILPA : les colloques, les séminaires, les réunions, les soutenances de thèses et les rencontres conviviales. Cette ambiance agréable est le résultat des efforts collectifs de l'équipe de direction, des professeurs, et également des doctorants. Je suis très heureuse d'avoir eu la chance de m'intégrer dans ce groupe magnifique.

Je voudrais aussi accorder mes vifs remerciements à madame Jun Miao, qui m'a proposé les textes bruts de mon corpus de thèse. En effet, le roman-fleuve ***Jean-Christophe*** en français et en chinois n'a pas de version complète libre d'accès en format brut. Elle a donc tapé manuellement ***Jean-Christophe*** en français et en chinois pour les besoins de sa propre thèse et m'a ensuite proposé sans réserve cette version brute. Sans elle, je n'aurais donc pas eu la possibilité de travailler sur ces textes.

J'aimerais par ailleurs remercier monsieur Thomas Boutonnet, madame Marie Bizais, madame Daniéla Campo, du Département d'études chinoises de l'Université de Strasbourg, qui ont décidé de me recruter en tant que maître de langue, cela étant mon premier poste universitaire. Merci, à ce propos, à mes chères collègues et amies Dingxi Liu et Jianhui Wang, qui m'ont beaucoup aidée à faire mes premiers pas dans l'enseignement du chinois en France.

Je remercie en outre les instituts qui ont soutenu mes cursus universitaires en France : la Faculté des Lettres, l'École Doctorale des Humanités 520 et l'Université de Strasbourg, dans lesquels j'ai eu l'opportunité de suivre tous types de formations disciplinaires et inter-disciplinaires. C'est grâce à ces formations que j'ai maîtrisé des logiciels bureautiques et que j'ai appris l'anglais et l'allemand.

Je ne dois pas oublier de remercier le China Scholarship Council qui m'a proposé un financement de quatre ans pour ce long travail doctoral.

Enfin, les remerciements de mon cœur vont à mes parents que j'aime et qui m'aiment depuis toujours et pour toujours.

Abréviations (gloses, tableaux, figures et annexes)

- 1P(s), 2P(s), 3P(s) : pronom de la première / deuxième / troisième personne (au pluriel).
- AUX : catégorie grammaticale « auxiliaire » du chinois.
- BA : 把, transcription phonétique en pinyin /ba/, mot grammatical et auxiliaire propre du chinois ; marqueur de la structure de double complément d'objet.
- CDN : complément du nom.
- CDO : complément d'objet.
- Ch / fr : chinois / français.
- CMP : comparatif, comparaison.
- CR : chaîne(s) de référence.
- DE : 的, transcription phonétique en pinyin /de/, marqueur grammatical et auxiliaire polyvant propre du chinois.
- Dét : déterminant.
- ER : expression(s) référentielle(s).
- INTERR : catégorie grammaticale « interrogatif ».
- INTERJ : catégorie grammaticale « interjectif ».
- JC : personnage du roman Jean-Christophe.
- JM : personnage du roman Jean-Michel.
- L : personnage du roman Louisa.
- LE : 了, transcription phonétique en pinyin /le/, mot grammatical et auxiliaire propre du chinois, marqueur du passé composé.
- M : personnage du roman Melchior.
- Nbre : nombre.
- Nco / npr : nom(s) commun(s) / nom(s) propre(s).
- NEG : négatif, négation.
- OCC : occurrence(s).
- ONOMA : catégorie grammaticale « onomatopée ».
- PREP : catégorie grammaticale « préposition ».
- Réf : référent(s).

Sommaire

Remerciements 1

Abréviations 3

Sommaire 5

Introduction 7

Partie I : Problèmes théoriques relatifs aux chaînes de référence dans les approches contrastives franco-chinoises 15

Chapitre 1 : Revue des définitions de la notion de chaîne de référence et notre position relativement aux aspects franco-chinois 17

Chapitre 2 : Trois notions-clés fondamentales pour examiner les problèmes des chaînes de référence relatifs à l'approche franco-chinoise 43

Partie II : Problèmes méthodologiques relatifs aux approches contrastives et à la constitution d'un corpus 65

Chapitre 3 : Motivations et représentativité en termes du choix d'un corpus 67

Chapitre 4 : Outils pour la constitution d'un corpus bilingue annoté en expressions référentielles et en chaînes de référence 103

Chapitre 5 : Processus du traitement outillé des expressions référentielles et des chaînes de référence en français et en chinois 125

Partie III : Observations sur le corpus Jean-Christophe : régularités et spécificités contrastives des chaînes de référence en français et en chinois 179

Chapitre 6 : Régularités des chaînes de référence dans un roman-fleuve français : aspects quantitatifs et qualitatifs 181

Chapitre 7 : Propriétés des chaînes de référence dans un roman-fleuve traduit en chinois : approches contrastives franco-chinoises 261

Chapitre 8 : Synthèse comparative 321

Conclusion 339

Bibliographie 343

Table des matières 365

Liste des tableaux 373

Liste des figures 377

Liste des annexes 379

Introduction

Notre thèse porte sur la notion de *chaîne de référence* en français et en chinois. Cette notion linguistique se base sur celle, à la fois linguistique et philosophique, de *référence*. Les réflexions qui ont été consacrées à cette dernière commencent avec Platon (428 av. J.-C. – 348 av. J.-C.), et, en passant par des philosophes de l'époque moderne, comme Leibniz (1646 – 1716), se sont poursuivies jusqu'à l'époque contemporaine, avec notamment des auteurs comme Frege (1848 – 1925), Russell (1872 – 1970) et des linguistes philosophes actuels, comme Donnellan (1913 – 2015) et Strawson (1919 – 2006). Définie comme relation entre une entité du monde réel et l'expression d'une langue, la référence est une fonction universelle du langage d'humain (*cf*. Jakobson, 1963 : 78). Avec, pour socle commun, la notion de *référence* sous son double aspect philosophique et linguistique, les chaînes de référence se situent à la croisée de plusieurs domaines comme la sémantique, la linguistique textuelle, la linguistique contrastive, la linguistique outillée et la linguistique cognitive.

La notion de *chaîne de référence* a été définie par plusieurs auteurs en diverses langues, comme Chastain (1975) en anglais, et étudiée notamment par Corblin (1985, 1999), Charolles (1988, 1997, 2002, 2007) et Schnedecker (1992, 1997, 2012, 2014, 2017, 2019, 2021) en français. L'importance des études contrastives sur les variations des chaînes de référence a été soulignée par de nombreux auteurs. Par exemple, Schnedecker (1997) propose trois critères contrastifs dont la fonction est de préciser les spécificités des chaînes selon leur contexte : le genre discursif, la diachronie d'une langue et le système linguistique. Les deux premiers concernent les approches contrastives intra-linguistiques, alors que le troisième touche aux approches inter-linguistiques. Quelques études relatives à cette dernière facette portent sur des aspects étroitement liés à notre recherche. Par exemple, la thèse de Baumer (2012) aborde certaines questions contrastives des chaînes de référence en français et en anglais. Une série d'études de Miao (2008, 2011, 2015) présentent les observations de stratégies de traductions franco-chinoises concernant le roman-fleuve *Jean-Christophe* de Romain Rolland et plusieurs versions de traductions vers le chinois ; celle de Hou (2021) compare la saillance référentielle en français et en chinois, et celle de Kim (2023) sur une comparaison des chaînes de référence en français et en coréen. Notre propre travail se concentre donc sur les questions suivantes : quelles variations apparaissent dans la composition des chaînes de référence en français et en chinois ? À quels niveaux peuvent-elles être inventoriées et illustrées ?

Cette approche comparative s'inspire des deux principes de Corblin (1995 : 29) sur la variation des expressions coréférentielles ; la comparaison entre le français et le chinois nous intéresse particulièrement :

> *(i) Il n'existe pas de système linguistique unifié pour l'identité de référence ;*
>
> *(ii) L'interprétation par identité référentielle ne semble définitoire pour aucune forme linguistique.*
>
> *(Corblin, 1995 : 29)*

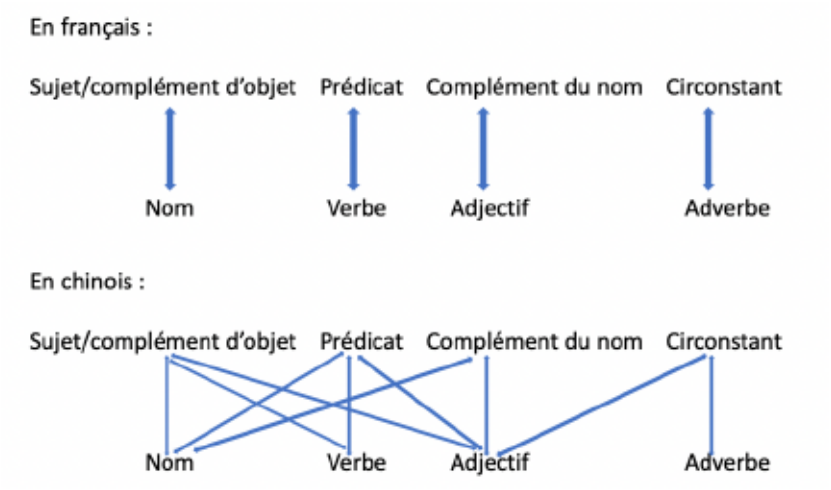
Les expressions référentielles, composants des chaînes de référence, sont propres à chaque langue. Dans les typologies des familles de langues, le français et le chinois n'appartiennent pas à la même famille linguistique : la première est une langue indo-européenne de la famille des langues romanes, tandis que la seconde constitue à elle seule

une famille linguistique sans parenté avec d'autres langues (*cf*. Dubois *et al*., 2012). Le français est une langue alphabétique dont les mots constituent des unités élémentaires en lettres latines : les mots se composent de lettres et les espaces marquent le début et la fin des mots (*cf*. Dubois *et al*., 2012). Une expression référentielle française peut comporter un seul mot (« je », « maman », « Paul ») ou plusieurs (« quelques amis », « un gros chien », « des fleurs dans le jardin »). À la différence du français, le chinois est une langue idéographique : les unités élémentaires sont les caractères en idéogrammes sémantico-phonétiques, unités toutes faites, inséparables et invariables qui représentent, d'une part, le concept ou l'idée, et, d'autre part, la prononciation (*cf*. Xu, 许余龙, 2010 ; Rabut *et al*., 2012 ; Lu, 陆俭明, 2013). Un caractère chinois peut être un *morphème libre* qui constitue un mot à lui tout seul ; il peut aussi être un *morphème non libre* qui fait seulement partie d'un mot (*cf*. Rabut *et al*., 2012). Un mot chinois, sans espace entre les caractères (*cf*. Xia *et al*., 2000), peut être composé d'un seul caractère (*che* 车 'voiture'), de deux caractères (*ka-che* 卡车 'camion'), de trois caractères (*zi-xing-che* 自行车 'vélo'), de quatre caractères (*gong-gong-qi-che* 公共汽车 'bus') et même de plus de quatre. Tout comme en français, une expression référentielle chinoise peut être composée d'un seul mot (*wo* 我 'je', *mama* 妈妈 'maman', *Bauluo* 保罗 'Paul') ou de plusieurs (*wode-hei-zixingche* 我的黑自行车 'mon vélo noir'), avec un nombre de caractères très variable.

L'œuvre cruciale relative à la présentation des expressions référentielles en français est celle de Charolles (2002), qui indique que les catégories principales sont les noms propres (« Aristote »), les SN définis (« le chat »), démonstratifs (« ce chat »), indéfinis (« un chat ») et les pronoms personnels (« il »). Les expressions référentielles en chinois sont étudiées dans leurs divers aspects par de nombreux auteurs, comme Chen (陈平, 1987), Liu (刘丹青, 2002), Xu (徐赳赳, 2003), Dong (董秀芳, 2010), Xiong (熊岭, 2012), Li (李先银, 2016) et Cao (曹道根, 2018), entre autres, selon lesquels les catégories principales ici sont les pronoms personnels (人称代词, *ni* 你 'tu'), les noms propres (专有名词, *Fa-Xiaoyan* 范小燕 'Fan Xiaoyan'), les SN démonstratifs (这 / 那 (量) 名词, *zhe-(ge)-ren* 这 (个) 人 'ce (classificateur) personne'), les SN sans déterminant (光杆普通名词, *ren* 人 'personne'), les SN quantifiés (数词 (量) 名词, *san-(ge)-ren* 三 (个) 人 'trois (classificateur) personne'), les SN indéfinis singuliers (一 (量) 名词, *yi-(ge)-ren* 一 (个) 人 'un (classificateur) personne') et les SN déterminés de classification (量词 名词, *(lai)-ge-ren* (来) 个人 '(venir) classificateur personne') (*cf*. Chen, 陈平, 1987 : 81). La comparaison des deux listes des catégories principales d'expressions référentielles dans les deux langues permet de supposer que les composants des chaînes de référence partagent des points communs et se diversifient selon les langues.

Au niveau syntaxique, le français et le chinois possèdent six fonctions syntaxiques élémentaires : le sujet, le verbe, le complément d'objet, le complément du nom, l'attribut et le circonstant ; cinq d'entre elles peuvent être occupées par des expressions référentielles : le sujet, le complément d'objet, l'attribut, le complément du nom et le circonstant (*cf*. He *et al*., 何卫 *et al*., 2013 : 80). De manière très sommaire, Lu (陆俭明, 2013 : 10) propose la figure suivante pour comparer les correspondances entre la catégorie grammaticale et la fonction syntaxique dans les deux langues ; on y constate que les corrélations tendent à être univoques en français par rapport à celles du chinois (*cf*. aussi Xia *et al*., 2000 ; Guo, 郭锐, 2001 : 496). Nous supposons ainsi que les influences de la fonction syntaxique sur la variation d'expressions référentielles chinoises ne sont pas les mêmes que celles du français, supposition qui devra être confirmée par les résultats de l'annotation d'un corpus.

Figure (0.1) : Corrélations entre les catégories grammaticales et les fonctions syntaxiques en français et en chinois selon Lu (陆俭明, 2013 : 10, nous traduisons)



Face aux multiples discussions relatives aux chaînes de référence, nous nous posons quelques questions sur cette notion selon l'approche contrastive franco-chinoise :

- Quelle divergence compositionnelle peut-on éventuellement observer entre les chaînes de référence en français et en chinois et pour quelles raisons ?
- Au vu des configurations des chaînes de référence en français déjà constituées, est-il possible de reprendre les mêmes pour celles du chinois, ou faudrait-il en constituer d'autres selon les particularités de chaque langue ?
- À quelle position phrastique privilégiée et à quelle fréquence dans les chaînes de référence en français et en chinois les expressions référentielles et coréférentielles apparaissent-elles, et pour quelles raisons ?
- Excepté les éléments phrastiques, quels autres facteurs pourraient influencer l'emploi de catégories d'expressions référentielles en français et en chinois et de quelle manière ?
- Quelles sont les questions et les difficultés informatiques essentielles rencontrées pendant le traitement outillé des chaînes de référence et quelles peuvent être les solutions éventuelles ?

Ces questions étant posées, notre thèse se fixe trois objectifs principaux. Le premier a trait à l'aspect définitoire : comment définir la notion de *chaîne de référence* dans la linguistique chinoise ? Le deuxième est relatif à l'aspect méthodologique : comment constituer un corpus franco-chinois annoté en chaînes de référence ? Il nous semble indispensable de proposer un guide des conventions d'annotation des chaînes de référence dans un corpus franco-chinois qui pourrait être appliqué sur d'autres corpus similaires. Le troisième objectif concerne l'observation des faits : comment se réalisent les chaînes de référence en français et en chinois dans un corpus ? Nous tenterons d'observer les variations de la composition des maillons et des chaînes de référence en français et en chinois par une étude basée sur un corpus constitué et annoté selon notre guide. En résumé, nous chercherons à examiner et à rectifier les caractéristiques communes et distinctes des expressions référentielles et coréférentielles en français et en chinois, et ce à partir des

théories sur les notions de *référence* et de *chaîne de référence*, d'une part, et, de l'autre, de l'observation de textes parallèles dans un corpus constitué par nous.

Les trois moments principaux du présent travail correspondent à ces trois objectifs. Dans un premier temps, une définition en chinois de la notion de *chaîne de référence* sera proposée, ce qui permettra de compléter cette notion dans la linguistique chinoise. Dans un deuxième temps, nous chercherons à constituer une méthode conventionnelle et reproductible pour le dépouillement des expressions référentielles et coréférentielles dans les corpus franco-chinois. Enfin, dans un troisième temps, de la comparaison des chaînes de référence obtenues dans notre corpus, nous dégagerons des particularités communes et différentes des chaînes de référence en français et en chinois dans des conditions textuelles et discursives spécifiques.

- Présentation du plan

Afin de répondre aux questions susmentionnées, nous divisons notre thèse en trois parties et en huit chapitres.

Partie I : Problèmes théoriques relatifs aux chaînes de référence dans les approches contrastives franco-chinoises. Les premières études sur les chaînes de référence ne datent pas de plus de cinquante ans : Chastain (1975) signale, le premier, la notion de *chaîne anaphorique* pour l'anglais ; d'autres auteurs proposent par la suite des définitions analogues et leurs apports sont remarquables dans plusieurs domaines linguistiques, comme en pragmatique (Li & Thompson, 1979 ; Givón, 1983), en sémantique (Corblin, 1985, 1995), en linguistique textuelle (Charolles, 1988 ; Xu, 徐赳赳, 1990 ; Schnedecker, 1992, 1997, 2021) et en linguistique informatique (Victorri, 2005 ; Landragin, 2011, 2014). Le ***Chapitre 1 : Revue des définitions de la notion de chaîne de référence et notre position relativement aux aspects franco-chinois*** propose une présentation des notions relatives aux chaînes de référence dans les deux langues, ce qui peut susciter des équivoques selon les auteurs et selon les langues. Nous évoquerons et comparerons les définitions de plusieurs auteurs dans différentes langues et selon les approches comparatives, pour ensuite proposer notre propre définition synthétique de cette notion de sorte qu'elle s'adapte à la fois au français et au chinois. Étant donné que définir ce qu'est une *chaîne de référence* s'avère complexe, il convient d'examiner trois autres notions qui paraissent fondamentales pour mieux comprendre notre sujet : l'*endophore*, l'*exophore* et la *coréférence*. Nous analyserons donc, dans le ***Chapitre 2 : Trois notions-clés fondamentales pour examiner les problèmes des chaînes de référence relatifs à l'approche franco-chinoise***, le rôle que jouent ces notions dans l'établissement des chaînes et nous examinerons par quelles formes elles sont représentées en français et en chinois.

Partie II : Problèmes méthodologiques relatifs aux approches contrastives et à la constitution d'un corpus. Les chaînes de référence sont beaucoup discutées comme sujet d'étude dans le domaine de la linguistique et de la linguistique informatique. Certains auteurs cherchent à mettre en évidence les apports des chaînes de référence dans le domaine du TAL : les chaînes sont l'un des indices essentiels de la détection automatique des thèmes discursifs (*cf.* Longo & Todirascu, 2012), permettant d'améliorer la navigation textuelle (*cf.* Nomoto & Matsumoto, 1996), la recherche d'informations (*cf.* Longo & Todirascu, 2012), les outils d'indexation et les moteurs de recherche. Par ailleurs, les comportements des chaînes de référence influencent également la lisibilité des textes (*cf.* Todirascu *et al.*, 2017 : 36). Faute de pouvoir aborder la question des chaînes de référence de manière générale dans les deux

langues, nous nous contentons d'amorcer des discussions à partir d'une étude de cas, dont le corpus comporte des textes parallèles de la plus grande taille possible que nous avons pu traiter de la manière la plus exhaustive possible. Les textes pertinents pour constituer un corpus bilingue franco-chinois sont contraignants pour des raisons scientifiques et juridiques. Scientifiquement, tous les genres de textes ne permettent pas de constituer des textes parallèles en relation de traduction. Légalement, toutes les œuvres littéraires ne peuvent pas être citées partiellement ou intégralement en raison des droits de publication. Le roman-fleuve *Jean-Christophe* de Romain Rolland et sa traduction en chinois par Lei Fu sont donc choisis pour cette raison : ces œuvres datant de plus de 80 ans, les questions de droits de réserve de l'auteur et du traducteur ne se posent plus. Cela nous permettra d'en publier des extraits ou l'intégralité des sections sans souci juridique. Notre objectif est donc de constituer un corpus annoté en expressions référentielles et en chaînes de référence. Cette partie présentera la constitution d'un corpus parallèle franco-chinois à partir des extraits du roman-fleuve *Jean-Christophe*, y compris les raisons du choix des textes, leur traitement outillé, ainsi que les conventions d'annotation reproductibles pour les deux langues. Nous envisageons de répondre aux questions méthodologiques suivantes :

- Quel est l'état actuel des corpus franco-chinois ?
- Quels sont les traitements automatiques et manuels accessibles pour dépouiller les expressions référentielles en français et en chinois ?
- Pourquoi *Jean-Christophe* constitue-t-il un corpus pertinent pour étudier les expressions référentielles et les chaînes de référence dans les deux langues ?
- Quelle démarche de traitement adopter pour le dépouillement du corpus ?

Le ***Chapitre 3 : Motivations et représentativité en termes du choix d'un corpus*** présente les raisons pour lesquelles nous choisissons cette œuvre comme corpus. Les outils de traitement de corpus sont abordés dans le ***Chapitre 4 : Outils pour la constitution d'un corpus bilingue annoté en expressions référentielles et en chaînes de référence***. Après s'être formée à l'utilisation de ces outils, nous entamerons, dans le ***Chapitre 5 : Processus du traitement outillé des expressions référentielles et des chaînes de référence en français et en chinois***, l'annotation du corpus selon le guide des conventions d'annotation élaboré et justifié par nous.

Partie III : Observations sur le corpus* Jean-Christophe *: régularités et spécificités contrastives des chaînes de référence en français et en chinois. Cette partie, dédiée aux résultats, cherche à montrer, à analyser et à expliquer les points communs et les différences entre les chaînes de référence selon les approches quantitatives, qualitatives et comparatives au sein du corpus bilingue *Jean-Christophe*. Les faits observés seront exposés en trois temps. Tout d'abord, d'après les chiffres et les exemples obtenus, nous présenterons de manière quantitative et qualitative les particularités des chaînes de référence dans *Jean-Christophe* en français, cela constituant le ***Chapitre 6 : Régularités des chaînes de référence dans un roman-fleuve français : aspects quantitatifs et qualitatifs***. Dans le chapitre suivant, consacré aux faits en chinois, ***Chapitre 7 : Propriétés des chaînes de référence dans un roman-fleuve traduit en chinois : approches contrastives franco-chinoises***, les points communs entre les chaînes de référence en chinois et en français seront évoqués de manière très brève ; nous mettrons en relief leurs différences en proposant des explications par diverses approches. Pour terminer, le ***Chapitre 8 : Synthèse comparative*** proposera une synthèse des propriétés contrastives des chaînes de référence dans le corpus sur la base des chiffres et des exemples déjà évoqués dans les deux chapitres précédents.

- Encadrement de la thèse

Notre recherche s'inscrit dans le cadre du projet *Description et MOdélisation des Chaînes de Référence : outils pour l'Annotation de corpus (en diachronie et en langues comparées) et le Traitement automatique* (*DEMOCRAT*) de l'ANR (ANR-15-CE-38-0008), dont l'objectif principal aura été, d'une part, de constituer le plus grand corpus annoté de chaînes de référence en français, et, d'autre part, d'améliorer la détection automatique des expressions coréférentielles dans les outils (*cf*. Landragin, 2021). Ce projet a duré quatre ans (2016-2020) et a couvert les premières années de notre projet de thèse.

Le corpus DEMOCRAT comprend une sélection de textes écrits en diachronie entre le 9ème et le 21ème siècle, issus de divers genres discursifs et en plusieurs langues. Quatre objectifs sont visés dans ce projet :

> *(i) Construire un modèle théorique de la référence et de la coréférence, qui tienne compte des aspects discursifs des CR et réponde aux critiques sur l'état de l'art linguistique ;*
>
> *(ii) Constituer un corpus annoté en CR, comprenant plusieurs états de la langue française, depuis le Xème jusqu'au XXIème siècle, intégrant des textes de différents genres textuels ;*
>
> *(iii) Améliorer et adapter les outils de linguistique de corpus pour qu'ils tiennent mieux compte d'annotations complexes comme celles des CR ;*
>
> *(iv) Réaliser un ou plusieurs – en fonction des techniques utilisées – systèmes de TAL pour l'identification automatique des CR.*
>
> *(Schnedecker* et al.*, 2017 : 11)*

Les travaux scientifiques se sont déroulés sur quatre axes : (i) analyses linguistiques de la diachronie ; (ii) analyses linguistiques entre les genres ; (iii) analyses linguistiques entre les langues ; (iv) analyses informatiques. Notre thèse appartient au troisième axe du projet.

Ses membres viennent de trois laboratoires : LaTTICe de Paris III, LiLPa de l'Université de Strasbourg et ICAR de l'ENS de Lyon. Plusieurs thèses doctorales sont affiliées à ce projet, comme celle de Delaborde (2020) sur les chaînes de coréférence non-stricte, celle de Hou (2021) sur la saillance référentielle en français et en chinois, celle de Kim (2023) sur les chaînes de référence et les expressions référentielles en français et en coréen, et la nôtre.

- Apports de notre thèse

Notre thèse aborde la notion linguistique dite *chaîne de référence* par l'approche comparative franco-chinoise appliquée à un corpus spécifique constitué d'extraits du roman-fleuve *Jean-Christophe.* Il nous semble que cette notion n'est pas suffisamment définie et étudiée dans la linguistique chinoise et dans le monde sinophone ; le premier apport du présent travail est donc de la présenter en en donnant une définition synthétique, de sorte qu'elle soit utile pour les études de l'analyse textuelle du chinois. En proposant une définition de cette notion, nous espérons que notre guide des conventions d'annotation pourra servir à des études ultérieures. Il est souhaitable que ce guide puisse être examiné et discuté par les linguistes informaticiens afin que soient améliorés les algorithmes des outils pour la détection des chaînes de référence, du topique discursif et d'autres facteurs cruciaux dans différents textes. Enfin, avec les chaînes de référence obtenues dans notre corpus en français et en

chinois, il est fort possible que nos résultats aident à améliorer les explications relatives à la cohérence et à la cohésion dans la grammaire textuelle, ce d'une manière plus claire et plus évidente, pour que les apprenants de langues comprennent mieux le fonctionnement des mots, des expressions, des SN, des propositions, des phrases, des séquences textuelles et celui du discours en général.

Partie I

Problèmes théoriques relatifs aux chaînes de référence dans les approches contrastives franco-chinoises

Chapitre 1 : Revue des définitions de la notion de *chaîne de référence* et notre position relativement aux aspects franco-chinois

La notion de ***chaîne de référence*** a été discutée par les linguistes dans différentes langues et sous diverses dénominations ; plusieurs définitions convergentes en ont été données, avec certaines nuances divergentes. Nous aborderons dans cette partie quelques notions similaires à celle de ***chaîne de référence*** et leurs définitions plus ou moins différentes selon les auteurs et qui ont évolué au fils du temps, à partir desquelles nous proposerons la nôtre, pour le français et pour le chinois.

1.1. *Chaîne anaphorique* de Chastain (1975)

La notion de ***chaîne anaphorique*** est la première notion similaire à celle de ***chaîne de référence***. Précisons que la notion proposée dans cette étude de Chastain constitue la base des études ultérieures relatives à la ***chaîne***. Chastain (1975) a défini pour la première fois, avant l'apparition du terme de ***chaîne de référence***, ce qu'est une ***chaîne anaphorique*** (en anglais ***anaphoric chain***), notion similaire à celle de ***chaîne de référence***, qui a conditionné cette dernière. La définition de cette première en anglais est la suivante :

> *Let an* anaphoric chain *be a sequence of singular expressions occurring in a context, such that if one of them refers to something then all of the others also refer to it. (Chastain, 1975 : 204 – 205)*

Voici la traduction française proposée par Corblin (1995) :

> *Appelons* chaîne anaphorique *une séquence d'expressions singulières apparaissant dans un contexte telles que si l'une de ces expressions réfère à quelque chose, toutes les autres y réfèrent également. (Chastain, 1975 : 204 – 205, traduit par Corblin, 1995 : 151)*

Selon cette définition de Chastain, l'extrait [1] comporte 4 chaînes anaphoriques en anglais, qu'il annote comme suit :

- chaîne 1 « an ARVN officer » renvoyant au premier personnage (3 expressions) : « an ARVN officer », « he » et « the officer » ;
- chaîne 2 « a young prisoner » renvoyant au second personnage (5 expressions) : « a young prisoner », « the prisoner », « the prisoner », « him » et « the prisoner » ;
- chaîne 3 « a wall » renvoyant au même objet (2 expressions) : « a wall » et « the wall » ;
- chaîne 4 « the beating » renvoyant à la même action (2 expressions) : « the beating » et « the beating ».

[1] At eleven o'clock that morning, [**an ARVN officer**] stood [**a young prisoner**], bound and blindfolded,
chaîne-1-expression-1 *chaîne-2-expression-1*
up against [**a wall**]. [**He**] asked [**the prisoner**] several questions, and, when [**the**
chaîne-3-expression-1 *chaîne-1-expression-2* *chaîne-2-expression-2* *chaîne-2-expression-3*
prisoner] failed to answer, beat [**him**] repeatedly. An American observer who saw [**the beating**]
chaîne-2-expression-4 *chaîne-4-expression-1*
reported that [**the officer**] « really worked [**him**] over. » After [**the beating**], [**the prisoner**] was
chaîne-1-expression-3 *chaîne-2-expression-4* *chaîne-4-expression-2* *chaîne-2-expression-5*
forced to remain standing against [**the wall**] for several hours.
chaîne-3-expression-2

(extrait d'un livre cité par Chastain, 1975 : 205, nous annotons et soulignons)

D'après ces exemples proposés par Chastain, une ***chaîne anaphorique*** comprend au minimum deux expressions qui constituent ensemble une ***anaphore*** ; il s'agit là de la première notion élémentaire d'une chaîne, au sujet de laquelle de nombreuses discussions seront présentées dans le ***Chapitre 2*** du présent travail. D'autre part, les composants des chaînes anaphoriques renvoient à des entités de différentes catégories sémantiques : deux personnes (« un officier », « un prisonnier »), un objet (« un mur ») et une action (« un battement »). Ces trois types d'entités existantes dans le monde réel sont appelés ***référents*** ; le verbe ***référer*** et le substantif ***référence*** constituent donc le second groupe de notions élémentaires, que nous présenterons dans le ***Chapitre 2***.

1.2. *Chaîne topicale* de Li & Thompson (1979) et de Givón (1983)

À la différence de Chastain, Li & Thompson (1979) et Givón (1983) abordent la notion de ***chaîne*** selon une approche discursive-fonctionnelle en introduisant la notion de ***topique*** comme élément fondamental. Nous présentons tout d'abord la définition de ***topique*** et ensuite celle de ***chaîne topicale*** selon ces auteurs.

1.2.1. Définitions : *topique de discours* et *chaîne topicale*

Avant de parler de ***chaîne topicale***, il convient d'examiner en premier lieu ce qu'est un ***topique***. De nombreuses définitions de cette notion discursive-fonctionnelle figurent dans les études antérieures ; considérons-en une plus récente comme l'est celle de la ***Grande Grammaire du français*** :

> *[Le topique de discours] Il peut s'agir d'**un référent sur lequel le discours va apporter**, dans son déroulement, un ensemble d'informations ; il peut également s'agir d'une question plus générale, du sujet que traite le discours et qui n'est pas obligatoirement exprimé par un syntagme particulier. Que le topique soit **une unité particulière ou la question traitée par le discours**, il peut y avoir, au fil du texte, une progression qui se déploie en sous-topiques, lesquels en constituent des aspects particuliers (sous-parties, propriétés…). (Abeillé & Godard (éds), 2021 : 1931, nous soulignons)*

Un discours peut donc comporter deux types de topique : un référent exprimé, ou un sujet non exprimé. Cette notion fondamentale de topique conduit Li & Thompson (1979) et Givón (1983) à proposer la notion de ***chaîne topicale***[1] : toutes les expressions renvoyant au topique d'un discours constituent une chaîne topicale. Dans l'exemple de Li & Thompson (1979), une chaîne topicale est composée majoritairement de zéros : la première expression d'une chaîne sert à indiquer le référent topical qui n'est pas repris par des formes explicites

[1] Dans les deux études écrites en anglais, les termes d'origine sont ***topic chain*** selon Li & Thompson (1979) et ***topic strand*** selon Givón (1983). Les deux termes sont traduits par ***chaîne topicale*** en français.

dans le discours suivant, comme le montre l'extrait [2][2] : la chaîne topicale « Yang-Zhi » est composée d'un nom propre à la place du premier maillon et d'une suite de zéros dans le reste du discours :

[2]	**[杨志]**	取路。		不是	日，	**[Ø]**	来到	东京，	**[Ø]**	入得	城来，	**[Ø]**	选	个	客店，				
	yangzhi	*qulu*		*bushi*	*ri*		*daidao*	*dongjing*		*rude*	*chenglai*		*xuan*	*ge*	*kedian*				
	Yang-Zhi	take-to-the-road		not-many	day		arrive	Dong-Jing		enter	city		find	a	hotel				
	npr					*zéro-1*			*zéro-2*			*zéro-3*							
	[Ø]	安息	下。																
		anxi	*xia*																
		settle	down																
	zéro-4																		

'Yang-Zhi took to the road. In a few days, (he) arrived in Dong-Jing. (He) entered the city. (He) found a hotel. (He) settled down.'

(Shi Nai'an, 施耐庵, *Shui-Hu-Zhuan*, cité par Li & Thompson, 1979 : 313)

Selon les exemples de Givón (1983), une ***chaîne topicale*** doit tout d'abord inclure au minimum deux expressions qui renvoient à une même entité. Cette caractéristique ressemble à celle de la chaîne anaphorique de Chastain. Comme le montrent ses exemples en anglais, une chaîne topicale peut être composée de « John ... he » dans l'extrait [3] et de « he ... John » dans l'extrait [4] :

[3] **[John]**, **[he]** came yesterday.
[4] **[He]** came yesterday, **[John]**.

(Givón, 1983 : 6)

Ensuite, Givón indique que les catégories des composants potentiels d'une chaîne topicale sont nombreuses : SN[3] (extraits [5]), pronom (extraits [6]), zéro (extraits [7]) ainsi que les accords grammaticaux[4] peuvent tous être maillons d'une chaîne topicale.

[5] ... Now **[the man]**, **[he]** never joined ...
SN *pro*

[6] (... **[He]** came in;) **[he]** then sat down ...
pro *pro*

[7] (... **[He]** came in and) **[Ø]** sat down ...
pro *zéro*

(Givón, 1983 : 6, nous annotons)

D'après les exemples proposés par ces auteurs, le topique des ***chaînes topicales*** appartient plutôt au premier type de la définition de la *Grande Grammaire du Français*, à savoir un référent spécifique qui se situe à la place la plus importante par rapport à celle de tous les autres référents dans le même discours.

[2] Dans notre thèse, nous annotons les exemples en chinois par trois lignes : la première, en italique, indique la prononciation des caractères chinois ; la deuxième donne la traduction littérale mot-à-mot ; la troisième, en italique, mentionne annotations des parties que nous aimerions mettre en relief dans chaque exemple.

[3] Le SN, *i.e.* syntagme nominal, désigne « le type particulier de syntagme dont la tête catégorielle est un nom et dont l'autre constituant obligatoire est un déterminant, » (Riegel, 2000 : 5) avec certains modifieurs facultatifs.

[4] Givón ne propose pas d'exemples d'accords grammaticaux dans son article. Il nous paraît donc susceptible de considérer les accords grammaticaux comme parties des chaînes topicales.

1.2.2. Deux positions face aux notions de *topique* et de *chaîne topicale*

Les deux catégories discursives évoquées dans la ***Grande Grammaire du Français*** mises à part, la notion de ***topique*** est très discutée dans de nombreuses études ; deux positions sont le plus souvent évoquées. La première est celle de Givón (1983), qui propose, dans un premier temps, la notion de ***chaîne topicale*** dans le cadre phrastique, à savoir qu'une chaîne topicale peut se situer à l'intérieur d'une seule phrase en distinguant deux parties : ***topique*** et ***propos***[5]. La continuité du référent dépend non seulement de la forme morpho-syntaxique, mais aussi de la fonction syntaxique de l'expression ; le topique peut ainsi être appelé ***topique phrastique*** selon ce point de vue. Sachant que le chinois est une langue dominée par le topique au lieu du sujet syntaxique, le topique phrastique y est souvent comparé avec le sujet syntaxique en français. Les expressions concernées se rapprochent relativement du sujet syntaxique (*cf.* Rohde & Kehler, 2014 : 912), surtout les topiques phrastiques en chinois (*cf.* Fang, 方梅, 2018 ; Xu, 许余龙, 2003 ; He, 2013). La ***Grande Grammaire du Français*** définit ***topique*** et ***propos*** comme suit :

> ***Dans le contenu principal de la phrase, on peut généralement distinguer ce dont parle la phrase, appelé*** topique***, et ce qu'elle en dit, qui est appelé*** propos***. (Abeillé & Godard (éds), 2021 : 99)***

La seconde position est celle que Givón (1983) développe dans un second temps, ainsi que celle de nombreux autres auteurs dans une perspective inter-phrastique, comme notamment Combettes (1983), Kleiber (1994a) et Adam (1994, 2015). Selon ce point de vue, une chaîne topicale se situe dans plusieurs phrases ou dans l'ensemble d'un texte et elle peut garantir la continuité et la cohérence du discours ; le topique est ainsi appelé ***topique discursif*** ou ***textuel***. Cette position est également adoptée par de nombreux linguistes chinois, comme Xu (徐赳赳, 2003, 2014), Qu (屈承熹, 2006), He (2013) Fang (方梅, 2018) et Hou (2021). Nous les présenterons dans la partie suivante. Citons un exemple français d'Adam (2015) et un exemple chinois de He (2013) de chaînes topicales qui se situent dans plusieurs phrases selon les auteurs :

[8] **[Trois forçats]** [**se**] procurent des revolvers. [**Ils**] tuent [**leur**] geôlier et [**Ø**] s'emparent des clefs de la prison. [**Ils**] se précipitent hors de leurs cellules et [**Ø**] tuent quatre gardiens dans la cour.
(Adam, 2015 : 76, nous annotons)

[9] [**我**] 没 有 借口，[**我**] 无 法 解释，[**我**] 不 敢 求 你 谅佑。

wo	*mei*	*you*	*jiekou*	*wo*	*wu*	*fa*	*jieshi*	*wo*	*bu*	*gan*	*qiu*	*ni*	*liangyou*
1P	NEG	avoir	excuse	1P	NEG	pouvoir	expliquer	1P	NEG	oser	demander	2P	pardonner[6]

'**Je** n'ai pas d'excuse. **Je** ne peux pas m'expliquer. **Je** n'ose pas te demander de me pardonner.'
(Zhong-Shu Qian, 钱锺书, *Fortress Besieged*, 1947, cité par He, 2013 : 345, nous traduisons et annotons)

---

[5] Selon les cadres théoriques de divers auteurs, le *topique* est aussi appelé « thème » (Halliday & Hasan, 1976) et « information connue » (Prince, 1981, cité et traduit en français par Combettes, 1986 et Kleiber, 1994) ; le *propos* est aussi appelé « commentaire » (Givón, 1983 ; Adam, 1994, 2015), « focus » (Gundel *et al.*, 1993), « rhème » (Halliday & Hasan, 1976) et « information nouvelle » (Prince, 1981, cité et traduit en français par Combettes, 1986 et Kleiber, 1994).

[6] La traduction littérale des exemples chinois respecte la catégorie morpho-syntaxique d'expressions en chinois, c'est-à-dire que si une expression chinoise ne contient pas de désinence de nombre et de genre, la glose ne présente pas non plus ces marqueurs morpho-syntaxiques, qui ne sont pas propres au chinois. Toutes les gloses des exemples dans notre thèse respectent ce principe. Voir les explications des signes abrégés dans les ***Abréviations***.

Dans l'énoncé [8], la suite de sept expressions anaphoriques (« trois forçats », « se », « ils », « leur », « Ø », « ils » et « Ø ») constitue une chaîne topicale. Dans l'énoncé [9], les trois '1P' ***wo*** 我 constituent une chaîne topicale en chinois dans les trois phrases successives.

1.2.3. *Chaîne topicale* en linguistique chinoise

La ***chaîne topicale*** semble être une notion intéressante et pertinente pour deux raisons. Premièrement, de nombreux auteurs ont étudié les chaînes topicales spécifiquement pour le chinois, langue dominée par le topique, contrairement au français, dominé par le sujet. L'adaptation de cette notion au chinois par Xu (徐赳赳, 2003, 2014) attire particulièrement l'attention en raison des étiquettes et configurations formelles des composants dans les chaînes. Xu nomme les expressions de chaîne ***nœuds*** (节点) (2014 : 316). La première expression est un ***nœud de source*** (根节点) ; les expressions suivantes sont des ***nœuds de reprise*** (子节点). D'après ses définitions, une chaîne topicale est composée d'au moins un nœud de source et d'un nœud de reprise, comme le montre l'extrait [10] qui contient une chaîne renvoyant à la personne « Xin-Zhen-Bei » : un nœud - source 'Xin-Zhen-Bei, un élève de cinquième année de l'Ecole Huai-Rou-Yan-Xi' ***Huairouyanxi-xuexiao-wu-nianji-xuesheng-Xinzhenbei*** 怀柔雁栖学校五年级学生信振北 et un nœud de reprise 'Xin-Zhen-Bei, 12 ans' ***shiersui-de-Xinzhenbei*** 12 岁的信振北 :

[10]前天，	**[怀柔雁栖**		**学校**	**五**		**年级**	**学生**	**信振北]**	为	救	落水				
qiantian	*Huairouyanxi*		*xuexiao*	*wu*		*nianji*	*xuesheng*	*Xinzhenbei*	*wei*	*jiu*	*luoshui*				
avant-hier	Huai-Rou-Yan-Xi		école	cinquième		année	élève	Xin-Zhen-Bei	pour	secourir	tomber-eau				
	nœud-source														
同学	宋春阳		而	坠落	水	中，	宋春阳		得救，	但	**[12**	**岁**	**的**		
tongxue	*Songchunyang*		*er*	*zhuiluo*	*shui*	*zhong*	*Songchunyang*		*dejiu*	*dan*	*shier*	*sui*	*de*		
camarade	Song-Chun-Yang		donc	tomber	eau	dedans	Song-Chun-Yang		être-secouru	mais	12	an	DE		
											nœud-reprise				
信振北]	溺水身亡。														
Xinzhenbei	*nishuishenwang*														
Xin-Zhen-Bei	se-noyer														

'Avant-hier, Xin-Zhen-Bei, un élève de cinquième année de l'école Huai-Rou-Yan-Xi, a plongé dans l'eau pour sauver son camarade Song-Chun-Yang. Song-Chun-Yang a été secouru, mais Xin-Zhen-Bei, 12 ans, s'est noyé.'

(Le 4 mai 2007, *Quotidien Jinghua*, 京华时报, cité par Xu, 徐赳赳, 2014 : 316, nous traduisons et annotons)

Xu précise que les SN[7], les pronoms ainsi que les zéros peuvent jouer le rôle de ***nœuds*** d'une chaîne topicale, comme le montre l'extrait [11] où celle de « Bi-Zhong » est composée de nœuds en forme de pronom, de nœuds zéro et d'un nœud en forme de nom propre :

[11]**[他]**，	不	是	孩子	了，	**[Ø]**		已经	20	岁	了；	**[他]**		不稚嫩	了，					
ta	*bu*	*shi*	*haizi*	*le*			*yijing*	*ershi*	*sui*	*le*	*ta*		*buzhinen*	*le*					
3P	NEG	être	enfant	LE			déjà	20	an	LE	3P		immature	LE					
nœud-source-pro					*nœud-reprise-zéro*						*nœud-reprise-pro*								
[Ø]	身高		1	米	87，	**[Ø]**		体重	达		110	公斤。	尽管	这样，					
	shengao		*yi*	*mi*	*baqi*			*tizhong*	*da*		*yibaiyishi*	*gongjin*	*jinguan*	*zheyang*					
	se-mesurer		1	mètre	87			se-peser	jusque		110	kilogramme	malgré	cela					
nœud-reprise-zéro						*nœud-reprise-zéro*													

[7] Les SN dont Xu parle ici concernent les noms communs et les noms propres puisque ces deux types de noms ne se distinguent pas par l'initiale de leur forme.

[他] 毕竟 是 个 孩子，总 [Ø] 喜欢 做梦，[Ø] 总是
ta *bijing* *shi* *ge* *haizi* *zhong* *xihuan* *zuomeng* *zongshi*
3P quand-même être CL enfant toujours aimer rêver toujours
nœud-reprise-pro *nœud-reprise-zéro* *nœud-reprise-zéro*

想 看 那 71 米 08 的 纪录，[Ø] 总是 盼着 超过 这 个 纪录。
xiang *kan* *na* *qishiyi* *mi* *lingba* *de* *jilu* *zongshi* *panzhe* *chaoguo* *zhe* *ge* *jilu*
vouloir voir ce-là 71 mètre 08 DE record toujours espérant dépasser ce CL record
nœud-reprise-zéro

怨 谁 呢？谁 叫 [他] 爱上 了 链球？一 个 优秀 运动员 谁
yuan *shui* *ne* *shui* *jiao* *ta* *aishang* *le* *lianqiu* *yi* *ge* *youxiu* *yundongyuan* *shui*
se-plaindre qui INTERR qui demander 3P adorer LE marteau un CL excellent athlète qui
nœud-reprise-pro

不 想 破 纪录，拿 金 牌？这 是 很 自然的 事情。这 就是 [毕忠]。
bu *xiang* *po* *jilu* *na* *jin* *pai* *zhe* *shi* *hen* *zirande* *shiqing* *zhe* *jiushi* *Bizhong*
NEG vouloir dépasser record gagner or médaille ce être très naturel chose ce être Bi-Zhong
nœud-reprise-npr

这 就是 他的 链球 梦。
zhe *jiushi* *tade* *lianqiu* *meng*
ce être son marteau rêve

'Ce n'est plus un enfant, il a déjà 20 ans, il n'est plus immature, il mesure 1m 87 et pèse 110 kilos. Malgré cela, c'est un enfant après tout, il aime toujours rêver, il a toujours en vue le record de 71m 08 et il a toujours l'espoir de battre ce record. À qui se plaindre ? Qui l'a fait tomber amoureux du marteau ? Un athlète d'excellence qui ne veut pas battre des records et gagner une médaille d'or ? C'est une chose naturelle. C'est Bi-Zhong. C'est son rêve, le marteau.'

(Xu, 徐赳赳, 2003 : 109, nous traduisons et annotons)

La deuxième raison pour laquelle la notion de ***chaîne topicale*** paraît intéressante et appropriée est qu'elle est abondamment discutée en linguistique chinoise dans la perspective de l'approche contrastive. Du point de vue discursivo-fonctionnel, le topique est particulièrement étudié en chinois, par opposition avec les autres langues. Dans la typologie des langues de Li & Thompson (1976 : 466), le français appartient au groupe des langues dites « régies par le sujet » (***subject dominate language*** ou ***subject prominent language***), tandis que le chinois appartient à celui des langues dites « régies par le topique » (***topic dominate language*** ou ***topic prominent language***)[8]. Dans les langues du premier groupe, le sujet syntaxique est obligatoirement exprimé afin d'avoir une structure prototypique de phrase ; il l'emporte sur l'importance de la structure pragmatique du discours. Les analyses des langues régies par le sujet portent donc le plus souvent sur les fonctions syntaxiques. Au contraire, le sujet syntaxique, dans les langues du second groupe, est optionnel : sa place n'est habituellement remplie par aucune expression explicite, comme le montrent les nombreux zéros dans l'extrait [11]. Par conséquent, la comparaison des chaînes topicales du chinois avec celles du français pourrait présenter des différences radicales lorsqu'il s'agit de la densité référentielle, de la longueur moyenne des maillons et de la proportion des composants, que nous observerons dans la ***Partie III*** de notre thèse.

[8] Dans la typologie analogue de Tsao (1979, cité par Xu, 许余龙, 2010 : 159) et Zhang (张伯江, 2013), les auteurs signalent que les langues dominées par le topique sont orientées par le discours (discourse-oriented), alors que les langues dominées par le sujet le sont par la phrase (sentence-oriented). Dans la typologie similaire de Huang (1994 : 159), les langues du groupe « subject prominent language » sont dites du ***style anglais*** (***English-style topic structure***) ; celles du groupe « topic prominent language » sont dites du ***style chinois*** (***Chinese-style topic structure***).

1.3. *Chaîne de référence* de Corblin (1985, 1995)

Le terme de ***chaîne de référence***, en français, est utilisé pour la première fois par Corblin (1985). Dans une de ses œuvres ultérieures (1995), l'auteur signale deux sous-types de chaînes de référence (1995 : 174) : les ***chaînes d'identité***, équivalentes aux chaînes de référence que nous projetons d'étudier dans notre thèse, et les ***chaînes d'association***. Les premières se basent sur la notion d'***anaphore*** dite ***stricte*** (« un garçon … le garçon », Corblin, 1995 : 165) ; la relation inter-maillonnaire des secondes est constituée par l'***anaphore*** dite ***associative*** (« une maison … le toit », Corblin, 1995 : 165). L'anaphore associative se distingue de l'anaphore stricte puisque les expressions concernées dans cette première ne renvoient pas au même référent ; ce type de chaîne non coréférentielle ne constitue donc pas notre sujet d'étude. Nous présentons la définition des chaînes de référence du premier type de Corblin, à savoir les ***chaînes d'identité référentielle***. La définition ainsi que les discussions sur les notions d'***identité référentielle*** et de ***coréférence*** seront détaillées dans le ***Chapitre 2*** du présent travail.

En s'appuyant sur la première définition de ***chaîne anaphorique*** de Chastain, Corblin (1985) réajuste et développe cette notion en lui conférant une nouvelle étiquette, celle de ***chaîne de référence***, qu'il définit comme suit :

> ***Convenons d'appeler*** chaîne de référence ***la suite des expressions d'un texte entre lesquelles l'interprétation construit une relation d'identité référentielle. (Corblin, 1985 : 123)***

Selon cette définition, l'extrait [12] en français comporte 2 chaînes de référence :

- chaîne 1 « Segalen » (6 expressions), renvoyant au même personnage : « Segalen », « sa », « ses », « sa », « Segalen » et « lui-même » ;
- chaîne 2 « Flaubert » (3 expressions), renvoyant au même personnage : « Flaubert », « l'ermite de Croisset » et « son ».

[12]Une des plus belles œuvres de [**Segalen**] est certainement comme c'est le cas pour [**Flaubert**],
chaîne-1-expression-1 *chaîne-2-expression-1*

[**sa**] correspondance (en très grande partie inédite). Souvent éloigné de [**ses**] amis les plus
chaîne-1-expression-2 *chaîne-1-expression-3*

proches, voire de [**sa**] femme, [**Segalen**] leur écrivait et, tout comme [**l'ermite de Croisset**]
chaîne-1-expression-4 *chaîne-1-expression-5* *chaîne-2-expression-2*

livrait les secrets de [**son**] œuvre, levait le voile qui masquait certaines régions de [**lui-même**].
chaîne-2-expression-3 *chaîne-1-expression-6*

(Claude Courtot, *Victor Segalen*, cité par Corblin, 1985 : 123, nous annotons)

Ces exemples montrent qu'une chaîne de référence, selon la définition de Corblin, doit répondre à un impératif de quantité, c'est-à-dire qu'elle doit comporter au moins trois expressions, alors qu'une chaîne anaphorique peut en comporter seulement deux, comme nous l'avons vu avec les chaînes anaphoriques de « a wall » et de « the beating » dans l'extrait [1]. Cette différence quantitative entre une anaphore et une chaîne de référence est soulignée par Corblin comme suit :

> ***Chaîne*** ***permet de dépasser les contextes de simple succession de deux termes auxquels se limite le plus souvent le linguiste qui sort du domaine phrastique. (Corblin, 1987 : 7, nous soulignons)***

Les composants des chaînes sont appelés ***maillons*** par de nombreux auteurs, comme notamment Corblin (1985, 1995), Charolles (1988), Schnedecker (1997, 2014, 2017, 2019,

2021), Cornish (1998) et également par le ***Manuel d'annotation du projet DEMOCRAT***[9] (2018) ; nous utilisons nous aussi ce terme pour désigner les composants en question, car il permet de distinguer les expressions incluses dans les chaînes de celles qui ne sont que simplement référentielles.

Selon la définition de Corblin (1985), une chaîne peut être anaphorique lorsque la première expression est nominale et qu'elle est suivie par des expressions non nominales ; si la chaîne est initiée par une expression non nominale, elle n'est pas anaphorique parce que les deux premières expressions ne constituent pas une anaphore.

Afin de bien marquer la différence entre la ***chaîne anaphorique*** de Chastain et la ***chaîne de référence*** de Corblin, précisons que si les deux notions permettent d'établir une relation d'identité référentielle à l'intérieur d'un contexte linguistique, la seconde permet en outre de constituer une relation entre les éléments linguistiques et les entités extra-linguistiques dans le monde réel. Par ailleurs, une ***chaîne anaphorique*** se distingue d'une ***chaîne de référence*** par la relation primordiale entre les expressions concernées : celles des chaînes anaphoriques sont liées par la relation d'***anaphore***, alors que celles des chaînes de référence le sont par la relation d'identité référentielle.

Au niveau textuel, Corblin (2005) signale en outre que les chaînes de référence garantissent la cohérence textuelle et permettent de maintenir l'attention des locuteurs sur une entité dans le discours.

1.4. *Chaîne référentielle* de Charolles (1988)

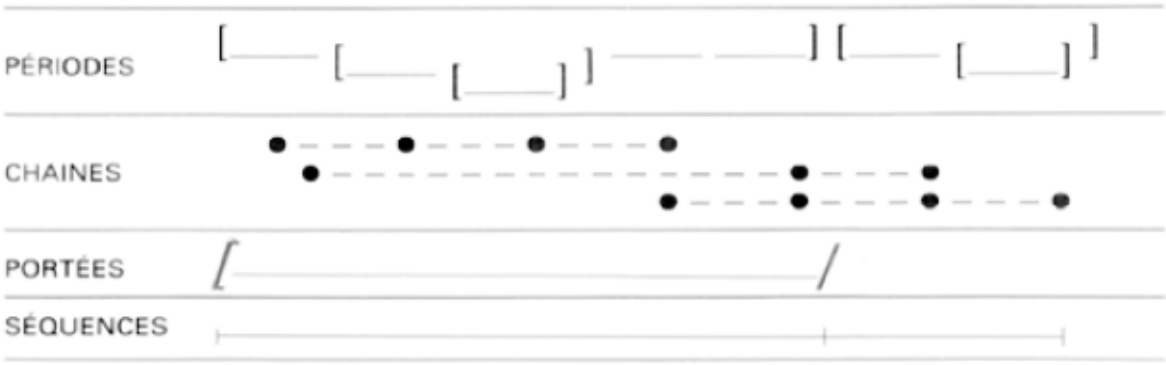
Charolles (1988) définit quatre types d'unités textuelles qui permettent de désigner l'organisation de la textualité selon différents niveaux, parmi lesquels la notion de ***chaîne*** prend sa place avec la ***période***, la ***portée*** et la ***séquence***. Leur définition s'inscrit dans le cadre textuel, comme suit :

> *La période est une unité d'énonciation dont les membres ou composants (phrastiques) entretiennent des rapports de dépendance*
>
> ***Les chaînes sont constituées par des « suites d'expressions coréférentielles.***
>
> *(La portée est) toute portion de texte dont l'interprétation est indexée comme devant s'effectuer dans un certain cadre (ou espace) de véridiction.*
>
> *Les séquences résultent du découpage du matériau discursif. Il y a création d'une séquence lorsque le sujet, par exemple écrivant, indique comment il fractionne son texte. À l'écrit, les séquences sont faciles à repérer, elles correspondent notamment au découpage en paragraphes qui est une marque linguistique comme une autre, faisant sens au même titre par exemple qu'une expression relationnelle quelconque.*
>
> *(Charolles, 1988, nous soulignons)*

La relation entre les quatre notions de Charolles est schématisée par la figure (1.1). Les quatre unités textuelles permettent ensemble d'établir la structure d'un texte. Les expressions dans les chaînes se présentent en tant que nœuds et liens garantissant la continuité textuelle.

[9] Ci-après *Manuel DEMOCRAT*.

Figure (1.1) : Les *périodes*, les *chaînes*, les *portées* et les *séquences* dans l'organisation textuelle selon Charolles (1988 : 10)



Analogue à celle de Corblin, la définition de Charolles inclut aussi la notion de *coréférence*. Selon la définition de ce dernier, l'extrait [13] cité par Charolles comporte une chaîne de « Cunégonde » formée de 8 expressions : « Cunégonde », « sa », « mademoiselle Cunégonde », « elle », « elle », « elle », « elle » et « la sienne ».

[13]Un jour, [**Cunégonde**]$_{M-1}$ en se[10] promenant auprès du château, dans le petit bois qu'on appelait « parc », vit entre des broussailles le docteur Pangloss qui donnait une leçon de physique expérimentale à la femme de chambre de [**sa**]$_{M-2}$ mère, petite brune très jolie et très docile. Comme [**mademoiselle Cunégonde**]$_{M-3}$ avait beaucoup de dispositions pour les sciences, [**elle**]$_{M-4}$ observa, sans souffler, les expériences réitérées dont [**elle**]$_{M-5}$ fut témoin ; [**elle**]$_{M-6}$ vit clairement la raison suffisante du docteur, les effets et les causes, et s'en retourna toute agitée, toute pensive, toute remplie du désir d'être savante, songeant qu'[**elle**]$_{M-7}$ pourrait bien être la raison suffisante du jeune Candide, qui pouvait aussi être [**la sienne**]$_{M-8}$.

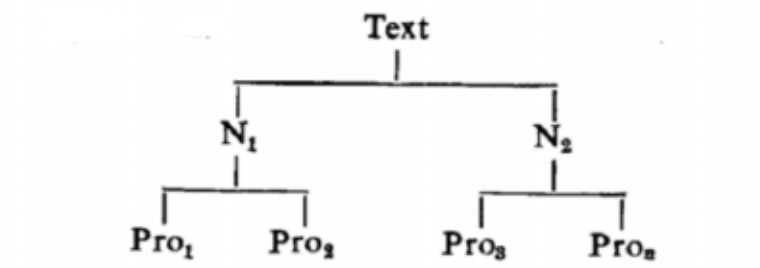
(*Candide*, cité par Charolles, 1988 : 8, nous annotons)

[10] Les pronoms réflexifs français ne sont pas considérés comme maillons des chaînes de référence dans notre thèse. Ce pour deux raisons. Premièrement, ils sont constituants des verbes pronominaux français dont la variation de la partie pronominale peut être considérée comme une partie d'accord de conjugaison. Deuxièmement, Corblin (1995) définit un type de chaînes particulières dites *liées*, c'est-à-dire que « la construction de la chaîne est exclusivement régie par la syntaxe » (1995 : 157), dont l'exemple le plus typique est celui des pronoms réflexifs français. Ces chaînes liées présentent d'autres particularités que nous ne développerons pas ici. Pourtant, les pronoms réciproques peuvent constituer des maillons de chaînes. Ils partagent les mêmes formes que les pronoms réflexifs, mais avec des interprétations sémantiques complètement différentes. En effet, les formes des pronoms réflexifs « assimilent le verbe à une relation qui unit un élément à lui-même ou chacun des éléments d'un ensemble à lui-même » (Riegel *et al.*, 2016 [1994] : 457).

1.5. *Chaîne pronominale* de Xu (徐赳赳, 1990)

À la suite des discussions portant sur la notion de *chaîne topicale*, Xu (徐赳赳, 1990) propose celle de *daici-danwei* 代词单位, signifiant littéralement 'pronom – unité', que nous traduisons par *chaîne pronominale* pour définir les séquences des expressions coréférentielles initiées par un SN, suivi par des pronoms qui coréfèrent à la même entité. C'est d'un côté ce type de répartition des SN et des pronoms, et, de l'autre, la récurrence de ces chaînes pronominales, qui permettent la continuité référentielle du texte : les SN donnent des images concrètes des entités dans l'esprit des lecteurs, alors que l'interprétation référentielle des pronoms dépend de celle des SN. Nous empruntons la figure (1.2) de Xu pour illustrer cette hiérarchie dominante entre les deux catégories de composants des chaînes dans le texte.

Figure (1.2) : Configuration des *chaînes pronominales* dans les articles journalistiques chinois selon Xu (徐赳赳, 1990 : 330)



Si une chaîne topicale est définie de manière discursive, cette définition de la chaîne pronominale fait plutôt ressortir la question de la catégorie grammaticale des composants dans les chaînes. Citons directement Xu :

> *这里的"代词单位"，我们指的是一个名词出现后，接着又出现一个或几个"他"，直到这个名词（可以同形也可以异形，只要同指就行）再出现为止。名词加上这些"他"就构成一个"代词单位"。零形式多少不予考虑。(Xu, 徐赳赳, 1990 : 330)*

> *Une « chaîne pronominale » est une séquence initiée par un SN, suivi par un ou plusieurs pronoms personnels de la troisième personne 他 'il'. Cette séquence se termine au moment où un autre SN coréférentiel apparaît, quelle que soit sa forme. Le SN initial et les pronoms suivants constituent ensemble une chaîne pronominale, sans tenir compte du nombre des zéros. (nous traduisons)*

Selon cette définition, une chaîne pronominale se termine quand un nouveau SN surgit dans le discours, qu'il soit coréférentiel au premier ou non. Nous proposons de configurer les chaînes pronominales de Xu par la formule suivante : le premier « SN1 » signifie le SN qui initie une chaîne ; le deuxième « SN1 » représente un SN coréférentiel de forme identique au premier « SN1 » ; « SN1' » signifie que ce SN est coréférentiel au premier « SN1 » mais avec une autre forme ; « SN2 » correspond à un SN non coréférentiel au premier « SN1 ». Dans les trois cas, la séquence « SN1 … pronom » est la première chaîne ; la séquence « SN1 / SN1' / SN2 … pronom » est la deuxième chaîne.

SN_1 … $pronom_1$ … $pronom_2$ … $pronom_n$ … SN_1 / SN_1' / SN_2 … pronom …

Comme le montre l'extrait [14], d'après les notions que nous avons présentées *supra*, toutes les expressions entre crochets forment une chaîne de référence selon les définitions de Corblin (1985) et de Charolles (1988), tandis que le premier nom propre, *Guwei* 顾威, et les pronoms qui suivent, constituent la première chaîne pronominale d'après la définition de Xu (徐赳赳, 1990) ; le second nom propre, 'Guwei, ayant 48 ans' *sishiba-sui-de-Guwei* 48 岁的顾威, démarre la seconde chaîne pronominale avec la suite du texte.

[14]"出力 不 讨好" 这 句 俗语 正 可 用 在 [**顾威**] 身 上。[他] 明知
chuli bu taohao zhe ju suyu zheng ke yong zai Guwei shen shang ta mingzhi
aider NEG apprécier ce CL proverbe justement pouvoir utiliser PREP Gu-Wei corps dessus 3P conscient
SN-1-chaîne-1 *pro-1*

当 演员 要 比 当 导演 容易 出名； 可 [他] 还是 非
dang yanyuan yao bi dang daoyan rongyi chuming ke ta haishi fei
être acteur devoir par-rapport être metteur-en-scène facile célèbre mais 3P toujours malgré-tout
pro-2

改行 当 导演： 许多 人 劝 [他] 别 自讨苦吃； 但 [他]
gaihang dang daoyan xuduo ren quan ta bie zitaokuchi dan ta
se-réorienter être metteur-en-scène beaucoup-de gens persuader 3P NEG se-chercher-difficulté mais 3P
pro-3 *pro-4*

终于 还是 半路出了家。[**48 岁 的 顾威**] 是 学表演的， …
zhongyu haishi banluchujiale sishiba sui de Guwei shi xuebiaoyande
enfin toujours se-réorienter-LE 48 an DE Gu-Wei être apprenant-acteur
SN-2-chaîne-2

'Le proverbe « proposer des aides sans aucun remerciement » décrit justement la situation de **[Gu-Wei]**. [Il] s'est déjà rendu compte que les acteurs sont souvent plus célèbres que les metteurs
SN-1-chaîne-1 *pro-1*

en scène, mais [il] s'est réorienté vers ce dernier métier malgré tout. Beaucoup de personnes
pro-2

[l'] ont persuadé de ne pas se créer des difficultés, mais [il] s'est décidé. **[Gu-Wei, ayant 48 ans]**,
pro-3 *pro-4* *SN-2-chaîne-2*

s'est formé comme acteur, …'
(Le 18 février 1989, *Quotidien de Pékin*, 北京日报, cité par Xu 徐赳赳, 1990 : 333, nous traduisons et annotons)

Les *chaînes pronominales* de Xu peuvent ainsi être considérées comme des séquences dans une chaîne de référence car leurs nœuds successifs coréfèrent à la même entité. Les nœuds des chaînes pronominales successives peuvent également coréférer à la même entité et constituent en conséquence une chaîne de référence.

En ce qui concerne les catégories grammaticales de nœuds incluses dans les chaînes pronominales en chinois, Xu ne considère que les SN, les noms propres et les pronoms comme composants, sans prendre en compte les zéros ou d'autres formes potentielles. D'après lui, les chaînes pronominales jouent un rôle important dans la structure textuelle parce qu'elles peuvent servir à marquer les frontières des *épisodes* (情节) dans la narration[11]. Xu précise qu'une narration est constituée par des épisodes, et que le premier maillon d'une chaîne pronominale marque souvent le début d'un nouvel épisode.

[11] Voir la définition d'*épisode* dans la section *5.1.2. Étape 2 : partition des textes en épisodes* du *Chapitre 5*.

1.6. *Chaîne de référence* de Schnedecker (1992, 1997, 2021)

La première étude sur les chaînes de référence de Schnedecker est parue en 1992 dans sa thèse doctorale en sciences du langage et en philosophie linguistique. Une série de plusieurs livres et articles sur le même sujet ont été publiés au fil du temps, notamment ceux de 1997, 2005, 2014, 2017, 2019 et 2021. Nous présenterons surtout sa thèse de 1992, son œuvre de 1997 et celle, récente, publiée en 2021, en incluant certains points tirés des articles publiés dans les intervalles.

1.6.1. Apports textuels

Tout comme Corblin (1985) et Charolles (1988), Schnedecker (1997) attire l'attention sur le rôle joué par les chaînes de référence dans la cohérence et la cohésion textuelles. Avant de proposer sa propre définition de ***chaîne de référence***, Schnedecker insiste sur l'importance de la relation entre les phrases : un texte ne se compose pas de phrases juxtaposées ; il ne suffit pas, en d'autres termes, de mettre des phrases les unes à la suite des autres pour former un texte, mais il s'agit aussi de parler de certaines entités plus saisissables que les autres par le moyen desquelles la continuité référentielle peut être réalisée. Un texte doit permettre que ***certaines entités ressortent de l'ensemble*** (Schnedecker, 1997 : 6). Pour ce faire, il faut que ces entités soient reprises ou développées d'une manière ou d'une autre au fil du discours ; si les entités restent en suspens, le texte risque d'être décousu. Schnedecker s'intéresse donc avant tout à l'ensemble des expressions qui permet la continuité référentielle, d'où la notion de ***chaîne de référence***.

Afin d'illustrer l'importance des chaînes de référence pour un texte, Schnedecker (1997) emprunte la rédaction d'un élève citée par Turco, qui constitue le fragment [15] suivant. Les expressions référentielles sont numérotées :

[15][**Sortie à [la dune]**]
réf-1 *réf-2-M-1*

[Nous] sommes partis de **[l'école]** à **[bicyclette]** pour aller à **[la côte]** découvrir **[la dune]**.
réf-3-M-1 *réf-4-M-1* *réf-5* *réf-6* *réf-2-M-2*

[L'euphorbe] est **[une plante toxique]**, **[elle]** fait gonfler **[la langue]** et **[Ø]** étouffe. **[Les plantes**
réf-7-M-1 *réf-7-M-2* *réf-7-M-3* *réf-8* *réf-7-M-4* *réf-9-M-1*

les plus répandues] sont **[[l'oyat] et [le chiendent]]**. **[Les plus vieilles dunes]** ont **[plus de**
réf-10 *réf-11* *réf-9-M-2* *réf-12* *réf-13*

cinq mille ans]. **[La criste marine]** est **[une plante comestible]**. **[Madame l'inspectrice]**
réf-14-M-1 *réf-14-M-2* *réf-15*

accompagnait **[M. Berger]**. **[Nous]** avons continué **[la visite]** avec **[eux]**. Puis **[nous]** sommes
réf-16 *réf-3-M-2* *réf-17* *réf-18* *réf-3-M-3*

rentrés à **[l'école]**.
réf-4-M-2

(Turco, cité par Reichler-Béguelin *et al.*, 1988 : 132, cité par Schnedecker, 1997 : 5, nous annotons)

Il est difficile de saisir une ou plusieurs entités qui ressortent de l'ensemble de ce texte, puisque la plupart d'entre elles sont simplement introduites les unes après les autres dans chaque phrase, sans continuité ni progression. Au total, 18 référents sont présents dans cette

rédaction ; la plupart sont éphémères, avec une seule apparition et sans aucune expression coréférentielle :

- 12 référents ne sont mentionnés qu'une seule fois : « sortie » (« réf-1 »), « bicyclette » (« réf-5 »), « la côte » (« réf-6 »), « la langue » (« réf-8 »), « l'oyat » (« réf-10 »), « le chiendent » (« réf-11 »), « les plus vieilles dunes » (« réf-12 »), « plus de cinq mille ans » (« réf-13 »), « Madame l'inspectrice » (« réf-15 »), « M. Berger » (« réf-16 »), « la visite » (« réf-17 ») et « eux » (« réf-18 ») ;
- 4 référents sont mentionnés deux fois : « la dune » (« réf-2 »), « l'école » (« réf-4 »), « les plantes les plus répandues » (« réf-10 ») et « la criste marine » (« réf-14 ») ;
- 2 référents sont mentionnés trois ou plus de trois fois : « nous » (« réf-3 », 3 fois) et « l'euphorbe » (« réf-7 », 4 fois).

1.6.2. Définition selon trois critères

Dans la définition de Schnedecker, trois critères sont fondamentaux pour caractériser une chaîne : ***numérique***, ***spatial*** et ***lexical***.

Le critère numérique indique le nombre minimal de composants d'une chaîne de référence : ce nombre doit être supérieur ou égal à trois. Schnedecker reprend la position de Corblin sur la quantité minimale de trois expressions, pour ne pas confondre la notion de chaîne avec celles de ***coréférence*** et d'***anaphore***, qui se basent sur la relation entre deux expressions.

> *Une chaîne comporte* ***au moins trois maillons****. (Schnedecker, 1997 : 24, nous soulignons)*

Le critère spatial est relatif à la distance inter-maillonnaire et la mesure de la longueur des chaînes ; il est le plus difficile à appliquer à la définition de chaîne de référence en raison des unités de mesures insuffisantes pour décrire les spécificités des chaînes.

Le critère lexical concerne les nombreuses catégories de maillons d'une chaîne. Les composants d'une chaîne sont censés être variables, avec diverses catégories nominales et non nominales : les SN peuvent mettre en relief les propriétés des référents ; les noms propres se chargent de la fonction de dénomination à leur première occurrence et de redénomination dans la chaîne à leur reprise. De nombreux auteurs confirment que les formes non nominales sont censées être les composants majoritaires des chaînes, quel que soit le genre de discours ou le type de texte. Une composition hypothétique est ainsi proposée par Schnedecker : une chaîne de référence est composée de noms propres, de SN et de pronoms coréférentiels :

> *On définira les* chaînes de référence *comme la suite des expressions référentielles (homogènes) encadrée par deux noms propres ou deux SN pleins identiques coréférents. (Schnedecker, 1997 : 190)*

Cette définition par le critère lexical nous paraît plus précise et plus stricte que les précédentes pour deux raisons. Premièrement, elle permet de clarifier les catégories de maillons qui marquent le début et la fin des chaînes. Pour constituer une chaîne, deux noms propres ou SN pleins doivent se coréférer ; ils délimitent ensemble les frontières d'une chaîne de référence. Deuxièmement, les catégories de maillons qui maintiennent une chaîne sont précisées : elles sont le plus souvent pronominales.

Pour illustrer sa définition selon ces trois critères, Schnedecker propose un exemple de chaîne de référence qu'on peut lire dans l'extrait [16] : le nom propre « Bony » démarre la chaîne et il est suivi d'une séquence de cinq pronoms « il » qui coréfèrent à la même personne.

[16]Le lendemain, [**Bony**]$_{M-1}$ resta au Trou du Pied-Bot. [**Il**]$_{M-2}$ fit la lessive et [**Ø**]$_{M-3}$ prit la peine de cuisiner quelque chose qui ressemblait à du ragoût. [**Il**]$_{M-4}$ remonta le sentier jusqu'au sommet de la falaise, le fusil sous le bras, espérant voir un kangourou car [**il**]$_{M-5}$ avait remarqué des empreintes récentes dans le trou. [**Il**]$_{M-6}$ passa deux heures à examiner les anciens journaux et documents qu'[**il**]$_{M-7}$ avait pris dans la vieille valise, sous le lit de Lonergan.

(Upfield, *L'homme des tribus*, cité par Schnedecker, 1997 : 10, nous annotons)

Cette définition est développée dans deux directions. D'une part, Schnedecker (1997 : 201) signale le rôle crucial des noms propres dans les chaînes : ce sont les séquences coréférentielles qui permettent la continuité référentielle dans le texte. Une succession de quatre chaînes de référence est configurée comme suit : chaque chaîne (« C1 », « C2 », « C3 » et « C4 ») est initiée par un nom propre (« Np1 », « Np2 », « Np3 » et « Np4 »), suivi de quelques pronoms (« Pro ») :

> *Dans notre perspective, la permanence référentielle se fonde sur une succession de chaînes coréférentes, suivant un schéma comme :*
>
> *C1.....C2 C3..... C4*
>
> *{Np1... Pro... Pro... / /{Np3... Pro... Pro...*
>
> *{Np2... Pro... Pro... {Np4 ... Pro ... Pro ...*
>
> *Dans ce dispositif, le Np est la cheville ouvrière. Sa récurrence fonde le système référentiel et lui permet de tourner. Ce sont en effet les différents Np qui servent de nœuds et corrèlent les unes aux autres les chaînes.*
>
> *(Schnedecker, 1997 : 201)*

D'autre part, si cette configuration permet de définir plus spécifiquement et plus précisément la notion de ***chaîne de référence***, la linguiste propose également, dans une étude plus récente, une définition plus générale des chaînes de référence :

> *Suite des expressions d'un discours renvoyant à une même entité. (Schnedecker, 2021 : 261)*

Dans la logique de cette définition, Schnedecker (2021) présente une chaîne de référence dans un contexte narratif, visible ci-dessous dans l'extrait [17] :

[17]Dans la plaine rase, sous la nuit sans étoiles, d'une obscurité et d'une épaisseur d'encre, [**un homme**]$_{M-1}$ suivait seul la grande route de Marchiennes à Montsou, dix kilomètres de pavé coupant tout droit, à travers les champs de betteraves. Devant [**lui**]$_{M-2}$, [**il**]$_{M-3}$ ne voyait même pas le sol noir, et il n'avait la sensation de l'immense horizon plat que par les souffles du vent de mars, des rafales larges comme sur une mer, glacées d'avoir balayé des lieus de marais et de terres nues. Aucune ombre d'arbre ne tachait le ciel, le pavé se déroulait avec la rectitude d'une jetée, au milieu de l'embrun aveuglant des ténèbres.
[**L'homme**]$_{M-4}$ était parti de Marchiennes vers deux heures. [**Il**]$_{M-5}$ marchait d'un pas allongé, grelottant sous le coton aminci de [**sa**]$_{M-6}$ veste et de [**son**]$_{M-7}$ pantalon de velours. Un petit paquet, noué dans un mouchoir à carreaux, [**le**]$_{M-8}$ gênait beaucoup ; et [**il**]$_{M-9}$ le serrait contre [**ses**]$_{M-10}$ flancs, tantôt d'un coude,

tantôt de l'autre, pour glisser au fond de [**ses**]$_{M-11}$ poches les deux mains à la fois, des mains gourdes que les lanières du vent d'est faisaient saigner. Une seule idée occupait [**sa**]$_{M-12}$ tête vide d'ouvrier sans travail et sans gîte, l'espoir que le froid serait moins vif après le lever du jour. Depuis une heure, [**il**]$_{M-13}$ avançait ainsi, lorsque sur la gauche à deux kilomètres de Montsou, [**il**]$_{M-14}$ aperçut des feux rouges, trois brasiers brûlant au plein air, et comme suspendus. D'abord, [**il**]$_{M-15}$ hésita, pris de crainte ; puis, il ne put résister au besoin douloureux de se chauffer un instant les mains.

(Zola, *Germinal, incipit*, cité par Schnedecker, 2021 : 7, nous annotons)

Cette chaîne de 15 maillons comporte trois catégories d'expressions : les SN (« un homme », « l'homme »), les pronoms (« lui », « il ») et les déterminants possessifs (« sa », « son »). Elle reflète quelques caractéristiques de la configuration de Schnedecker (1997) : elle est initiée par un SN indéfini, « un homme », correspondant à « SN1 » ; ce référent est repris par un SN défini « l'homme », correspondant à « SN1' » ; chaque expression nominale est suivie d'une séquence constituée de pronoms et de déterminants possessifs.

1.6.3. Configuration selon la saillance du référent

Les composants des chaînes de référence étant variables et une seule configuration typique ne paraissant pas suffisante pour décrire les chaînes dans différents contextes, quelques précisions sur diverses configurations de chaînes selon la ***saillance*** du référent s'imposent. Ainsi, il convient au préalable d'examiner cette notion de ***saillance*** (en anglais ***prominence***, *cf*. Rohde & Kehler, 2014 : 913) référentielle, qui est une notion créée dans le but de décrire « le statut de centralité de certains référents dans la conscience de l'énonciateur » (Neveu, 2000 : 100, cité par Schnedecker, 2011 : 31)[12]. Landragin (2011) précise cette notion :

> *Un élément linguistiquement saillant, c'est un élément qui ressort prioritairement lors de la compréhension d'un énoncé, du fait d'une importance cognitive particulière, et au point, par exemple, d'axer sur cet élément la réaction ou la réponse à l'énoncé. (Landragin, 2011 : 92)*

Autrement dit, la saillance est un indice qui reflète l'importance du référent dans l'esprit du locuteur.

Selon cette distinction des référents, les chaînes sont également subdivisées par Schnedecker (1997) en celles dont les référents sont ***saillants/continus*** et celles dont ils sont ***non saillants***. Une chaîne d'un référent saillant peut être configurée comme :

> ***Np1****… il1… il1… Np2 … il1… il1…* ***Np1'****…*
>
> *(Schnedecker, 1997 : 72)*

[12] D'autres auteurs définissent la saillance différemment. Par exemple, Chafe (1994, cité par Schnedecker, 2011 : 34) fait correspondre la saillance à l'identification du référent dans le contexte : « An identifiable referent is one that (a) assumed to be already shared, directly or indirectly by the listener, (b) verbalized in a sufficiently identifying way; and (c) contextually salient ». La définition de la notion d'*accessibilité*, très proche de celle de *saillance*, est résumée par Xu (许余龙, 2010 : 357) comme suit : « Accessibility: a psycholinguistic notion referring to the degree of relative ease or difficulty with which an entity can be retrieved from one's mind ».

Au contraire, une chaîne d'un référent non saillant peut se présenter comme :

> ***Np1**… il1… il1… il1… il1… **Np1'** … il1… il1… **Np1"**…*
>
> *(Schnedecker, 1997 : 72)*

Par la comparaison de ces deux configurations, Schnedecker met en relief le rôle des noms propres dans les chaînes de référence, en corrélation avec la saillance du référent : plus un référent est saillant, moins il est mentionné par les noms propres :

> *Théoriquement, plus un référent serait saillant, moins il serait redénommé ; moins un référent serait saillant, plus il aurait besoin d'être réinstancié par le Np. Bref,* la fréquence de la redénomination serait l'indicateur d'une moindre saillance. *(Schnedecker, 1997 : 73)*

Par ailleurs, les configurations susmentionnées représentent plutôt les chaînes des référents d'humains dans la narration ; celle des chaînes des référents saillants non humains dans les genres non narratifs doit donc être représentée autrement : celles-ci contiennent souvent, comme le notent Schnedecker & Landragin (2014), seulement des expressions nominales très distanciées entre elles[13] :

> *§1 SN1… §2 SN2… §N SNn…*
>
> *(Schnedecker & Landragin, 2014 : 13)*

1.7. Chaîne de coréférence des linguistes informaticiens

Dans le domaine du traitement automatique des langues, les linguistes informaticiens, notamment Mitkov *et al*. (2000), Victorri (2005), Boudreau & Kittredge (2006), Landragin (2011, 2014a, 2014b 2016b, 2017) et Grobol (2021), font référence à la *chaîne de coréférence*. Ce terme a pour but d'indiquer que c'est la coréférence qui constitue la relation principale des composants :

> *Les expressions qui réfèrent à un même élément sont dites coréférentielles, et elles forment dans un texte ce que l'on appelle une **chaîne de coréférence**. (Victorri, 2005 : 113, nous soulignons)*

Dans la définition de Victorri, une chaîne de coréférence comprend au moins deux expressions renvoyant à des individus, des objets, des événements, etc. Par exemple, l'extrait [18] contient trois chaînes qui renvoient à trois véhicules différents :

[18]J'étais ([**véhicule A**]) à l'arrêt au feu rouge derrière [**une 4L de la gendarmerie**]. Lorsque le feu
chaîne-1-M-1 *chaîne-2-M-1*
est passé au vert, [**la 4L**] a calé. J'attendais qu'[**elle**] redémarre lorsque j'ai été heurtée à l'arrière
chaîne-2-M-2 *chaîne-2-M-3*
par [**le véhicule B**]. Conduisant [**un véhicule aménagé pour la conduite manuelle**], j'ai sans
chaîne-3-M-1 *chaîne-1-M-2*
doute accentué la pression sur le frein principal que je tenais dans la main droite ce qui a causé une douleur du poignet mais qui m'a évité de heurter [**l'autre véhicule**]. Bien entendu ma tête a fait un
chaîne-3-M-2
hochement involontaire.

(Victorri, 2005 : 113, nous annotons)

[13] « § » : paragraphe.

1.8. Synthèse des définitions

Après avoir examiné l'histoire de la notion de ***chaîne de référence*** chez divers auteurs, il convient de proposer à présent une synthèse des définitions de cette notion qui se trouve au cœur de notre thèse. Malgré certaines nuances, les linguistes s'accordent généralement à dire qu'une ***chaîne de référence*** est une suite d'expressions référentielles renvoyant à une même entité dans un discours.

Les notions relatives à celle de chaîne de référence partagent des points communs et comportent des différences dans la définition de chaque auteur. Nous synthétisons ces points dans le tableau (1.1). Ces différences sont synthétisées selon quatre aspects : i) domaines et approches de départ ; ii) notions fondamentales pour décrire la relation entre maillons et celle entre maillons et entités du monde extra-linguistique[14] ; iii) nombre de maillons et iv) catégories grammaticales des maillons.

[14] Cet aspect nous paraît essentiel et a donné lieu à de nombreuses discussions, que nous présenterons dans le *Chapitre 2*.

Tableau (1.1) : Synthèse des définitions de *chaîne*

Auteur	Terminologie	Domaine ou approche	Notions de base	Nbre d'expressions	Composants	Langues étudiées
Chastain (1975)	**Chaîne anaphorique**	Sémantique	Anaphore, référence	≥ 2	SN, pronom	Anglais
Li & Tompson (1979)	**Chaîne topicale**	Pragmatique, sémantique	Topique, référence	≥ 3	SN, zéro	Chinois
Givon (1983) et Xu (徐赳赳, 2003, 2014)	**Chaîne topicale**	Fonctionnel, discursif	Topique, Anaphore	≥ 2	SN, pronom, zéro, accord grammatical	Anglais, chinois
Corblin (1985)	**Chaîne de référence**	Sémantique, textuel	Coréférence	≥ 3	SN, pronom, déterminant possessif	Français
Charolles (1988)	**Chaîne référentielle**	Textuel	Coréférence	≥ 3	SN, pronom	Français
Xu (徐赳赳, 1990)	**Chaîne pronominale**	Sémantique, textuel	Coréférence	≥ 3	Nom propre, SN, pronom	Chinois
Schnedecker (1997)	**Chaîne de référence**	Sémantique, textuel	Coréférence	≥ 3	Nom propre, SN, pronom, zéro	Français
Victorri (2005) et Landragin (2011)	**Chaîne de coréférence**	Informatique	Coréférence	≥ 2	SN, pronom	Français, anglais

1.8.1. Domaines et approches de départ

Il existe une distinction évidente entre le terme ***chaîne de coréférence*** et les autres : le premier est créé et utilisé plutôt pour l'informatique, et les autres pour les domaines traditionnels de la linguistique à proprement parler. L'objet du présent travail ne porte pas sur la résolution des questions dans le domaine du traitement automatique des expressions coréférentielles. Nous nous positionnons, relativement à la question des chaînes, sur les approches sémantiques, discursives et contrastives, pour observer et interpréter les phénomènes en français et en chinois. Par conséquent, entre ***chaîne de référence*** et ***chaîne de coréférence***, nous privilégions la première dénomination pour discuter linguistiquement cette notion.

Par ailleurs, la distinction entre ***chaîne topicale*** et les autres termes est également facile à établir : la première se base sur la notion discursivo-fonctionnaliste de ***topique***, alors que les autres notions de ***chaîne*** ne mettent pas en exergue la différence entre le topique du discours et les autres référents. La notion de ***topique*** nous paraît importante parce qu'elle permet de distinguer, entre les référents, ceux qui sont ***topiques*** et ceux qui ne le sont pas. Les ***chaînes topicales*** sont donc, d'après nous, un sous-type de chaîne de référence qui ne concerne que les chaînes des topiques du discours.

En bref, les structures de la grammaire, du discours et de l'information sont étroitement reliées les unes aux autres, en constituant ensemble les phrases et le texte. Le référent en position de sujet syntaxique peut correspondre au topique phrastique et au focus de l'attention en portant des informations anciennes. Le topique et le focus sont en même temps déplacés aux éléments d'autres fonctions syntaxiques dès qu'il y a changement de structure phrastique, en raison duquel les informations anciennes et les nouvelles ne sont plus les mêmes.

1.8.2. Nombres de maillons

Selon les différentes définitions que nous avons examinées, une chaîne peut contenir au minimum deux expressions ou trois maillons. Certains auteurs considèrent que deux expressions peuvent constituer une chaîne, comme la chaîne anaphorique de Chastain (1975), dans laquelle la première est l'antécédent et la seconde l'anaphorique ; ou comme la chaîne topicale de Givón (1983) et de Xu (徐赳赳, 2003, 2014), la première expression servant de source référentielle au topique et la seconde renvoyant au même topique ; ou encore comme la chaîne de coréférence, selon les linguistes informaticiens (Victorri, 2005 ; Landragin, 2011), dont la relation basique est la ***coréférence*** entre deux expressions. D'autres auteurs considèrent que deux maillons ne sont pas suffisants pour constituer une chaîne. D'après eux, cette dernière doit comporter trois maillons ou plus. Corblin (1985) aborde pour la première fois la question du nombre d'expressions dans la définition : deux expressions coréférentielles ne suffisent pas à former une ***chaîne,*** qui doit comporter au moins trois expressions. Cette idée est reprise dans la définition de chaîne de Schnedecker (1997), qui insiste sur le fait qu'une chaîne doit, pour se distinguer de la notion de ***coréférence*** et d'***anaphore***, contenir au minimum trois maillons. Charolles (1988) ne précise pas la question du nombre de maillons

dans sa définition de chaîne ; pourtant, d'après sa configuration dans le cadre textuel, et d'après ses exemples, une chaîne est susceptible de contenir au moins trois expressions.

Au niveau quantitatif, nous ne considérons pas les expressions coréférentielles en paire comme des chaînes, afin de ne pas confondre cette notion de ***chaîne*** avec celles d'***anaphore*** et de ***coréférence***. Nous adoptons ainsi la même position que Corblin (1985) et que Schnedecker (1997) sur la quantité minimale de trois maillons dans la définition de ***chaîne de référence***.

1.8.3. Formes des composants

Il n'y a pas de consensus entre les linguistes au sujet des composants des chaînes. Deux catégories semblent être relativement principales, à savoir les SN et les pronoms, mais, manifestement, les autres formes s'avèrent aussi pertinentes et acceptables, surtout les noms propres, les zéros et les déterminant possessifs, selon divers auteurs.

Les positions typiques des SN et des pronoms sont configurées par Xu (1990) ; celles des noms propres et des pronoms le sont par Schnedecker (1997). Les SN et les noms propres servent de source référentielle aux pronoms qui suivent ; les deux catégories constituent des séquences coréférentielles.

Les déterminants possessifs, les zéro et les accords grammaticaux sont également évoqués dans certaines définitions et sont illustrés par des exemples. Corblin (1985) annote, dans son exemple, les déterminants possessifs en tant que maillons. Cette catégorie est bien acceptée pour les chaînes de référence en chinois, comme le montrera ultérieurement l'extrait [20], dans lequel le déterminant possessif 'son' ***tade*** 她的 (« CR-3-M-3 ») est sans aucun doute un maillon de la chaîne 3 'mère' ***muqin*** 母亲. Xu (1990, 2003, 2014) évoque la question des zéros en tant que composants importants dans les chaînes topicales et pronominales pour le chinois. Nous adhérons à ce point de vue selon lequel les zéros peuvent être des composants potentiels pour les chaînes de référence à la fois en chinois et en français. Les accords grammaticaux sont considérés comme composants seulement dans les chaînes topicales de Givón (1983)[15], sans être illustrés par des exemples. Cette remarque est compréhensible parce qu'elle se base sur des analyses de l'anglais, langue analytique[16] marquée par de nombreux accords grammaticaux ; elle pourrait être pertinente pour le français, qui est également une langue analytique. Au contraire, le chinois est une langue synthétique, dépourvue d'accords grammaticaux systématiques. Par conséquent, nous considérons que les accords grammaticaux ne serviront pas à notre analyse contrastive des chaînes de référence en chinois et en français.

---

[15] Ariel (1990) évoque également l'importance de la référence des accord grammaticaux dans son *Echelle d'accessibilité référentielle* (voir la section *5.2.2.5. Forme : déterminant, tête lexicale et modifieur* du *Chapitre 5* pour la présentation de sa théorie).

[16] Définition du *Dictionnaire des sciences du langage* (Dubois *et al.*, 2012 : 472) : « On qualifie de *synthétiques* les langues flexionnelles comme le latin et les langues agglutinantes comme le vietnamien. *Synthétique* s'oppose ici à *analytique*. Est synthétique une langue qui tend à réunir en un seul mot plusieurs morphèmes. Le français est une langue analytique parce qu'il exprime les fonctions par des mots autonomes appelés *prépositions* et que dans une phrase chacune des unités reste relativement indépendante des autres ».

Pour résumer, nous adoptons une position accueillante vis-vis des formes de maillons en nous adaptant, d'un côté, aux définitions respectives de chaque auteur, et, de l'autre, aux caractéristiques typologiques propres au français et au chinois. Par conséquent, si une forme existe dans les deux langues, nous la considérons comme maillon potentiel de chaîne, comme les SN, les pronoms, les déterminants possessifs et les zéros. Les accords grammaticaux ne sont donc pas considérés comme maillons ou marqueurs référentiels dans l'annotation de notre corpus, que nous présenterons plus loin.

1.8.4. Langues de base des études

Sachant qu'un des objectifs de notre thèse est de construire une définition qui s'adapte au français et au chinois, il nous paraît important de voir sur quelles langues de base ces notions sont établies. Malgré les nombreuses définitions que nous avons passées en revue et les nombreuses études relativement récentes relatives au projet DEMOCRAT sur la comparaison des chaînes de référence dans diverses langues, comme l'anglais, l'allemand, l'ancien français, le tchèque, le hongrois et le coréen[17], aucun auteur n'aborde la question de la définition de cette notion par l'approche contrastive franco-chinoise. Ces termes sont créés, soit sur la base d'une seule langue, comme la chaîne anaphorique en anglais, la chaîne de référence en français et la chaîne pronominale en chinois, soit sur celle de deux langues, comme la chaîne topicale en anglais et en chinois, la chaîne de coréférence en anglais et en français. Par conséquent, nous cherchons à établir notre propre définition, qui s'adapte à la fois au français et au chinois, au fondement de notre corpus constitué de textes en ces deux langues.

1.9. Notre position relativement aux aspects textuels franco-chinois

Les définitions de la notion de chaîne de référence partagent des points communs universels ; elles présentent aussi des points divergents. Au fondement des études mentionnées dans ce chapitre, nous proposons une définition de chaîne de référence pour le chinois avec notre traduction en français :

> *指称链条指的是在同一语境中的三个及三个以上的共指表达所构成的链条，主要包括但不限于名词短语，代词和零形式。共指对象可以是话语中提及的任何对象，包括但不限于人，物，事件，时空概念，抽象概念等。**(nous définissons en chinois)***
>
> ***Une chaîne de référence est une séquence de trois ou de plus de trois expressions coréférentielles dans un même contexte. Les composants peuvent être principalement mais pas uniquement les SN, les pronoms et les zéro. Le référent peut être un objet de toute sorte, mentionné dans le discours, ce qui inclut les personnes, les objets, les événements, les notions spatio-temporelles, les notions abstraites, etc. (nous traduisons en français)***

[17] Voir les articles dans deux revues : *Langue française* (2017), numéro 195 ; *Travaux de Linguistique : Approches contrastives des chaînes de référence* (2021), numéro 82, et la thèse de Kim (2023).

Selon notre définition, trois chaînes de référence sont annotées dans les extraits parallèles [19] et [20] issus du roman-fleuve ***Jean-Christophe*** en français et en chinois :

- la chaîne 1 de « le nouveau-né » : 6 maillons en français (« le nouveau-né », « son », « l'enfant », « le », « le petit », « l'enfant ») et 5 maillons en chinois ('nouveau-né bébé' ***chushengde-yinger*** 初生的婴儿, 'enfant' ***haizi*** 孩子, '3P' ***ta*** 他, 'enfant' ***haizi*** 孩子, 'enfant' ***haizi*** 孩子) ;
- la chaîne 2 de « le vieux » : 14 maillons en français (« le vieux », « ses », « son », « le grand-père », « du vieux Jean-Michel », « sa », « son », « ses », « il », « son », « il », « ses », « lui », « s' ») et 5 maillons en chinois ('vieil homme' ***laoren*** 老人, 'grand-père' ***zufu*** 祖父, 'vieux Jean-Michel' ***lao-Yuehan-Mixier*** 老约翰·米希尔, '3P' ***ta*** 他, '3P' ***ta*** 他) ;
- la chaîne 3 de « la mère » : 9 maillons en français (« la mère », « son », « Louisa », « elle », « un blond presque blanc », « ses », « sa », « elle », « elle ») et 4 maillons en chinois ('mère' ***muqin*** 母亲, 'Louisa' ***Luyisha*** 鲁意莎, 'son' ***tade*** 她的, '3P' ***ta*** 她).

[19] **[Le nouveau-né]** *CR-1-M-1* s'agite dans **[son]** *CR-1-M-2* berceau. Bien que **[le vieux]** *CR-2-M-1* ait laissé, pour entrer, **[ses]** *CR-2-M-2* sabots à la porte, **[son]** *CR-2-M-3* pas a fait craquer le plancher : **[l'enfant]** *CR-1-M-3* commence à geindre. **[La mère]** *CR-3-M-1* se penche hors de **[son]** *CR-3-M-2* lit, afin de **[le]** *CR-1-M-4* rassurer ; et **[le grand-père]** *CR-2-M-4* allume la lampe en tâtonnant, pour que **[le petit]** *CR-1-M-5* n'ait pas peur de la nuit. La flamme éclaire la figure rouge **[du vieux Jean-Michel]** *CR-2-M-5*, **[sa]** *CR-2-M-6* barbe blanche et rude, **[son]** *CR-2-M-7* air bourru et **[ses]** *CR-2-M-8* yeux vifs. **[Il]** *CR-2-M-9* vient près du berceau. **[Son]** *CR-2-M-10* manteau sent le mouillé ; **[il]** *CR-2-M-11* traîne en marchant **[ses]** *CR-2-M-12* gros chaussons bleus.

[Louisa] *CR-3-M-3* **[lui]** *CR-2-M-13* fait signe de ne pas **[s']**approcher. **[Elle]** *CR-2-M-14* est d'**[un blond presque blanc]** *CR-3-M-4 CR-3-M-5* ; **[ses]** *CR-3-M-6* traits sont tirés ; **[sa]** *CR-3-M-7* douce figure mouton est marquée de taches de rousseur ; **[elle]** *CR-3-M-8* a des lèvres pâles et grosses, qui ne parviennent pas à se rejoindre et qui sourient avec timidité ; **[elle]** *CR-3-M-9* couve **[l'enfant]** *CR-1-M-6* des yeux – des yeux très bleus, très vagues, où la prunelle est un point tout petit, mais infiniment tendre.

[20]

[初生的	**婴儿]**	在	摇篮	里	扭动。	**[老人]**	进来	虽然	把	木靴	脱	在	
chushengde	*yinger*	*zai*	*yaolan*	*li*	*niudong*	*laoren*	*jinlai*	*suiran*	*ba*	*muque*	*tuo*	*zai*	
nouveau-né *CR-1-M-1*	bébé	PREP	berceau	intérieur	s'agiter	vieux-homme *CR-2-M-1*	entrer	malgré-tout	BA	sabot	laisser	PREP	

门	外，	走路	的	时候	地板	还是	格格的响：	**[孩子]**	哼啊嗐的	哭	了。	**[母亲]**	
men	*wai*	*zoulu*	*de*	*shihou*	*diban*	*haisi*	*gegedexiang*	*haizi*	*hengahaide*	*ku*	*le*	*muqin*	
porte	extérieur	marcher	DE	temps	plancher	encore	craquer	enfant *CR-1-M-2*	ONOMA	pleurer	LE	mère *CR-3-M-1*	

从	床	上	探出	身子	抚慰	**[他]**；	**[祖父]**	摸索着	点起	灯	来，免得	**[孩子]**	
cong	*chuang*	*shang*	*tanchu*	*shenzi*	*fuwei*	*ta*	*zufu*	*mosuozhe*	*dianzi*	*deng*	*lai*	*miande*	*haizi*
par	lit	dessus	sortir	corps	assurer	3P *CR-1-M-3*	grand-père *CR-2-M-2*	tâtonnant	allumer	lampe	AUX	évitant	enfant *CR-1-M-4*

在	黑夜	里	害怕。	灯光	照出	**[老**	**约翰·米希尔]**	红红的	脸，	粗硬的	白须，		
zai	*heiye*	*li*	*haipa*	*dengguang*	*zhaochu*	*lao*	*Yuehan Mixier*	*honghongde*	*lian*	*cuyingde*	*baixu*		
PREP	nuit	pendant	avoir-peur	lampe	allumer	vieux *CR-2-M-3*	Jean-Michel	rouge	visage	rude	barbe		

忧郁易怒的	表情，	炯炯有神的	眼睛。	**[他]**	走近	摇篮，	外套	发出	股	潮气，			
youyuyinude	*biaoqing*	*jiongjiongyoushende*	*yanjing*	*ta*	*zoujin*	*yaolan*	*waitao*	*fachu*	*gu*	*chaoqi*			
bourru	figure	vif	yeux	3P *CR-2-M-4*	s'approcher	berceau	manteau	sentir	CL	mouillé			

脚 下 拖着 双 大 蓝 布鞋。**[鲁意莎]** 做着 手势 叫 **[他]** 不要 走近。
jiao xia tuozhe shuang da lan buxie luyisha zuozhe shoushi jiao ta buyao zoujin
pied dessous traînant paire gros bleu chausson Louisa faisant signe demander 3P NEG s'approcher
CR-3-M-2 *CR-2-M-5*

[她的] 淡黄 头发 差不多 象 白的； 绵羊 般 和善的 脸 都 打皱了，
tade danhuang toufa chabuduo xiang baide mianyang ban heshande lian dou dazhoule
son jaune cheveux presque comme blanc mouton comme doux visage presque tiré
CR-3-M-3

颇有 些 雀斑； 没有 血色 的 厚 嘴唇 不大 容易 合拢， 笑起来 非常 胆怯；
poyou xie queban meiyou xuese de hou zuichun buda rongyi helong xiaoqilai feichang danqie
avoir des tâche-rousseur NEG couleur DE gros lèvre NEG facile se-rejoindre riant très timide

眼睛 很 蓝， 迷迷惘惘的， 眼珠 只有 极小的 一点， 可是 挺 温柔； — **[她]**
yanjing hen lan mimiwangwangde yanzhu zhiyou jixiaode yidian keshi ting wenrou ta
yeux très bleu vague prunelle seulement petit point mais très doux 3P
CR-3-M-4

不胜怜爱的 瞅着 **[孩子]**。
bushenglianaide chouzhe haizi
infiniment-tendre regardant enfant
CR-1-M-5

Nous utilisons, dans le présent travail, le terme de ***chaîne de référence*** selon notre définition pour désigner ces suites d'expressions coréférentielles dans un même contexte discursif.

1.10. Questions et discussions à la base des comparaisons des textes

À partir de l'examen des définitions et de leur comparaison, nous formulons deux questions : sachant que les chaînes de référence présentent leurs apports spécifiques à la structure textuelle, comment permettent-elles la cohésion textuelle, surtout par rapport à un texte décousu ? Comment les chaînes de référence se réalisent-elles dans les textes cohésifs en français et en chinois ?

1.10.1. Comparaison entre une copie d'élève et un extrait de roman-fleuve en français

Une copie d'élève est citée par Schnedecker (1997 : 5) dans le fragment [15], qui nous permet d'obtenir et d'explorer les questions relatives à l'influence du nombre d'expressions référentielles et de chaînes de référence sur la cohérence textuelle. En effet, la longueur d'un texte peut être un indice important qui influence certains paramètres référentiels. Il est plausible que les textes longs contiennent plus d'expressions référentielles et plus de chaînes de référence que les textes courts ; mais la comparaison entre les nombres absolus et entre les pourcentages des deux langues pourrait révéler des différences. Nous avons donc comparé les paramètres référentiels entre le fragment [15], texte non cohésif, et le fragment [19], texte cohésif en français issu du début du roman-fleuve *Jean-Christophe*, pour tenter de dégager des tendances. Le fragment [15] ***Sortie à la dune*** comporte 76 mots ; nous avons obtenu 27 expressions référentielles renvoyant à 18 référents, dont seulement deux forment des chaînes de référence ; les 16 autres référents sont soit des singletons, soit des paires ; la chaîne la plus

longue contient 4 maillons (« L'euphorbe », « une plante toxique », « elle », et un zéro). Le fragment [19] de ***Jean-Christophe*** contient au total 176 mots. Nous avons obtenu en tout 53 expressions référentielles qui renvoient à 25 référents, dont 22 sont évoqués seulement par une ou deux expressions, et trois sont exprimés par les chaînes :

- 22 référents en singleton : « son berceau », « ses sabots », « la porte », « son pas », « le plancher », « son lit », « la lampe », « la nuit », « la flamme », « la figure rouge du vieux Jean-Michel », « sa barbe blanche et rude », « son air bourru », « ses yeux vifs », « son manteau », « le mouillé », « ses gros chaussons bleus », « ses traits », « sa douce figure mouton », « taches de rousseur », « des lèvres pâles et grosses … » « des yeux … » et « la prunelle » ;
- 3 référents en chaîne : « le nouveau-né » (6 maillons), « le vieux » (14 maillons) et « la mère » (9 maillons).

La chaîne la plus longue de ce texte a 14 maillons et renvoie à la personne « Jean-Michel ».

À première vue, les deux fragments n'ont pas la même longueur : il se peut que les chiffres obtenus ne soient pas comparables dans deux textes de longueur largement différente. Nous avons ainsi calculé le pourcentage de chaque paramètre ; les chiffres sont présentés dans le tableau (1.2) suivant.

Tableau (1.2) : Comparaison entre une copie d'élève et un extrait d'un roman-fleuve en français

Texte	Copie d'élève	Extrait de roman-fleuve
	Fragment [15] (***Sortie à la dune***)	Fragment [19] (***Jean-Christophe*** en français)
Mots	76	176
ER	27 (36 %)	53 (30 %)
Réf	18 (24 %)	25 (14 %)
CR	2 (11 %)	3 (12 %)
Maillons de la chaîne la plus longue	4 (5 %)	14 (8 %)

La proportion d'expressions référentielles est plus élevée dans le texte non cohésif (36 %) que dans le cohésif (30 %) ; c'est également le cas pour celle des référents (24 %, contre 14 %). Par contre, les proportions concernant les chaînes montrent des tendances inverses : celle des référents en chaîne est légèrement plus élevée dans le texte cohésif (12 %) que dans le non cohésif (11 %) ; celle des maillons dans la chaîne la plus longue est plus élevée dans le texte cohésif (8 %) que dans le non cohésif (5 %). Cette comparaison nous permet d'affirmer, d'une part, que la proportion des expressions référentielles et des référents a tendance à baisser dans les textes cohésifs, et, d'autre part, que la proportion des chaînes affiche une tendance contraire : elle s'élève dans ces mêmes textes.

Comparer deux textes uniquement conduit à des résultats et à des tendances qui peuvent être aléatoires et non représentatifs. Il convient donc de poursuivre nos observations au sein d'un corpus élargi, comportant un grand nombre de textes, et de nous demander si les proportions constatées montrent ainsi des tendances plus ou moins stables dans les textes cohésifs. Les textes ne seraient-ils pas plus cohérents dès que le nombre des chaînes de référence dépasse un certain seuil, malgré l'emploi de tous les autres moyens qui garantissent en même temps la cohérence textuelle ?

1.10.2. Comparaison entre deux extraits de roman-fleuve en français et en chinois

Étant donné le fait que ce sont des paramètres comme le nombre de référents, d'expressions référentielles et de chaînes, qui renforcent la continuité textuelle en plus de tous les autres moyens cohésifs, ceux-ci présentent-ils des différences en français et en chinois ? Notre seconde hypothèse est donc relative aux indices référentiels dans les deux langues : les textes cohésifs en français et en chinois comportent-ils les mêmes proportions d'expressions référentielles, de référents et de chaînes ? Pour le savoir, nous avons comparé les chiffres absolus et les proportions dans les textes cohésifs parallèles [19] et [20] issus du roman-fleuve *Jean-Christophe* en français et en chinois. Le fragment [20] contient 251 caractères chinois. Nous avons obtenu au total 44 expressions référentielles qui renvoient à 32 référents, dont 29 sont évoqués une ou deux fois ; 3 référents constituent des chaînes.

- 29 référents en singleton : 'berceau' ***yaolan*** 摇篮, 'berceau-intérieur' ***yaolan-li*** 摇篮里, 'sabot' ***muxue*** 木靴, 'porte' ***men*** 门, 'porte-extérieur' ***men-wai*** 门外, 'plancher' ***diban*** 地板, 'lit' ***chuang*** 床, 'lit-dessus' ***chuang-shang*** 床上, 'corps' ***shenzi*** 身子, 'lampe' ***deng*** 灯, 'nuit' ***heiye*** 黑夜, 'nuit-pendant' ***heiye-li*** 黑夜里, 'flamme' ***dengguang*** 灯光, 'rouge figure de vieux Jean-Michel' ***lao-Yuehan-Mixier-honghongde-lian*** 老约翰·米希尔红红的脸, 'barbe rude et blanche' ***cuyingde-baixu*** 粗硬的白须, 'air bourru' ***youyuyinude biaoqing*** 忧郁易怒的表情, 'yeux vifs' ***jiongjiongyoushende-yanjing*** 炯炯有神的眼睛, 'manteau' ***waitao*** 外套, 'mouillé' ***gu-chaoqi*** 股潮气, 'pied' ***jiao*** 脚, 'pied-dessous' ***jiao-xia*** 脚下, 'paire gros bleu chausson' ***shuang-da-lan-buxie*** 双大蓝布鞋, 'signe' ***shoushi*** 手势, 'ses blonds cheveux' ***tade-danhuang-toufa*** 她的淡黄头发, 'figure douce comme mouton' ***mianyang-ban-heshande-lian*** 绵羊般和善的脸, 'des taches de rousseur' ***xie-queban*** 些雀斑, 'grosses pâles lèvres' ***meiyou-xuese-de-hou-zuichun*** 没有血色的厚嘴唇, 'yeux' ***yanjing*** 眼睛 et 'prunelle' ***yanzhu*** 眼珠 ;
- 3 référents en chaîne : 'nouveau-né bébé' ***chushengde-yinger*** 初生的婴儿 (5 maillons), 'vieux' ***laoren*** 老人 (5 maillons) et 'mère' ***muqin*** 母亲 (4 maillons).

  Les deux chaînes les plus longues comportent chacune cinq maillons.

  Les chiffres absolus et les proportions sont présentés dans le tableau (1.3) ci-dessous.

Tableau (1.3) : Comparaison entre deux extraits d'un roman-fleuve en français et en chinois

Langue	Fr	Ch
Nbre de	Fragment [19] (*Jean-Christophe*)	Fragment [20] (*'Jean-Christophe' Kelisiduofu* 克利斯朵夫)
Mots / caractères	176	251
ER	53 (30 %)	44 (18 %)
Réf	25 (14 %)	32 (13 %)
CR	3 (12 %)	3 (9 %)
Maillons de la chaîne la plus longue	14 (8 %)	5 (2 %)

Les chiffres obtenus dans les deux textes parallèles en relation de traduction révèlent des écarts imprévus. En premier lieu, le nombre de mots en français est inférieur à celui du chinois : 176 mots français contre 251 caractères chinois ; les proportions seraient ainsi plus intéressantes que les chiffres absolus. En second lieu, les pourcentages en français sont plus élevés que ceux du chinois. La proportion des expressions référentielles en français (30 %) est beaucoup plus grande que celle du chinois (18 %) ; celle des référents en français (14 %) est légèrement supérieure à celle du chinois (13 %) ; celle des chaînes est plus élevée en français (12 %) qu'en chinois (9 %) ; celle enfin des maillons de la chaîne la plus longue est également plus significative en français (8 %) qu'en chinois (2 %).

Notre première hypothèse sur les tendances et les corrélations des paramètres référentiels dans les textes cohésifs et non cohésifs est donc remise en question : s'il y a des corrélations, celles-ci devraient toutes être très variables selon les langues. Il se peut que chaque langue présente des chiffres absolus et des pourcentages très différents dans les textes cohésifs même si ces textes sont en relation de traduction.

Récapitulons. La double comparaison que nous avons établie montre quelques caractéristiques potentielles reflétées par les chaînes de référence dans un vrai texte et dans un texte composé de phrases non cohésives en français et en chinois ; les chiffres obtenus sont rassemblés dans le tableau (1.4). Nous testerons ces indices sur une plus grande quantité de textes pour voir si ces tendances sont seulement aléatoires ou si nous obtenons des résultats similaires dans les autres textes.

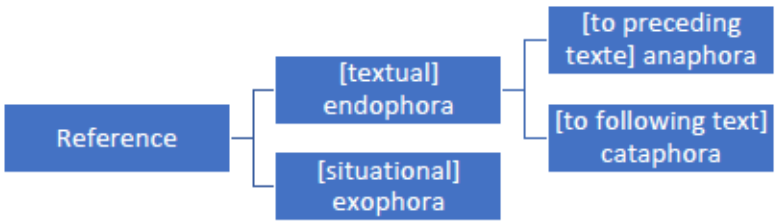
Tableau (1.4) : Comparaison entre une copie d'élève en français, un extrait d'un roman-fleuve en français et un extrait d'un roman-fleuve en chinois

Texte	Copie d'élève en français	Extrait de roman-fleuve en français	Extrait de roman-fleuve en chinois
	Fragment [15] (*Sortie à la dune*)	Fragment [19] (*Jean-Christophe*)	Fragment [20] (*'Jean-Christophe' Kelisiduofu* 克利斯朵夫)
Mots / caractères	76	176	251
ER	27 (36 %)	53 (30 %)	44 (18 %)
Réf	18 (24 %)	25 (14 %)	32 (13 %)
CR	2 (11 %)	3 (12 %)	3 (9 %)
Maillons de la chaîne la plus longue	4 (5 %)	14 (8 %)	5 (2 %)

Chapitre 2 : Trois notions-clés fondamentales pour examiner les problèmes des chaînes de référence relatifs à l'approche franco-chinoise

Dans les définitions des chaînes de référence que nous avons présentées dans le *Chapitre 1*, deux notions fondamentales sont évoquées par plusieurs auteurs, à savoir celles d'*anaphore* et de *coréférence*, qui sont, avec la notion de *chaîne de référence*, les notions les plus discutées et les plus comparées dans beaucoup d'études, entre autres celles de Schnedecker (2019, 2021). Par ailleurs, certaines autres notions homologues sont également mentionnées ; elles nous permettent de compléter les potentielles relations incluses dans les chaînes de référence de deux manières. En premier lieu, la notion d'*anaphore* à strictement parler ne nous paraissant pas suffisante pour décrire les relations inter-maillonnaires, nous adoptons la notion d'*endophore* de Halliday & Hasan (2013 [1976]) qui inclut deux ordres de composants, celui d'*anaphore* et celui de *cataphore* : la référence *endophorique* permet donc de décrire les deux relations entre les composants des chaînes de référence. Ce mode de référence endophorique s'inscrit dans le contexte linguistique, à savoir qu'il existe soit un texte précédent soit un texte suivant qui permet aux lecteurs d'identifier le référent dans un contexte linguistique. En second lieu, une autre notion, celle de *déixis*, est ajoutée, à savoir la référence *exophorique*, qui, selon Halliday & Hasan (2013 [1976]), permet d'établir une relation entre les expressions et les entités extra-linguistiques du monde réel. L'*endophore* et l'*exophore* sont donc les deux premières notions-clés essentielles à la compréhension de la notion de chaîne de référence que nous discuterons dans ce chapitre.

Figure (2.1) : Sous-types de la référence selon Halliday & Hasan (2013 [1976] : 33)



Ces deux notions sont résumées, d'un côté, par *endophore* et *exophore* selon le cadre de la typologie référentielle de Halliday & Hasan (2013 [1976]), comme le montre la figure (2.1), et, d'un autre côté, par deux types de connexions référentielles par Corblin (1995) en tant que connexions *linguistiques* et *communicatives* relativement à la coréférence : la coréférence endophorique et exophorique fait partie des connexions linguistiques ; la coréférence non endophorique et non exophorique fait partie des connexions communicatives :

> *La connexion référentielle entre des mentions dans le discours peut avoir deux fondements bien distincts :*
>
> *1. Certaines connexions sont construites sur des bases **linguistiques**. Les facteurs communicatifs et pragmatiques affectent alors seulement la sélection du contexte pertinent et de l'antécédent. Tel est le cas pour les **relations antécédent / pronom**, les **GN démonstratifs**, et les **usages « anaphoriques » du défini**.*

> *2. Certaines connexions s'établissent sur des bases* ***communicatives****. Des expressions sont associées à un designatum en vertu de connaissances contingentes, et si nous les traitons comme équi-référentes, c'est uniquement en fonction de connaissances constituées dans la communication. Tel est le lien qui unit par exemple des occurrences de* ***nom propre****, de* ***descriptions définies complètes****, et* ***d'indéfinis****.*
>
> *(Corblin, 1995 : 167, nous soulignons)*

L'*endophore*, l'*exophore* ainsi que la *coréférence* semblent donc permettre de compléter l'explication des phénomènes concernés dans les chaînes de référence.

Outre le domaine sémantique, ces trois notions sont également importantes dans l'approche sémiologique. Hamon (1972) considère en effet que les modes de référence endophorique, exophorique et non endo-exophorique peuvent être incarnés par les relations entre les signes (*i.e.* les expressions) et les significations (*i.e.* la référence) fixes ou flottantes : les trois types de signes correspondent aux trois relations incluses dans les chaînes de référence : les *signes référentiels*, les *signes déictiques* et les *signes anaphoriques*, définis comme suit :

> *1. Les signes qui renvoient à une réalité du monde extérieur (table / girafe / Picasso / rivière…) ou à un concept (structure / apocalypse / liberté…). On peut les appeler,* **référentiels****. Ils font tous référence à un savoir institutionnalisé ou à un objet concret* appris*. (Sémantique plus ou moins « stable » et permanente.) On* reconnaît *de tels signes, ils sont définis par le dictionnaire.*
>
> *2. Les signes qui renvoient à une instance d'énonciation, signes à contenu « flottant » qui ne prennent sens que par rapport à une situation concrète de discours (hic et nunc), que par rapport à un acte historique de parole déterminé par la contemporanéité de ses composants (je / tu / ici / demain / ceci…). Ce sont les « circonstanciels égocentriques » de Russel, les «* ***déictiques*** *», ou les « embrayeurs » de Jakobson, non définis « sémantiquement » par le dictionnaire.*
>
> *3. Les signes qui renvoient à un signe disjoint du même énoncé, proche ou lointain, soit antécédent dans la chaîne parlée – ou écrite –, soit postérieur. Leur fonction est essentiellement cohésive, substitutive, et économique. Ils abaissent en effet le coût du message, ainsi que sa longueur. On peut les appeler globalement* **anaphoriques** *(le nom propre et l'article dans certains de leurs emplois, la plupart des pronoms, les substituts divers, le verbe « vicaire »* faire*, etc.). Leur contenu, lui aussi flottant et variable, est uniquement fonction du contexte auquel ils renvoient.*
>
> *(Hamon, 1972 : 94-95, nous soulignons)*

Dans ce chapitre, nous présenterons, sous l'angle de l'approche franco-chinoise, les discussions relatives à ces trois notions-clés pour comprendre celle de *chaîne de référence*. Cette démarche permettra de mieux examiner les composants réguliers et spécifiques des chaînes de référence dans chaque langue.

2.1. Notion de *référence*

Avant de présenter la typologie des modes de référence et de coréférence, il est important de rendre limpide la notion de *référence* et ses dérivés, à savoir le substantif *référent* et le verbe *référer*.

Depuis la première définition de *chaîne*, la *référence* est, selon Chastain, présente comme notion fondamentale dans la *chaîne anaphorique*. Une *référence* se fonde sur l'idée de *valeur de vérité* de Leibniz (cité et traduit par Roussarie, 2007) et se développe ensuite dans la théorie du philosophe allemand Frege (1892) par le biais du terme *Bedeutung*, traduit par *référence* en français dans de nombreuses études[1]. Dans la *Grammaire Méthodique du français*, un acte de *référence* est défini comme suit :

> *L'**acte de référence** consiste à utiliser des formes linguistiques (mots, syntagmes, phrases) pour évoquer des entités (objets, personnes, propriétés, procès, événements) appartenant à des univers réels ou fictifs, extérieurs ou intérieurs. (Riegel* et al.*, 2016 [1994] : 959, nous soulignons)*

Afin de proposer une définition générale et synthétique, disons que la *référence* décrit une relation entre des signes linguistiques et des entités réelles dans le monde extra-linguistique[2]. Dans une relation référentielle, l'entité du monde réel s'appelle *référent* ; le signe linguistique s'appelle *expression référentielle*. Considérons la définition que donne Charolles (2002) des expressions référentielles en français :

> *Expressions référentielles : toutes les expressions à même de faire allusion à une entité extralinguistique (concrète, abstraite, générique, spécifique, etc.) existant dans un monde quelconque. (Charolles, 2002 : 247)*

Le référent est donc l'*entité* dans la définition de *chaîne anaphorique* de Chastain : une entité quelconque de qui ou de quoi on parle. Cette entité ne se limite à un objet concret, mais peut présenter tout type de réalités : une personne (« un officier »), un objet (« un mur », un état abstrait (« l'emprisonnement ») et un geste (« un battement »), comme c'est le cas dans l'extrait [1] du *Chapitre 1*.

Avec ce groupe de notions basiques, nous poursuivons l'examen des différents modes de référence dans les chaînes de référence.

2.2. Références endophoriques : *anaphore* et *cataphore*

L'anaphore est une notion à la fois rhétorique et discursive pour décrire un procédé de reprise vers l'arrière. Elle est définie par TLFi (*Trésor de la Langue Française informatisé*) comme « procédé visant à un effet de symétrie, d'insistance ou propositions successives[3]. » Cette notion implique un ordre spécifique entre les expressions concernées, comme Corblin (1985) le rappelle :

> *On parle d'anaphore lorsqu'un élément, par exemple un pronom, exige d'être interprété par emprunt à un terme du contexte proche, lorsqu'il y a dépendance*

---

[1] *Bedeutung* a plusieurs traductions en français selon les auteurs, comme *référence* (Black, 1952) et *dénotation* (Roussarie, 2007).

[2] *Cf.* aussi Ducrot & Todorov (1972), Dubois *et al.* (1973, 2012), Lyons (1995), Kleiber (1981, 1984, 1997a), Charolles (2002), Gary-Prieur (2011), Li (李虹, 2015), entre autres.

[3] Voir http://atilf.atilf.fr/dendien/scripts/tlfiv5/advanced.exe?8;s=2623751355 ; dernière consultation en août 2022.

nécessaire d'un « anaphorique » à un antécédent qui se comporte comme une source. (Corblin, 1985 : 126)

Une entité est antérieurement mentionnée par la première expression, *i.e. antécédent*, et elle est reprise par la deuxième, *i.e. reprise anaphorique*, comme le montre l'énoncé [1] de Reichler-Béguelin (1998 : 17) : « Pierre – il » constitue une anaphore, dont « Pierre » est l'antécédent et « il » est la reprise anaphorique.

[1] J'ai rencontré [**Pierre**]. [**Il**] m'a parlé de toi.
antécédent *anaphorique*

(Reichler-Béguelin, 1988 : 17)

L'*anaphore* est une notion essentielle discutée dans un grand nombre d'études en plusieurs langues. Elle se retrouve dans la définition de plusieurs termes connexes à la *chaîne*, surtout dans celle de *chaîne anaphorique* selon Chastain. Dans la *Grammaire Méthodique du français* (Riegel *et al.*, 2016 [1994]), *anaphore* et *expression anaphorique* sont définies de la manière suivante :

*L'**anaphore** se définit traditionnellement comme toute reprise d'un élément antérieur dans un texte.*

*Une **expression** est **anaphorique** si son interprétation référentielle dépend nécessairement d'une autre expression qui figure dans le contexte antérieur. (Riegel* et al.*, 2016 [1994] : 1029, nous soulignons)*

Dans une anaphore, le premier élément s'appelle *antécédent* et constitue la source de cette relation entre deux éléments linguistiques ; le second est *anaphorique*, dont l'interprétation dépend du premier. Précisons que la relation anaphorique entre les deux éléments en question n'est ni symétrique, ni transitive : l'interprétation de la seconde expression est dépendante de celle de la première (*cf.* Corblin, 1985, 1995 ; Cornish, 1998 ; Schnedecker, 2019, entre autres). Citons les définitions des deux expressions dans la *Grande Grammaire du Français* :

Certains mots, et plus généralement certaines expressions linguistiques, sont dits anaphoriques*, c'est-à-dire renvoyant à un antécédent, et nécessitent la prise en compte des phrases précédentes pour être interprétées. Ce type de mot permet à plusieurs expressions de renvoyer au même individu ou plus généralement à la même entité, et assure également la cohérence du discours. (Abeillé & Godard (éds), 2021 : 98)*

Dans l'ordre inverse, on obtient une *cataphore*, ce qui sous-entendant un ordre contraire à celui de l'anaphore. Citons la définition de Kesik (1989) :

Nous appellerons cataphore *la relation qui s'établit entre une expression indexicale (ou en emploi indexical) et le (une partie du) contexte linguistique subséquent, et qui permet d'identifier le référent de cette expression.*

L'expression indexicale sera désormais appelée cataphorique *et le contexte corrélatif portera la dénomination de* subséquent*. (Kesik, 1989 : 37)*

Les deux expressions concernées dans une cataphore sont appelées, la première, *cataphorique*, et la seconde, *subséquent*. Comme le montre l'énoncé [2] : « il – Pierre » constitue une cataphore : l'interprétation du pronom cataphorique « l' » dépend de celle du nom propre subséquent « Pierre » :

[2] Quand je [**l'**] ai rencontré, [**Pierre**] a paru surpris de me voir.
cataphorique *subséquent*

(Reichler-Béguelin, 1988 : 17)

Les études consacrées à l'anaphore sont beaucoup plus nombreuses que celles qui le sont à la cataphore. Cela s'explique, d'une part, par le fait que la différence majeure des deux relations est l'ordre des mots, à savoir que les expressions concernées peuvent être *a priori* les mêmes ; d'autre part, l'anaphore est beaucoup plus régulièrement remarquée que la cataphore dans les études qui sont au fondement des corpus[4] ; qui plus est, la cataphore est un phénomène plutôt récemment observé en français, à la fin du XIXème siècle (*cf*. Combettes, 2017). Pour ces trois raisons, la plupart des études dédiées à l'endophore se concentrent sur les phénomènes relatifs à l'anaphore.

L'anaphore et la cataphore sont deux types de référence dits *non autonomes*[5] puisque l'interprétation d'une des deux expressions dépend fortement de celle de l'autre : comme le montrent les énoncés [1] et [2], sans « Pierre », il n'est pas possible d'identifier le référent des pronoms « il » et « l' », quel que soit l'ordre des mots. L'anaphore permet d'établir une relation avec un autre mot dans le contexte linguistique à gauche (*cf*. Cornish, 2015), contrairement à la cataphore, dont le mot relatif se situe dans le contexte à droite. Dès qu'un référent est mentionné dans le contexte linguistique, il est potentiellement repris, ce qui amène à l'interprétation des expressions endophoriques. C'est la raison pour laquelle la référence endophorique est également appelée référence *textuelle* (Halliday & Hasan, 2013 [1976] : 56).

En ce qui concerne notre étude sur les chaînes de référence, le nombre d'expressions concernées doit être supérieur à trois, à savoir que la reprise anaphorique peut constituer ensuite une cataphore avec le SN suivant. Par exemple, les trois « il » de l'extrait [3] constituent à la fois trois anaphores avec le premier maillon « Christophe » (« M-1 ») et trois cataphores avec le dernier maillon « Christophe » (« M-5 »). L'extrait parallèle chinois [4] montre les mêmes relations.

[3] [**Christophe**]$_{M-1}$ en était le général, [**il**]$_{M-2}$ marchait devant elles, [**il**]$_{M-3}$ leur donnait l'exemple, [**il**]$_{M-4}$ montait à l'assaut des talus. Quand la branche était flexible, elle se transformait en fouet. [**Christophe**]$_{M-5}$ montait à cheval, [**Ø**]$_{M-6}$ sautait des précipices.

[4]

[克利斯朵夫]$_{M-1}$	是	将军，	[他]$_{M-2}$	以身作则，	[Ø]$_{M-3}$	跑	在	前面，	[Ø]$_{M-4}$	冲上	山坡	去	袭击。	要是	树枝	柔软	的话，	便
Kelisiduofu	*shi*	*jiangjun*	*ta*	*yishenzuoze*		*pao*	*zai*	*qianmian*		*chongshang*	*shanpo*	*qu*	*xiji*	*yaoshi*	*shuzhi*	*rouruan*	*dehua*	*bian*
Christophe	être	général	3P	donner-exemple		courir	PREP	devant		monter	colline	pour	attaquer	si	branche	flexible	si	donc

可	做	一	条	鞭子。	[克利斯朵夫]$_{M-5}$	骑着	马	跳过	危崖绝壁。									
ke	*zuo*	*yi*	*tiao*	*bianzi*	*Kelisiduofu*	*qizhe*	*ma*	*tiaoguo*	*weiyajuebi*									
pouvoir	faire	un	CL	fouet	Christophe	montant	cheval	sauter	précipice									

[4] Givón (2017) : « Subject referents in Mandarin, claiming 82.9 % of zero-anaphora in the text, exhibits higher cataphoric persistence – thus higher topicality – in 79.0 % of their occurrence in text. In contrast, direct objects, claiming only 6.2 % of zero anaphora in the text, exhibit lower cataphoric persistence – thus lower topicality – in 86.7 % of their occurrence in text ».

[5] La référence *non autonome* est le contraire de la référence *autonome*. Cette dernière est définie par Charolles (2002) comme suit : « On dit d'une expression qu'elle est autonome référentiellement quand elle peut, par ses seules ressources, évoquer un référent. » (2002 : 247).

Après avoir passé en revue ces définitions, nous nous intéressons aux formes linguistiques en français et en chinois : concrètement, comment les références endophoriques sont-elles réalisées dans ces deux langues ? Quels sont les points communs et les différences théoriques entre elles ?

2.2.1. Expressions endophoriques typiques : nominales, pronominales et zéros

La typologie des anaphores est similaire en français et en chinois. Les trois types principaux sont : i) l'anaphore *nominale*, composée d'un antécédent nominal et d'un anaphorique nominal, ii) l'anaphore *pronominale*, composée d'un antécédent nominal et d'un anaphorique pronominal et iii) l'anaphore *zéro*, composée d'un antécédent nominal et d'un anaphorique zéro (*cf*. Cornish, 1985 ; Jiang, 姜望琪, 2011 ; Xu, 许余龙, 2013).

Les formes impliquées dans ces relations ne sont pas identiques en français et en chinois, surtout en ce qui concerne les parties nominales de chaque langue. Les linguistes français tendent à classer ces dernières par leur déterminant, d'où l'anaphore *définie* et l'anaphore *démonstrative* (*cf*. Cornish, 1985, 1995 ; Adam, 2015) ; les linguistes chinois, quant à eux, classent en priorité les distinctions formelles selon la tête lexicale, d'où l'anaphore *fidèle* et l'anaphore *infidèle* (*cf.* Xu & Webster, 1999). Une anaphore nominale peut être formulée de manière simpliste comme *SN – SN*, hormis que les SN en question sont variables et spécifiques en termes de forme dans chaque langue.

Pour ce qui est des anaphores nominales en français, les expressions impliquées sont les SN définis, marqués par les articles définis « le / la / les / l' », et les SN démonstratifs, marqués par les déterminants démonstratifs « ce (-là) / cet (-là) / cette (-là) / ces (-là) ».

L'emploi anaphorique des SN définis a pour but d'extraire un individu d'un ensemble antérieurement mentionné ou généralement existant dans l'esprit du locuteur, ce qui est le cas dans l'exemple [5].

[5] Une princesse était amoureuse d'[**un jeune homme**] qui ne plaisait pas au roi son père.
antécédent

[**Le jeune homme**] …
SN-déf

(Corblin, 1995 : 68)

Selon l'étymologie, *démonstratif* signifie *montrer un objet* (Ducrot & Todorov, 1972 : 322 ; Gary-Prieur & Léonard, 1998 : 12). Les SN démonstratifs anaphoriques réfèrent à un individu défini et spécifique par un *principe de proximité* (Corblin, 1987 : 195, 1995 : 51 ; Chen, 陈平, 1987 : 81 ; Reichler-Béguelin, 1998 : 96 ; Kleiber, 1999 : 74) quand ils sont en position de reprise anaphorique et coréférentielle, c'est-à-dire qu'ils coréfèrent à l'entité la plus récente mentionnée, comme on peut le voir dans l'exemple [6]. Schnedecker (1997 : 34) l'appelle *stratégie de distance minimale* ; l'individu est extrait de la classe antérieurement mentionnée la plus proche.

[6] Une princesse était amoureuse d'[**un jeune homme**] qui ne plaisait pas au roi son père.
antécédent

[**Ce jeune homme**] …
SN-dém

(Corblin, 1995 : 68)

En ce qui concerne le chinois, les expressions anaphoriques nominales peuvent être marquées par les déterminants démonstratifs, comme ceux de proximité, 'ce / cet / cette' *zhe* 这 et 'ces' *zhexie* 这些, et ceux d'éloignement, 'ce / cet / cette N-là' *na* 那 et 'ces N-là' *naxie* 那些. Pour les expressions anaphoriques nominales « non marquées », à savoir sans déterminant, c'est la reprise nominale fidèle complète, fidèle partielle et infidèle qui constitue les anaphores nominales. Citons l'exemple [7] avec une anaphorique fidèle partielle démonstrative (dont le déterminant démonstratif n'est pas obligatoire pour la constitution d'une relation anaphorique) et une anaphorique fidèle partielle :

[7] 芬兰 并不是 以 金矿 储存量 和 淘金业 闻名于世， 但 在 她的 北方 小镇
fenlan bingbushi yi jinkuang chucunliang he taojinye wenmingyushi dan zai tade beifang xiaozhen
Finlande NEG par or-mine quantité et orpaillage connu mais PREP sont nord village

萨瑞沙卡 却 建有 [一 座 淘金 博物馆]。[（这 个） 博物馆] 详尽地 反映 了
Saruishaka que jianyou yi zuo taojin bowuguan zhe ge bowuguan xiangjinde fanying le
Saariselkä mais établi un CL orpaillage musée ce CL musée détaillé exposer LE
antécédent *anaphorique-fidèle-partielle(-dém)*

芬兰 和 世界各地 人类 从古到今 的 淘金 历史， 生动地 体现 了
fenlan he shijiegedi renlei conggudaojin de taojin lishi shengdongde tixian le
Finlande et monde-entier humain de-ancienneté-à-aujourd'hui DE orpaillage histoire réellement exposer LE

世界各地 淘金工人 的 劳动 和 生活。 那天， 我们 遇到 前来 [博物馆]
shijiegedi taojingongren de laodong he shenghuo natian women yudao qianlai bowuguan
monde-entier orpailleur DE travail et vie ce-jour-là 1Ps rencontrer venant musée
anaphorique-fidèle-partielle

上 自然 和 地理 课 的 一 群 当地 小学生。
shang ziran he dili ke de yi qun dangdi xiaoxuesheng
avoir histoire-naturel et géologie cours DE un groupe local écolier

'La Finlande n'est pas connue pour ses mines d'or ou pour son orpaillage, mais un petit village situé au nord, Saariselkä, possède un musée d'orpaillage. Le (/ ce) musée présente de manière détaillée l'histoire de l'orpaillage en Finlande et dans le monde entier, des temps anciens jusqu'à aujourd'hui. Il expose réellement le travail et la vie des orpailleurs dans le monde entier. Un jour, nous avons rencontré un groupe d'écoliers locaux qui avaient leurs cours d'histoire naturelle et de géologie au musée.

(Xu & Webster, 1999 : 96, nous traduisons)

Le deuxième sous-type d'anaphore peut être représenté par la formule *SN – pronom*. L'énoncé [1] montre ce cas (« Pierre – il »), dans lequel l'expression anaphorique est un pronom[6]. Le « pronom » impliqué dans cette anaphore renvoie surtout à ceux de troisième personne. Les pronoms singuliers de troisième personne s'avèrent en effet pouvoir indiquer régulièrement la maintenance référentielle dans n'importe quel genre de discours (*cf.* Combettes, 1986 : 81 ; Kleiber, 1994a : 111 ; Schnedecker, 1997). Les pronoms de première et de deuxième personnes fonctionnent autrement ; nous les classons comme expressions déictiques et nous les examinerons dans la section *2.3. Référence exophorique : déixis*.

Le troisième sous-type d'anaphore concerne le zéro, à savoir le fait que la reprise anaphorique n'est pas exprimée par une forme explicite. Comme le montre l'extrait [4], 'Christophe' *Kelisiduofu* 克利斯朵夫 (« M-1 ») est l'antécédent dans deux relations anaphoriques zéro respectivement avec « M-3 » et « M-4 ». Ce phénomène est remarqué dans les deux langues, alors que le chinois est par ailleurs particulièrement connu pour la grande fréquence de son anaphore zéro, comme ce que nous avons présenté les zéro dans les

[6] Bianco (1992 : 95, citée par Schnedecker, 1997 : 31) signale la différence entre une anaphore nominale et une anaphore pronominale : « D'une manière générale, la lecture d'une phrase anaphorique est plus rapide lorsque l'antécédent est répété plutôt que pronominalisé ».

chaînes topicales dans *Chapitre 1*[7]. Leurs occurrences sont susceptibles d'être affectées par divers facteurs textuels, comme le signale Huang (1996) :

> *A fundamental assumption in generative theory was that the reference of pronouns and anaphors (including reflexives) depends crucially on the distance and relative prominence of their intended antecedents in relation to them, and that the domain in which to define the reference of anaphors is in complementary distribution to the domain in which to define the reference of pronouns. (Huang, 1996 : 73)*

En résumé, dans les trois types d'anaphore, les expressions endophoriques impliquées dans les chaînes de référence peuvent être les noms propres et les SN indéfinis au rôle d'antécédent, et les SN définis, les SN démonstratifs, les SN sans déterminant, les pronoms et les zéros peuvent être à la place de reprise anaphorique.

2.2.2. Expressions endophoriques exceptionnelles : verbales et adverbiales

La fonction référentielle des expressions verbales (y compris les verbes seuls et les groupes verbaux) et des expressions adverbiales (y compris les adverbes seuls et les groupes adverbiaux) est très discutable. Il se peut pourtant que ces expressions deviennent exceptionnellement des composants de chaînes de référence : d'un côté, ce sont les deux catégories qui ne réfèrent pas selon la grammaire, mais d'un autre côté, les entités désignées par ces expressions peuvent être reprises anaphoriques dans le texte. Donc, étant composants des chaînes anaphoriques, les expressions verbales et adverbiales peuvent, d'après certains linguistes, présenter une identité référentielle renvoyant à des gestes, à des événements et à des repères spatio-temporels, et ceci constitue un point similaire entre le français et le chinois.

En ce qui concerne les verbes, Gardelle (2012), Huyghe (2012) ainsi que la *Grammaire méthodique du français* (Riegel *et al.*, 2016 [1994]) indiquent leur emploi anaphorique en français et en anglais. Avant d'être mentionnés pour la troisième fois pour constituer enfin une chaîne de référence, les gestes sont repris anaphoriquement dans le discours par les groupes verbaux dans les exemples suivant :

[8] Please [**complete this form in full**]. Failure to [**do so**] will result in your application to join the scheme being rejected.
antécédent *reprise-anaphorique-verbale*
'Veuillez [**remplir ce formulaire au complet**]. Si vous ne [**le faites**] pas, votre demande d'adhésion au régime sera rejetée.'

(Gardelle, 2012 : 25, nous traduisons)

[9] Claire ne sait pas [**réparer sa voiture**], mais Florence peut [**le faire**].
antécédent *reprise-anaphorique-verbale*

(Riegel *et al.*, 2016 [1994] : 1041)

Pour le chinois, cette fonction anaphorique des verbes est aussi récurrente (*cf.* Shen, 沈家煊, 2013), puisque les verbes chinois sont polyvalents et peuvent être utilisés comme noms selon leur position syntaxique, sans marqueur formel entre la catégorie nominale et la catégorie verbale. Examinons nos exemples en chinois : *jiehun* 结婚 peut exprimer « (se) marier » en position prédicative, comme dans les énoncés [10] et [11], et « mariage » en

[7] *Cf.* Li & Thompson (1981), Huang (1994, 1996), Gundel & Fretheim (2004), Hou & Sun (侯敏 & 孙建军, 2004), entre autres.

position de sujet, comme dans l'énoncé [12]. Quand le mot est interprété comme verbe, les deux caractères peuvent être séparés et être modifiés par un déterminant quantitatif interposé que l'on appelle *verbe sécable* (*cf*. Rabut *et al*., 2012 ; Zhang, 张旺熹, 2013), comme le montre l'énoncé [13]. Dans ce dernier cas, le premier caractère *jie* 结 garde sa fonction prédicative, pouvant être interprété comme « se relier », et le deuxième *hun* 婚 sa fonction nominale « mariage ». La partie nominale du verbe sécable peut être reprise anaphoriquement.

[10]他 **[结婚]** 了。他 为 **[此]** 组织 了 隆重的 晚宴，大家 也 为

ta jiehun le ta wei ci zuzhi le longzhongde wanyan dajia ye wei

3P marier LE 3P pour cela organiser LE solennel dîner tout-le-monde aussi pour

antécédent-verbal reprise-anaphorique-pro

[此] 纷纷 送上 祝福。

ci fenfen songshang zhufu

cela tous donner vœux

reprise-anaphorique-pro

'Il est marié. Il a organisé un dîner solennel pour cela. Tout le monde lui a fait des vœux.'

[11]他们俩 **[结婚]** 了。他们 为 **[此]** 组织 了 隆重的 晚宴， 大家

tamenlia jiehun le tamen wei ci zuzhi le longzhongde wanyan dajia

3P-deux se-marier LE 3Ps pour cela organiser LE solennel dîner tout-le-monde

antécédent-verbal reprise-anaphorique-pro

也 为 **[此]** 纷纷 送上 祝福。

ye wei ci fenfen songshang zhufu

aussi pour cela tous donner vœux

reprise-anaphorique-pro

'Ils se sont mariés. Ils ont organisé un dîner solennel pour cela. Tout le monde leur a fait des vœux.'

[12]**[结婚]** 是 件 人生 大 事。老人们 认为 越 早 **[结婚]**

jiehun shi jian rensheng da shi laorenmen renwei yue zao jiehun

se-marier / mariage être CL vie grand affaire vieux-personnes considérer plus tôt se-marier / mariage

antécédent-verbal/nominal reprise-anaphorique-verbale/nominale

越 幸福， 但是 年轻人 越来越 不 想 **[结婚]**。

yue xingfu danshi nianqingren yuelaiyue bu xiang jiehun

plus heureux mais jeune-personne de-plus-en-plus NEG vouloir se-marier / mariage

reprise-anaphorique-verbale/nominale

'Se marier / Le mariage est une affaire importante dans la vie. Les personnes âgées trouvent que plus tôt on se marie, plus on est heureux, mais les jeunes ne veulent plus se marier / le mariage.'

[13]他 已经 **结** 了 **[两 次 婚]**。**[这 两 次]** 的 结局 都 很 痛苦。 但是 经过

ta yijing jie le liang ci hun zhe liang ci de jieju dou hen tongku danshi jingguo

3P déjà se-marier LE deux fois mariage ce deux fois DE fin tous très douloureux mais après

antécédent-nominal reprise-anaphorique-dém

[这 两 次 不幸]， 他 又 结 了 第三次 婚。

zhe liang ci buxing ta you jie le disanci hun

ce deux fois malheur 3P encore se-marier LE troisième mariage

reprise-anaphorique-dém

'Il s'est déjà marié deux fois. Ses deux mariages ont eu une fin douloureuse. Mais après ces deux malheurs, il a effectué un troisième mariage.'

(nos exemples)

En conséquence, les expressions verbales françaises et chinoises peuvent engager le rôle anaphorique lorsque les gestes et les événements sont repris. Lors de la troisième expression renvoyant au même geste, les trois expressions d'identité référentielle constituent des chaînes de référents verbaux et spatio-temporels. Ces deux groupes d'expressions sont ainsi composants exceptionnels des chaînes de référence.

Pour ce qui est de l'emploi anaphorique des adverbes et des groupes adverbiaux, il est aisé à comprendre puisque les endroits peuvent être référents potentiels et ensuite être repris dans la suite du texte sous forme nominale ou adverbiale (*cf*. Kleiber, 1981, Gardelle, 2012 ; Longo & Todirascu, 2014 ; Riegel *et al*., 2016 [1994]). Comme le montrent les énoncés suivants, les adverbes « there » et « là » renvoient certainement à des endroits repérables dans leur

situation d'énonciation : donc, lorsque ces référents spatiaux seront repris à la troisième fois, ces expressions pourront constituer une chaîne :

[14]I put the keys [**in the top drawer**] ; they should still be [**there**].
antécédent *reprise-anaphorique-adv*

'J'ai mis les clés dans le tiroir du haut ; elles devraient encore être là.'

(Gardelle, 2012 : 26, nous traduisons)

[15]Il est [**une contrée**] […]. C'est [**là**] qu'il faut aller vivre, c'est [**là**] qu'il faut aller mourir.
antécédent *reprise-anaphorique-adv* *reprise-anaphorique-adv*

(Baudelaire, cité par Riegel *et al.*, 2016 [1994 : 1041)

En réalité, lorsque les expressions adverbiales font partie des chaînes de référence, il s'agit également de leur fonction déictique, que nous examinerons dans la partie suivante.

2.3. Référence exophorique : *déixis*

Le terme de *déixis* provient du grec signifiant *montrer* ou *indiquer*. Les expressions déictiques s'inscrivent dans une situation d'énonciation[8] où sont présentes les entités auxquelles elles renvoient. Cette situation d'énonciation peut avoir lieu à l'oral ; elle peut également s'établir à l'écrit par le biais du contexte linguistique précédent. Kleiber (1986) précise la dépendance de l'interprétation référentielle d'expressions déictiques au contexte situationnel comme suit :

> *Une expression déictique est une expression dont le sens implique nécessairement que le référent soit présent dans la situation d'énonciation. (Kleiber, 1986 : 11)*
>
> *Les déictiques sont des expressions qui renvoient à un référent dont l'identification est à opérer nécessairement au moyen de l'entourage spatio-temporel de leur occurrence. La spécificité du sens indexical est de « donner » le référent par le truchement de ce contexte. (Kleiber, 1986 : 19)*

À la différence de l'anaphore, la déixis permet de créer directement un lien entre une expression et une entité dans la situation d'énonciation sans passer par le contexte linguistique (*cf.* Kleiber, 1986, 1991 ; Reichler-Béguelin, 1988 ; Corblin, 1995 ; Leeman, 2004 ; Cornish, 2015). Kleiber (1991) explicite ainsi les différences entre les deux notions :

> *- Une expression est anaphorique si son référent se trouve mentionné dans le contexte linguistique ;*
>
> *- Une expression est déictique si son référent réside dans le site d'énonciation immédiat ;*
>
> *- Une expression est anaphorique si son référent est (nécessairement) identifié par l'intermédiaire du contexte linguistique ;*
>
> *- Une expression est déictique si son référent est (nécessairement) identifié grâce au site d'énonciation immédiat.*
>
> *(Kleiber, 1991 : 9)*

[8] Au sujet de la définition de la *situation d'énonciation*, voir aussi Strawson (1950), Searle (1969), Karttunen (1976), Schwarz (1979), Corblin (1987), Reichler-Béguelin (1988, 1998), Ariel (1990, 2013), Charolles (2002), Wilson & Sperber (2004), Chen (陈小平, 2012) et Rohde & Kehler (2014).

Selon de nombreuses études[9], trois principaux types d'entités sont concernés : les interlocuteurs (« je », « tu »), les objets (« ceci », « cette voiture ») et les repères spatio-temporels (« ici », « là-bas », « maintenant », « en ce moment ») de l'énonciation. Ces trois types d'entités correspondent à trois types d'expressions : les pronoms de première et deuxième personnes renvoient aux interlocuteurs ; les SN démonstratifs et les pronoms démonstratifs renvoient aux objets mentionnés dans la situation ; les adverbes spatio-temporels renvoient aux lieu et temps de l'énonciation[10].

Les SN démonstratifs sont présentés dans la section précédente comme expressions anaphoriques, alors qu'ils peuvent également être utilisés comme expressions déictiques lorsqu'ils se réalisent dans une situation d'énonciation actualisée, sachant que les particules démonstratives « -ci » et « -là » sont dérivées respectivement des adverbes spatiaux « ici » et « là » (*cf.* Cornish, 2017). Kleiber (1986) propose les exemples suivants pour comparer ces deux types de SN démonstratifs français : l'interprétation référentielle de « cet homme » de l'énoncé [16] dépend de l'expression du contexte linguistique « un homme » ; donc le premier est anaphorique. Celle de l'énoncé [17] s'inscrit dans un contexte extra-linguistique, à savoir qu'il n'y a pas d'expression coréférentielle explicite. Si cette expression permet d'identifier une personne particulière, c'est parce qu'elle est présente dans une situation d'énonciation spécifique.

[16][Un homme] entra, un chapeau de femme sur la tête. [**Cet homme**] était fou.
SN-dém-anaphorique

[17]Mais [**cet homme**] est fou ! (en pointant sur un référent de la situation d'énonciation)
SN-dém-déictique

(Kleiber, 1986 : 6)

Les énoncés [18] et [19] illustrent l'emploi de trois types d'expressions déictiques dans deux contextes spécifiques : « moi » (déictique personnel) renvoie au locuteur de l'énoncé, « là » (déictique spatial) au lieu et « aujourd'hui » (déictique temporel) au temps des deux situations d'énonciation spécifiques.

[18]Laisse-[**moi**] [**là**] !
déictique-personnel *déictique-spatial*

(Poirot-Delpech, 1984, cité par Abeillé & Godard (éds), 2021 : 1995)

[19][**Aujourd'hui**] ils ont mijoté de l'écureuil en ragoût, comme hier, comme demain.
déictique-temporel

(Rambaud, 1997, cité par Abeillé & Godard (éds), 2021 : 1995)

Toutes les expressions déictiques ne permettent certainement pas d'établir une relation coréférentielle, comme le montre « cet homme » de l'exemple [17], puisque la personne n'est pas reprise dans son contexte discursif. Cependant, si l'énoncé apparaît comme un tour de parole dans un dialogue de roman et si la personne est rementionnée dans son contexte, « cet homme » peut sans nul doute être composant de la chaîne de cette personne. Avec les autres reprises, toutes les expressions coréférentielles peuvent constituer de manière certaine des relations anaphoriques entre elles ; par conséquent, cette expression démonstrative peut être à la fois déictique et anaphorique. C'est pour cette raison que nous

[9] *Cf.* Ducrot & Todorov (1972), Quine (1977), Lyons (1978), Kleiber (1986, 1989, 1991), Reichler-Béguelin (1988), Charolles (2002), Dubois *et al.* (2012), Cornish (2015), Riegel *et al.* (2016 [1994]) et Abeillé & Godard (éds, 2021), entre autres.

[10] Excepté ces trois types d'expressions, les SN définis peuvent également être déictiques, *cf.* Hawkins (2015).

présentons une notion intermédiaire entre l'anaphore et la déixis dans la prochaine partie : l'*ana-déixis*.

2.4. Entre l'endophore et l'exophore : l'*ana-déixis*

Comme nous l'avons noté dans la section précédente, les SN démonstratifs peuvent être utilisés de manière anaphorique ou déictique selon le contexte. La frontière entre ces deux types d'expressions n'est donc pas tellement rigide. Les SN démonstratifs ne sont pas les seules expressions de ce cas : Kleiber (1986 : 6) indique, en plus des pronoms *anaphoriques* (« il ») et *déictiques* (« je », « tu »), un troisième groupe *mixte*, à savoir « nous », pronom qui combine le sens du pronom déictique « je » et celui du pronom anaphorique « il(s) ». Pour nommer le même type d'expressions qui combinent les caractéristiques des anaphoriques et des déictiques, Cornish (2007, 2010, 2013, 2017) propose le terme d'*ana-déixis*, avec une échelle de degrés indexicaux en fonction de la forme des expressions pour le français et l'anglais, comme le montre la figure (2.2). Les catégories d'expressions sont classées selon leur tendance à être déictique ou anaphorique : les expressions déictiques sont à gauche de l'échelle, les expressions anaphoriques à droite ; les expressions ana-déictiques sont au milieu :

Deux groupes sont déictiques et non anaphoriques :

- « p1ère/2èmep » : les pronoms de première et deuxième personne, comme « je / me » et « tu / te » ;
- « AdvdmP » : les adverbes démonstratifs proximaux, comme « ici » et « maintenant » ;

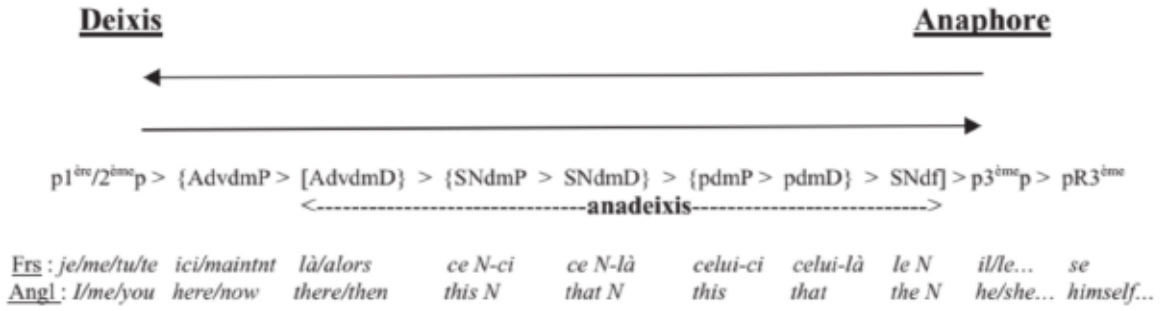
Cinq groupes sont ana-déictiques et concernent la plupart des formes de l'échelle :

- « AdvdmD » : les adverbes démonstratifs distaux, comme « là » et « alors » ;
- « SNdmP » : les SN démonstratifs proximaux, comme « ce / cet / cette / ces N-ci » ;
- « SNdmD » : les SN démonstratifs distaux, comme « ce / cet / cette / ces N-là » ;
- « pdmP » : les pronoms démonstratifs proximaux, comme « celui-ci / celle-ci / ceux-ci / celles-ci » ;
- « pdmD » : les pronoms démonstratifs distaux, comme « celui-là / celle-là / ceux-là / celles-là » ;

Trois groupes sont anaphoriques et non déictiques :

- « SNdf » : les SN définis, comme « le / la / l' / les N » ;
- « p3èmep » : les pronoms de troisième personne, comme « il / le », « elle / la » et « ils / elles / les » ;
- « pR3ème » : les pronoms réfléchis de troisième personne, comme « se ».

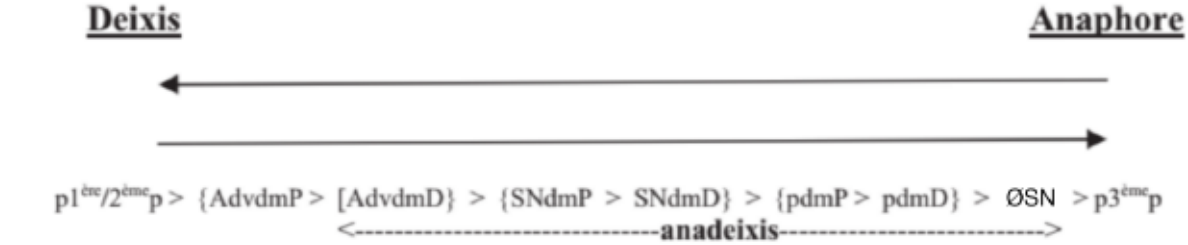
Figure (2.2) : Échelle d'indexicalité encodée par certaines catégories d'expressions indexicales selon Cornich (201 : 219)[11]



Frs : *je/me/tu/te*	*ici/maintnt*	*là/alors*	*ce N-ci*	*ce N-là*	*celui-ci*	*celui-là*	*le N*	*il/le…*	*se*
Angl : *I/me/you*	*here/now*	*there/then*	*this N*	*that N*	*this*	*that*	*the N*	*he/she…*	*himself…*

À partir de cette échelle pour le français et l'anglais, nous essayons de formuler une échelle pour le chinois, comme le montre la figure (2.3), dans laquelle le « SNdf » (SN défini) est remplacé par « ØSN » signifiant SN sans déterminant ; le « pR3ème » a disparu puisque cette catégorie n'existe pas en chinois.

Figure (2.3) : Échelle d'indexicalité encodée pour le chinois



Ch : *wo* 我	*zheli* 这里	*nali* 那里	*zhege N* 这个N	*nage N* 那个N	*zhege* 这个	*nage* 那个	N	*ta* 他
ni 你	*xianzai* 现在	*gangcai* 刚才						*ta* 她 *ta* 它

Ces deux échelles d'expressions ana-déictiques pour le français et le chinois pourront servir de critères formels pour délimiter les composants des chaînes de référence dans les dialogues des textes.

2.5. Notion de *coréférence*

Dans la définition de *chaîne de référence* de Corblin, le terme d'*anaphore* n'est pas présent. L'auteur met plutôt en relief la notion d'*interprétation d'identité référentielle*, qui permet d'établir des relations de *coréférence* entre les expressions. Cette notion d'*identité* est aussi utilisée par Halliday & Hasan (2013 [1976]) pour expliquer la relation dite de *coréférence*

[11] Cette figure apparaît pour la première fois dans l'ouvrage de Cornish (2007) ; elle a été révisée et légèrement modifiée par l'auteur dans ses publications ultérieures en 2010, 2013 et 2017. Nous citons la version la plus récente, celle de 2017.

entre les expressions. La *Grande Grammaire du français* donne la définition d'expression coréférentielle suivante :

> *Une expression peut avoir la même référence qu'une autre déjà introduite sans être anaphorique. (Abeillé & Godard (éds), 2021 : 2027)*

Selon cette définition, les expressions coréférentielles ne sont pas nécessairement anaphoriques. Les deux types d'expressions peuvent sans doute être constituants des chaînes de référence puisque leur relation de base est de renvoyer à la même entité, mais il existe des différences entre elles.

Tout d'abord, la coréférence implique une relation symétrique entre les expressions d'une chaîne, c'est-à-dire que l'interprétation référentielle de chaque expression n'est pas forcément dépendante des autres expressions. À l'inverse, les expressions anaphoriques présentent une relation asymétrique : une expression joue le rôle d'antécédent, à savoir la source référentielle, et l'interprétation des autres expressions dépend de cette première.

Ensuite, il n'y a aucun doute sur le fait que la relation « Pierre – il » dans l'énoncé [1] est anaphorique et coréférentielle. Autrement dit, les expressions anaphoriques peuvent être en même temps coréférentielles, ce qui leur permet de constituer des maillons dans les chaînes de référence. Or, toutes les endophores n'assurent pas la coréférence. Les anaphores peuvent concerner également des relations non coréférentielles, comme l'anaphore *associative*[12] (exemple [20]), l'anaphore *mentionnelle* (exemple [21]), l'anaphore *conceptuelle* (ou *lexicale*[13]) (exemple [22]), l'anaphore *générique* (exemple [23]) et l'anaphore *possessive* (exemple [24]). Dans tous ces cas, aucune paire d'expressions ne présente une relation coréférentielle.

[20](Pierre visita [un village].) [**L'église**] était triste.
(Corblin, 1987 : 128)

[21]Ne prends pas [ce vieux chapeau], prends [**le neuf**].
(Corblin, 1995 : 114)

[22]Paul a tué [trois lions]. Pierre [**en**] a tué [**cinq**].
(Milner, 1982, cité par Schnedecker, 2019 : 3)

[23]Paul a acheté [une Toyota], car [**elles**] / [**ces voitures**] sont robustes.
(Webber, 1983, cité par Kleiber, 1991, cité par Schnedecker, 2019 : 3)

[24][Paul] … [**son chien**] …
(Schnedecker, 2019 : 3)

Réciproquement, certaines expressions non anaphoriques peuvent également être coréférentielles, comme on l'a vu dans une section précédente au sujet des expressions verbales et adverbiales. Par ailleurs, les noms propres peuvent être coréférentiels entre eux mais ils ne sont ni anaphoriques, ni déictiques (*cf*. Schnedecker & Landragin, 2014). Corblin (1995 : 31) signale aussi les SN indéfinis, qui peuvent seulement jouer le rôle de l'antécédent dans l'anaphore (et de subséquent dans la cataphore). Comme le montre l'exemple [25], la relation « Pierre – Pierre » est coréférentielle non anaphorique et non déictique parce que l'interprétation référentielle des deux éléments ne requiert ni le contexte discursif ni la situation d'énonciation. La relation coréférentielle entre les deux expressions en question est symétrique et transitive.

---

[12] Les expressions impliquées dans une anaphore associative peuvent être composants des *chaînes associatives* selon la définition de Corblin (1995) que nous avons présentée dans le *Chapitre 1*.

[13] *Cf*. Milner (1982, cité par Schnedecker, 2019).

[25][**Pierre**] est venu. [**Pierre**] repartira demain.
coréférentiel-non-anaphorique

(Corblin, 1995 : 31, nous soulignons et annotons)

Dans cette partie, nous parlons donc des deux catégories d'expressions coréférentielles qui présentent des spécificités par rapport aux autres expressions anaphoriques et déictiques déjà abordée : les noms propres et les SN indéfinis.

2.5.1. Les noms propres : *désignateurs autonomes*

La référence des noms propres n'est ni endophorique ni exophorique puisque l'interprétation référentielle ne dépend ni du contexte discursif ni de la situation d'énonciation. Selon les termes de Kripke (1972, 1982, cité par Charolles, 2002 : 55), les noms propres sont des *désignateurs rigides*[14], contrairement aux autres expressions qui désignent de manière moins rigide dans la mesure où leur interprétation référentielle dépend plus ou moins du contexte.

> *Les noms propres étant liés à leur porteur par une convention qui est indépendante de leurs caractéristiques substantielles peuvent continuer à référer à ceux-ci quels que soient les changements susceptibles de les affecter. Ils constituent, de ce fait, ce que Kripke (1972, 1982) appelle des* ***désignateurs rigides****. (Charolles, 2002 : 55, nous soulignons)*

En raison de cette indépendance référentielle, ce mode de référence est ainsi appelé *autonome* par Charolles :

> *Les* ***noms propres*** *sont considérés comme autonomes référentiellement dans la mesure où ils suffisent à instancier un particulier, encore que le contexte intervienne toujours pour savoir si ce particulier est un humain, un bateau, un système informatique, etc. (Charolles, 2002 : 247, nous soulignons)*

La référence autonome des noms propres se fonde sur les *connaissances encyclopédiques* (*cf.* Kripke, 1972 : 269 ; Ariel, 1990 : 33 ; Charolles, 2002 : 143 ; Adam, 2015 : 106). Ces connaissances peuvent être les connaissances communes et générales sur le monde (« la Terre », « le soleil ») et les pratiques sociales et culturelles d'une communauté (« Pâques », « Noël »). Kleiber (1994b) élargit cette définition de *connaissances* à la *mémoire* du locuteur : ce sont non seulement les connaissances constantes, mais aussi des mémoires courtes ou longues :

> *Contexte linguistique, situation extralinguistique, connaissances générales se retrouvent tous traités mémoriellement : ils ont tous le statut de représentation interne, même s'ils se différencient quant à l'origine et au niveau de la représentation (mémoire courte, mémoire longue, etc.). (Kleiber, 1994(b) : 19)*

Contrairement aux noms communs, les noms propres peuvent référer à des entités de manière directe, sans avoir besoin de contexte linguistique ou situationnel. En comparant les énoncés [26] et [27], on peut s'interroger sur le référent de « un enfant », de [26] : quel enfant ? Qui est cet enfant ? Comment est cet enfant ? Il faudrait un contexte spécifique pour identifier l'enfant en question. En revanche, de telles questions ne se posent pas dans le cas de « Pierre » de [27], puisque le locuteur suppose que l'interlocuteur connaît une personne

[14] *Cf.* aussi Gary-Prieur & Noailly (2019).

spécifique qui est nommée « Pierre ». C'est dans ce sens que l'on dit que les noms propres peuvent référer de manière autonome.

[26][**Un enfant**] mange devant la télé.
[27][**Pierre**] mange devant la télé.

(nos exemples)

En raison de cette capacité de désigner des entités réelles dans le monde extra-linguistique, les noms propres, même s'ils ne sont ni anaphoriques ni déictiques, peuvent sans aucun doute être constituants des chaînes de référence. En tant que tels, ils présentent des caractéristiques que d'autres expressions référentielles ne possèdent pas. Dans une relation coréférentielle, ils peuvent, en premier lieu, servir à la désambiguïsation dans les situations de concurrence référentielle ; en deuxième lieu, *dénommer* er *redénommer* le référent ; et en troisième lieu, *faciliter considérablement la tâche de coréférentiation* (*cf*. Schnedecker, 1997 : 50).

Les études de Schnedecker (1992, 1997) et de Capin *et al*. (2021) remarquent, à partir de leur corpus, que les noms propres peuvent non seulement introduire un nouveau référent, le plus souvent un nouveau personnage ayant une certaine importance pour la continuité référentielle, mais qu'ils peuvent également coïncider avec le début d'une unité discursive, comme un paragraphe, pour marquer la structuration textuelle. Cette catégorie est moins souvent utilisée comme deuxième maillon d'une chaîne en raison de son processus cognitivement plus coûteux par rapport à la séquence « Npr … Pro », selon Kleiber (1994) :

> *L'emploi du nom propre en deuxième position fait […] opérer un calcul plus important (que celui du pronom), puisqu'il oblige à réeffectuer l'opération déjà réalisée par la première mention du nom propre et que cette opération est ici plus coûteuse que celle du pronom dans la mesure où elle ne tient pas compte du résultat référentiel obtenu par la première mention du nom propre.*
>
> *(Kleiber, 1994, cité par Schnedecker, 1997 : 202)*

Pourtant, les noms propres sont aussi remarqués en position non initiale des chaînes. L'usage peut impliquer trois fonctions dans le discours, d'après Schnedecker (1997) :

> *Le nom propre indiquerait qu'il faut (ou qu'on peut) déconnecter la proposition p_n[15] de son occurrence de la proposition p_{n-1} qui la précède : « disjoignez p_n de p_{n-1} » ;*
>
> *(ii) faire abstraction du contenu descriptif transféré sur les pronoms qui ont précédé son occurrence : « fermez le fichier » ;*
>
> *(iii) démarrer une « nouvelle chaîne », coréférant, visant à re-saisir de façon homogène le référent : « opérez une nouvelle saisie du référent ».*
>
> *(Schnedecker, 1997 : 153)*

Pour montrer la différence qui existe entre une anaphore et une coréférence, citons les exemples de Kleiber (1988), qui permettent de synthétiser les cas en français où les expressions coréférentielles à la seconde position ne sont pas anaphoriques, comme les pronoms personnels déictiques [28], les noms propres [29], les SN génériques [30] et les noms *unica* [31].

[28][**Je**] suis venu [**te**] voir et [**tu**] [**m'**] as fermé la porte au nez.
réf-1 *réf-2* *réf-2* *réf-1*

[15] « p_n » signifie « phrase n » ; « p_{n-1} » signifie la phrase précédent la « phrase n ».

[29][**Alfred**] était arrivé avec [**Céline**]. [**Céline**] était gaie, [**Alfred**] triste.
réf-1 *réf-2* *réf-2* *réf-1*

[30]Les vignes rapportent beaucoup [**aux viticulteurs**] mais il faut dire que [**les viticulteurs**]
réf-1 *réf-1*
travaillent durement.

[31][**La terre**] tourne autour [**du soleil**]. [**Le soleil**] est en effet le centre de l'univers.
réf-1 *réf-2* *réf-2*

(Kleiber, 1988 : 3)

Schnedecker (1997) considère la *coréférence* et l'*anaphore* comme les deux relations primordiales entre les maillons d'une chaîne de référence. Dans une étude de 2019, la linguiste met l'accent sur la distinction entre les trois notions qui sont présentes dans les définitions des autres auteurs : l'*anaphore*, la *coréférence* et la *chaîne de référence*.

> *Les notions de* coréférence *et de* chaîne de référence *supposent une forme d'identité référentielle entre les référents évoqués par les expressions référentielles, qui ne repose, ni nécessairement ni exclusivement, sur la relation d'*anaphore*, comme l'illustre [32] où les trois premiers « maillons » (ou mentions) sont des expressions référentiellement autonomes, qui donnent, si l'on peut dire, « directement » accès au référent visé, tandis que le pronom établit une relation d'anaphore coréférentielle avec les mentions antérieures. (Schnedecker, 2019 : 3)*

[32][**Barak Obama**] … [**Le 44ème président des Etats-Unis**] … [**L'époux de Michelle Robinson**] … [**Il**] …

(Schnedecker, 2019 : 3)

Cette illustration montre la complexité des relations entre les maillons dans une chaîne de référence : ces derniers peuvent coréférer aussi bien de manière autonome que de manière non autonome.

Les remarques que nous venons de faire portent sur les noms propres français, alors que la fonction référentielle des noms propres chinois, dépourvus de spécificité formelle, est souvent discutée avec celle des autres SN ; ce groupe n'est pas non plus considéré comme catégorie particulière composante des chaînes de référence. Comme nous l'avons vu au sujet de la chaîne pronominale dans le *Chapitre 1*, Xu (徐赳赳, 1990) considère les noms propres chinois tout comme les autres SN (voir l'exemple [14] du *Chapitre 1*). Il convient ainsi d'observer les noms propres chinois en tant que maillons des chaînes dans un corpus pour examiner s'ils fonctionnent vraiment comme équivalents des SN, sans grande différence ou avec des particularités distinctes.

2.5.2. Les SN indéfinis

Les SN indéfinis sont marqués par divers déterminants indéfinis au singulier et au pluriel en français et en chinois. La référence des SN indéfinis est bien distincte des autres formes, comme le note Corblin (1987) :

> *Dans un groupe nominal indéfini, le nom fournit le type d'objet sur lequel s'effectue un dénombrement de valeurs individuelle ; dans un groupe nominal défini, N fournit le type d'objet qui à lui seul doit permettre de séparer un objet des autres, il fonctionne alors comme principe d'identification de l'objet désigné. (Corblin, 1987 : 113)*

Au sujet de l'existence du référent, Leeman (2004) indique la différence qu'il y a entre les descriptions définies et les SN indéfinis en français : seuls les SN indéfinis peuvent être utilisés dans les structures existentielles.

> *Les indéfinis permettent de poser l'existence de quelque chose (de l'introduire dans le discours), tandis que les définis la présupposent (la supposent déjà établie, posée).*

[33]Il existe **une** exception / **deux** exceptions / **quelques** exception(s).
[34]*Il existe **l'**exception / **ton** exception / **ces** exceptions.

(Leeman, 2004 : 59)

Quant à leur usage coréférentiel dans les chaînes de référence, les SN indéfinis introduisent des « référents nouveaux, présumés non activés dans l'esprit des destinataires » (Charolles, 2002 : 144). Ils peuvent servir à introduire un nouveau référent au début d'un discours, sans avoir besoin d'un contexte avec un antécédent ou d'une situation d'énonciation spécifique, ce qui constitue une différence avec la coréférence des SN définis incomplets et des SN démonstratifs. Ainsi, « un garçon », de l'énoncé [35] peut être interprété de deux façons : i) la personne en question est un garçon mais non une fille ou un adulte ; ii) c'est cette seule et même personne qui vient tous les matins. Ni le SN défini incomplet « le garçon » de [36]*, ni le SN démonstratif « ce garçon » de [37]*, ne permet d'initier une chaîne de cette personne.

[35]Chaque matin, [**un garçon**] montait son courrier à Marcel.
[36]*Chaque matin, [**le garçon**] montait son courrier à Marcel.
[37]*Chaque matin, [**ce garçon**] montait son courrier à Marcel.

(Charolles, 2002 : 162)

Ce type d'initiation d'une chaîne de référence par un SN indéfini coïncide fortement avec les phrases existentielles en français et en chinois (*cf.* Kleiber, 1981 ; Galmiche, 1983 ; Charolles, 2002, Yue, 乐耀, 2010[16]). Comme le montrent les extraits parallèles [38] et [39], il est prévisible que les référents exprimés par le pronom « ils » (*tamen* 他们) et le nom propre « Christophe » (*Kelisiduofu* 克利斯朵夫) seront déjà présentés dans le contexte, alors que ceux des SN indéfinis « un jardin » (*yi-ge-yuanzi* 一个园子), « deux enfants » (*liang-ge-haizi* 两个孩子), « un garçon » (*yi-nan* 一男) et « une fille » (*yi-nü* 一女) seront nouveaux et apparaîtront pour la première fois dans le texte.

[38][Ils] arrivèrent dans [**un jardin**], où [**deux enfants**] à l'air maussade, [**un garçon**] et [**une fille**], à
nouveau-réf-SN-ind *nouveau-réf-SN-ind* *nouveau-réf-SN-ind* *nouveau-réf-SN-ind*
peu près du même âge que [Christophe], semblaient se bouder l'un l'autre.
ancien-réf-npr

[39][他们] 到 了 [一 个 园子] 里， 那儿 有 [两 个 孩子] 沉着脸， [一 男]
tamen *dao* *le* *yi* *ge* *yuanzi* *li* *naer* *you* *liang* *ge* *haizi* *chenzhelian* *yi* *nan*
3Ps arriver LE un CL jardin intérieur là-bas avoir deux CL enfant maussade un mâle
nouveau-réf-SN-ind *nouveau-réf-SN-ind* *nouveau-réf-SN-ind*

[一 女]， 和 [克利斯朵夫] 差不多 年纪， 好象 正在 生气。
yi *nü* *he* *Kelisiduofu* *chabuduo* *nianji* *haoxiang* *zhengzai* *shengqi*
un femelle avec Christophe presque âge semblant en-train-de se-bouder
nouveau-réf-SN-ind *ancien-réf-npr*

[16] Le corpus d'observation de Yue est composé de 20 textes narratifs d'environ 8 000 caractères chinois. L'auteur choisit par tirage au sort 20 étudiants chinois pour voir le film *The Pear Stories* (Chafe, 1980) dont il leur demande ensuite d'écrire l'histoire.

2.6. Bilan

La chaîne de référence est une notion complexe puisqu'elle permet d'établir des relations à quatre niveaux avec différentes relations fondamentales. Au premier niveau, avec la notion basique de référence, les expressions sont reliées aux entités réelles dans le monde extra-linguistique. Au deuxième niveau, avec la notion d'endophore, les expressions de source référentielles servent à interpréter les autres expressions de reprises. Au troisième niveau, avec la notion d'exophore, les expressions indiquent directement les entités dans la situation d'énonciation. Au dernier niveau, avec la notion de coréférence, certaines expressions réfèrent de manière autonome, grâce aux connaissances communes des locuteurs, et d'autres de façon non autonome. Les études relatives à l'endophore se situent souvent dans le cadre phrastique et inter-phrastique, alors que celles consacrées à la coréférence et aux chaînes de référence relèvent à la fois du niveau phrastique, du niveau discursif et de la linguistique des genres (*cf*. Schnedecker, 2019). Ni l'anaphore ni la coréférence seules ne suffisent pour décrire la relation entre les expressions d'une chaîne de référence.

Nous observons souvent, dans les chaînes de référence, un composé de différents types d'expressions endophoriques, exophoriques et coréférentielles, comme le montre la chaîne « Jean-Christophe » dans les extraits parallèles [40] et [41] issus du roman-fleuve *Jean-Christophe*, avec des signes de diverses catégories : nom propre autonome 'Christophe' *Kelisiduofu* 克利斯朵夫, pronom anaphorique de troisième personne 'il' *ta* 他, pronom déictique de première et deuxième personnes 'je' *wo* 我 et 'tu' *ni* 你. Malgré divers modes de référence, toutes les expressions sont coréférentielles entre elles pour enfin constituer une chaîne de référence, cela étant la même chose pour un texte narratif ou dialogal, pour un texte en français ou en chinois.

[40]Louisa servait les petits : deux pommes de terre à chacun. Lorsque venait le tour de [**Christophe**],
coréférentiel

souvent il n'en restait que trois sur l'assiette, et [**sa**] mère n'était pas servie. [**Il**] le savait d'avance,
anaphorique *anaphorique*

[**Il**] les avait comptées, avant qu'elles arrivent à [**lui**]. Alors [**il**] rassemblait [**son**] courage, et d'un
anaphorique *anaphorique* *anaphorique* *anaphorique*

air dégagé :
– Rien qu'une, maman.
Elle s'inquiétait un peu.
– Deux, comme les autres.
– Non, [**je**] t'en prie, une seule.
déictique
– Est-ce que [**tu**] n'as pas faim ?
déictique
– Non, [**je**] n'ai pas grand'faim.
déictique

[41]鲁意莎 替 孩子们 分菜， 每人 两 个 马铃薯。 轮到 [**克利斯朵夫**]， 往往 盘子
Luyisha ti haizimen fencai meiren liang ge malingshu lundao Kelisiduofu wangwang panzi
Louisa aider enfants servir chacun deux CL pomme-de-terre au-tour-de Christophe souvent assiette
coréférentiel

里 只 剩 了 三 个， 而 母亲 自己 还 没 拿。 [**他**] 早已 知道， 没
li zhi sheng le san ge er muqin ziji hai mei na ta zaoyi zhidao mei
dedans seulement rester LE trois CL et mère soi encore NEG prendre 3P déjà savoir NEG
anaphorique

轮到 [他] 就已经 数过 了，[他] 便 鼓足 勇气，装做 满不在乎的 说：
lundao *ta* *jiuyijing* *shuguo* *le* *ta* *bian* *guzu* *yongqi* *zhuangzuo* *manbuzaihude* *shuo*
au-tour-de 3P déjà compter LE 3P donc rassembler courage prétendre dégagé dire
anaphorique *anaphorique*

“只 要 一 个，妈妈。”她 有 点 不放心 了。
zhi *yao* *yi* *ge* *mama* *ta* *you* *dian* *bufangxin* *le*
seulement prendre un CL maman 3P avoir un-peu s’inquiéter LE

“两 个 罢， 跟 大家 一样。”
liang *ge* *ba* *gen* *dajia* *yiyang*
deux CL INTERJ comme tout-le-monde pareil

“不，真的，[我] 只 要 一 个。”
bu *zhende* *wo* *zhi* *yao* *yi* *ge*
non vraiment 1P seulement prendre un CL
déictique

“[你] 不 饿 么？”
ni *bu* *e* *me*
2P NEG avoir-faim INTERR
déictique

“对啦，[我] 不 大 饿。”
duila *wo* *bu* *da* *e*
oui 1P NEG trop avoir-faim
déictique

Bilan de la *Partie I*

Dans cette première partie consacrée aux cadres théoriques, nous avons passé en revue toutes les définitions et les discussions relatives à la notion de *chaîne de référence*, ainsi que les notions fondamentales et les formes d'expressions concernées. La distinction entre l'anaphore et la cataphore nous permet de présumer que l'ordre des maillons de chaîne est déterminant pour comprendre et saisir la source référentielle ; celle entre l'endophore et l'exophore nous porte à supposer que les contextes linguistique et extra-linguistique sont aussi importants l'un que l'autre pour identifier le référent. Les différentes typologies des expressions concernées dans chaque type de relation montrent la grande variété entre le français et le chinois : la catégorie grammaticale, la présence de déterminants, la variation des têtes lexicales, sont tous des critères essentiels pour décrire et étudier la composition des chaînes de référence dans les deux langues. L'usage de chaque forme est conditionné par le contexte phrastique et discursif de chaque langue. Les points communs et les différences que nous avons constatés sont d'une importance capitale ; il conviendra d'en tenir compte au moment de leur application à notre corpus de textes spécifiques en français et en chinois. Dans la *Partie II* (*Chapitre 3*, *Chapitre 4* et *Chapitre 5*), nous appliquerons notre propre définition de chaîne de référence afin de constituer un corpus franco-chinois annoté d'expressions référentielles et de chaînes de référence.

Partie II

Problèmes méthodologiques relatifs aux approches contrastives

et à la constitution d'un corpus

Chapitre 3 : Motivations et représentativité en termes du choix d'un corpus

Certaines bases de données linguistiques de grande échelle sont déjà bien construites, comme le CNRTL (*Centre National de Ressources Textuelles et Lexicales*) et ChCorpus (语料库在线) ; néanmoins, leur moteur de recherche ne permet pas de récupérer les chaînes de référence de manière automatique, d'autant plus qu'il s'agit de corpus monolingues en français ou en chinois sans texte parallèle dans l'autre langue. Les questions méthodologiques liées à la constitution de notre propre corpus ont trait aux aspects suivants :

- le choix des textes du côté linguistique : pour quelle représentativité constituons-nous un corpus pour observer les chaînes de référence en deux langues ?
- le choix des outils pour le traitement des textes : du côté informatique, selon quelles motivations choisissons-nous les textes dans notre corpus bilingue ?
- le traitement des expressions référentielles et des chaînes de référence dans les textes : quels choix d'annotation ferons-nous pour ce corpus, et avec quelles justifications et quelles illustrations (extraits du corpus) ?

Il nous semble trop vaste d'étudier les chaînes de référence de manière générale dans une langue, en raison de leur grande variété selon l'époque (*cf*. Capin, 2014 ; Capin *et al*., 2021) et selon le genre de discours (*cf*. Schnedecker & Longo, 2012 ; Longo & Todirascu, 2012, 2014 ; Schnedecker, 2021). Il paraît tout aussi ambitieux d'étudier les chaînes de référence dans plusieurs textes du même genre de discours, en raison de leur nombre illimité en vue de la constitution d'un corpus. Par conséquent, nous avons opté pour les phénomènes des chaînes de référence dans un corpus bilingue constitué d'extraits issus d'un seul roman-fleuve, celui de Romain Rolland (1866-1944) intitulé *Jean-Christophe* (1904-1912) et sa traduction en chinois (1936-1941) par Lei Fu (傅雷, 1908-1966), dans le but d'observer les chaînes de référence de la manière la plus précise possible. Pourquoi la constitution de ce corpus nous permettra-t-elle d'examiner les chaînes de référence dans les deux langues ? Nous répondrons à cette question en plusieurs étapes.

Tout d'abord, trois questions se posent relativement au choix des textes en tant que corpus :

1) Quelle est la place de *Jean-Christophe* dans le domaine de la littérature ?

2) Quels sont les avantages de *Jean-Christophe* pour une étude linguistique contrastive franco-chinoise et quelles sont les caractéristiques linguistiques de la traduction de ce roman-fleuve par Lei Fu ?

3) Quelles sont les spécificités de notre corpus dans la typologie des corpus multilingues ?

Ensuite, les trois questions suivantes portent sur la caractérisation de *Jean-Christophe* :

4) Quelles sont les caractéristiques de *Jean-Christophe* en tant que *roman-fleuve* ?

5) Compte tenu de la typologie textuelle, comment définir *Jean-Christophe* comme type de texte ?

6) Quelles sont les caractéristiques des traductions littéraires relativement aux expressions référentielles ?

Nous essayons, dans cette partie, d'apporter à ces questions des réponses, desquelles émergent des perspectives relativement à la caractérisation des chaînes de référence dans ce roman-fleuve en deux langues : quelles hypothèses pouvons-nous formuler sur les chaînes de référence dans *Jean-Christophe* en français et en chinois ?

3.1. Caractéristiques de *Jean-Christophe* selon les aspects littéraires et culturels

Pour des lecteurs francophones, *Jean-Christophe* ne fait pas partie de la littérature la plus lue et la plus populaire. Ce livre a pourtant des influences non négligeables dans le domaine de la littérature, de la traduction, de la culture et de la linguistique. Nous aborderons dans cette section les motivations selon lesquelles nous choisissons les extraits de ce livre en français et en chinois comme corpus de comparaison des chaînes de référence.

3.1.1. *Jean-Christophe* et la traduction de Lei Fu dans le domaine de la littérature

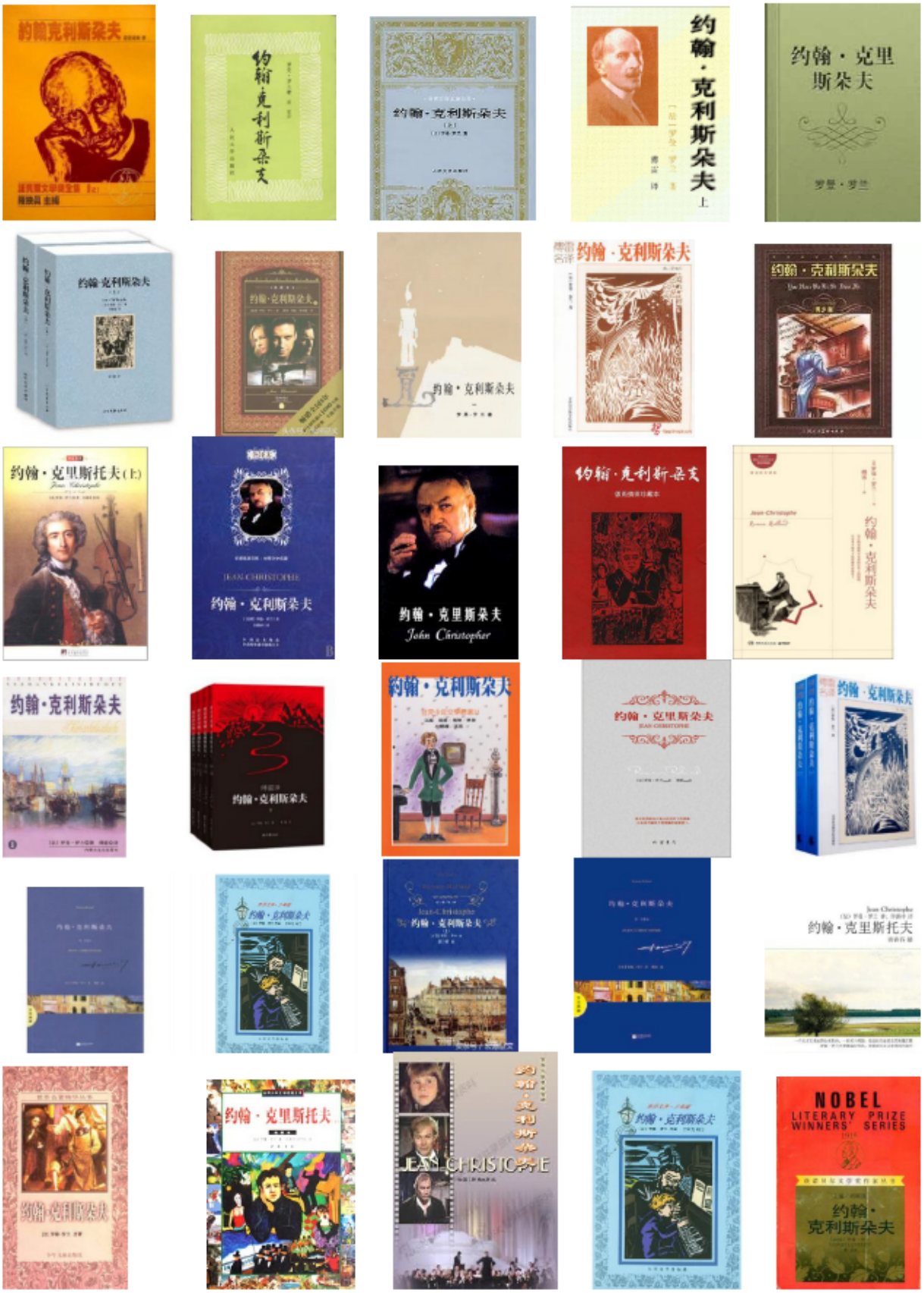
La qualité d'une œuvre littéraire est toujours difficile à évaluer en raison des commentaires souvent très diversifiés, selon les groupes de lecteurs chez les amateurs et selon les jurys des prix chez les professionnels. Pourtant, les prix littéraires peuvent, d'après nous, être considérés comme la garantie d'une certaine qualité d'écriture ou d'une écriture reconnue dans le domaine de la littérature, malgré les polémiques et les discussions que suscite chaque remise de prix. C'est aussi en raison de ces discussions que les œuvres littéraires lauréates sont sélectionnées pour être traduites dans les langues du monde entier. En ce sens, *Jean-Christophe* peut être considéré comme l'une des œuvres littéraires représentatives de la littérature française dans un pays étranger comme la Chine. En effet, ce roman-fleuve a reçu le Prix Femina en 1905, et, grâce à cette œuvre, Romain Rolland a reçu le prix Nobel de littérature de 1915 en guise d' « hommage rendu au grand idéalisme de ses écrits ainsi qu'à la sympathie et à la vérité avec lesquelles il a peint les différents types humains[1] ». Ces prix littéraires peuvent être considérés en quelque sorte comme gages de l'excellence du français de ce roman-fleuve et de l'importance de sa valeur littéraire pour la littérature française et européenne (*cf*. Brancy, 2005, 2014).

Jean-Christophe est reconnu comme chef-d'œuvre de la littérature française en Chine pour des raisons historiques et linguistiques relatives au domaine de la traduction franco-chinoise. Ce roman-fleuve a été traduit et publié pour la première fois par Lei Fu en chinois. Cette traduction se limitait au premier volume de *Jean-Christophe* en 1937 (*cf*. Fu, 傅雷, 2017 [1957] : VI) ; celle du roman-fleuve entier est parue la première fois en 1937 - 1941 et la seconde en 1952 - 1953. Les années 1930 et 1940 représentent une période importante pour la littérature française en Chine car c'est pendant celles-ci qu'un grand nombre d'œuvres littéraires françaises ont été traduites en chinois (Wu, 吴岳添, 2004). Lei Fu est l'un des

[1] ***France Archives*** : https://francearchives.fr/fr/commemo/recueil-2015/38720.

traducteurs les plus connus et représentatifs. Il doit cette notoriété avant tout aux études littéraires et esthétiques qu'il a faites à Paris de 1928 à 1931, au moment où peu de Chinois allaient en France étudier la littérature, et qui lui ont permis de bien connaître la culture et la civilisation de l'Hexagone. À partir de 1933 et jusqu'en 1947, il a traduit et publié 22 œuvres d'une dizaine d'auteurs français (dont les plus connus sont Romain Rolland et Honoré de Balzac), ce qui représente au total 3 683 000 caractères chinois. Toutes ses traductions ont été regroupées et publiées, après sa mort, dans un recueil de 15 volumes, sous le titre de *Traductions de Lei Fu*, recueil qui demeure aujourd'hui encore le plus important en matière de traduction d'œuvres littéraires françaises. La qualité de la traduction d'œuvres littéraires est aussi difficile à apprécier que celle de ces œuvres elles-mêmes (*cf*. Gallagher, 2007). Or, on peut aborder le problème sous plusieurs angles : le traducteur reste-t-il fidèle au texte original ? Restitue-t-il fidèlement le style de l'écrivain ? Adapte-t-il la langue source à la langue cible ? La traduction d'une œuvre littéraire peut-elle être considérée comme une réécriture ? Telles sont les questions que suscite le travail de traduction. Il existe de nos jours une trentaine de versions de *Jean-Christophe* en Chine, comme le montre la figure (3.1).

Figure (3.1) : Versions de *Jean-Christophe* en chinois (non exhaustives)



D'un côté, le nombre de publications et de rééditions reflète la popularité de ce roman-fleuve auprès des lecteurs chinois ; d'un autre côté, toutes ces publications ne sont pas l'œuvre d'un seul et même traducteur : outre celles de Lei Fu évoquées plus haut, deux autres traductions ont été effectuées à partir de la version de Fu par Hulin Han (韩沪麟, 1939 -...) et par Yuanchong Xu (许渊冲, 1921 - 2021).

Les premières traductions de Han et de Xu ont été publiées toutes les deux en 2000 par les Éditions Yilin (译林出版社) et les Éditions littéraires du Hunan (湖南文艺出版社). Elles constituent donc des versions plus récentes par rapport à celle de Lei Fu ; cependant, la version de Fu est toujours la plus lue et la plus rééditée de nos jours. Le fait qu'il y ait trois différentes traductions de *Jean-Christophe* permet de les comparer et de porter sur elles des appréciations relatives, ce qui évite les jugements absolus, qui sont souvent moins convaincants. Ces trois traductions sont utilisées comme corpus dans le cadre des études de lexicométrie de traductologie de Miao (2008, 2011), Gerdes & Miao (2008) et de Miao & Salem (2015). En comparant les trois versions grâce à la méthode lexicométrique, Miao (2011) signale que celle de Fu est la plus fidèle au texte original français et que le chinois utilisé est le meilleur parce qu'il est fidèle aux habitudes langagières des lecteurs chinois. Par ailleurs, les traductions de Lei Fu sont considérées comme de véritables chefs-d'œuvre littéraires en chinois, qui leur valent une réputation nationale selon de nombreux commentaires littéraires, comme notamment ceux de Chen (陈思和, 2008) et de Liu (柳鸣九, 2008, 2019).

Pour commémorer la remarquable contribution de Fu dans le domaine de la traduction, le *Prix Fu Lei de la traduction et de l'édition* a été créé en 2009 par l'ambassade de France en Chine. Ce prix est le plus prestigieux et le plus réputé dans le domaine de la traduction littéraire franco-chinoise en Chine de nos jours.

Le travail de Fu ne se limite pas à la traduction d'œuvres littéraires ; il a également créé son propre style et ses propres théories en matière de traductologie, comme l'indiquent Xiao & Xu (肖红 & 许均, 2002) et Miao (2011). Selon lui, une traduction idéale est censée provoquer les mêmes effets sur un lecteur étranger que ceux de l'original sur un lecteur natif. Une bonne traduction doit donc non seulement transmettre fidèlement le contenu du texte dans la langue cible mais également garder autant que possible les caractéristiques stylistiques de l'auteur de la langue source. Une telle traduction est censée répondre aux besoins des lecteurs natifs en permettant la transmission de l'esprit original de l'auteur du texte (Fu, 傅雷, 2017 [1951] : 1).

> *翻译以效果而论，应当像临画一样，所求的不在形似而在神似。(Fu, 傅雷, 2017 [1951] : 1)*
>
> *Au sujet des effets, les textes traduits doivent être comme un simulacre de dessin ; ce que l'on cherche à réaliser n'est pas de simuler leur forme, mais de simuler leur esprit. (nous traduisons)*

Jean-Christophe a été écrit par Romain Rolland de 1904 à 1912, et Lei Fu en a fait la traduction de 1937 à 1941, soit environ trente ans plus tard. Cette durée peut paraître minime dans toute l'évolution diachronique[2] ; la langue de l'écrivain français et celle du traducteur chinois peuvent donc être considérées comme appartenant à peu près à la même époque. La dernière grande évolution linguistique de la langue chinoise a eu lieu durant la première moitié du 20ème siècle. Après le *Mouvement de la Nouvelle Littérature* (新文化运动) autour de l'année

[2] L'évolution diachronique peut entraîner des changements dans les chaînes de référence, voir Capin (2014) et Capin ***et al.*** (2021).

1919, le chinois est passé de la phase « ancienne » à la phase « contemporaine » (*cf*. Luo & Lv, 罗常培 & 吕叔湘, 1956 ; Wei, 魏慧萍, 2008 ; Zhang, 张卫中, 2009). Le chinois de Fu a été produit juste après cette réforme linguistique et culturelle. Ainsi, le langage du roman-fleuve peut être considérée en quelque sorte comme la marque du début du chinois contemporain. Bien qu'il existe d'éventuelles variations du chinois de l'époque du roman-fleuve jusqu'à présent, Li & Thompson (1979) rappellent que la syntaxe du chinois n'a pas connu de modification majeure depuis le 16ème siècle jusqu'à nos jours.

3.1.2. Corpus parallèle *souple*

Les corpus bilingues peuvent se constituer de textes provenant du même domaine et relevant du même registre de langue, à savoir les *corpus comparables*, ou de textes en relation de traduction mutuelle, à savoir les *corpus parallèles* (*cf*. Teubert, 2001 ; Habert, 2005 ; Zimina-Poirot, 2005 ; Zheng & Chang, 郑茗元 & 常霜林, 2015 ; Guo, 郭纯洁, 2015). Ces deux types de corpus fournissent des possibilités de comparaison entre les textes ayant certaines ressemblances dans diverses langues ; ils sont ainsi utilisés par différents auteurs. Par exemple, plusieurs études contrastives sur les chaînes de référence se fondent sur des corpus comparables et parallèles dans le numéro 82 de *Travaux de Linguistique : Approches contrastives des chaînes de référence* (2021). Dvorak (2021) utilise un roman tchèque avec sa traduction en français pour constituer un corpus parallèle ; Baumer *et al*. (2021) rassemblent les *nouvelles d'actualité* extraites des presses journalistiques en français, en anglais et en allemand dans une *perspective comparative* (2021 : 19) ; Guo (2021) constitue un corpus comparable avec des textes encyclopédiques relatifs aux mêmes thèmes en français et en chinois.

Par rapport à un corpus comparable, un corpus parallèle présente des avantages quant à la comparaison des mêmes phénomènes dans plusieurs langues (Granger, 2003).

> *Translation corpora are an ideal resource for establishing equivalence between languages since they convey the same semantic content. (Granger, 2003 : 19)*

Ce qui est intéressant pour notre étude, c'est le fait qu'un corpus parallèle peut permettre d'observer les corrélations entre les usages, similaires ou différents, des expressions des mêmes référents. Il se peut que les expressions anaphoriques et coréférentielles équivalentes au niveau sémantique ne se correspondent pas complètement dans une langue ou dans l'autre, compte tenu de leur catégorie grammaticale, de leur fonction syntaxique et d'autres facteurs grammaticaux et discursifs.

Zimina-Poirot (2005) appelle également *corpus parallèle* un « bi-texte », à savoir la réunion d' « un texte original en langue source et de sa traduction en langue cible appuyée sur un système de mise en correspondance interactive entre les segments équivalents des deux volets bilingues » (2005 : 234). Selon cette définition, un roman-fleuve et sa traduction constituent sans nul doute un corpus parallèle parce que les textes sont en relation mutuelle.

Pourtant, en tant que traducteur et lettré, Fu a développé sa propre conception du travail de traduction littéraire. Selon lui, cette dernière est différente de celle des textes appartenant à d'autres genres de discours. Une traduction littéraire peut être même parfois

considérée comme une réécriture en raison d'un langage « fluent », « riche » et « variable[3] » (Fu, 傅雷, cité par Xiao & Xu, 肖红 & 许均, 2002) recréé et utilisé par le traducteur. Ce langage traduit est censé à la fois respecter la grammaire du chinois et emprunter la structure des phrases longues et complexes de la langue étrangère source afin d'en illustrer les particularités.

> *当中文不足以表达原作信息时，傅雷又提出在翻译中要"采用西洋长句"，"创造中国语言，加多句法变化"，使中国读者体会到异国文化的魅力.* **(Fu, *傅雷*, *cité par Xiao & Xu, 肖红 & 许均, 2002)***
>
> ***Quand le chinois ne suffit pas pour transmettre les informations des textes originaux, Fu propose d' « utiliser les phrases longues des langues occidentales » et de « créer un langage chinois en ajoutant des variations syntaxiques » pour que les lecteurs chinois fassent l'expérience du charme des cultures étrangères. (nous traduisons)***

Tout choix de traduction reflète en quelque sorte les conditions syntaxiques, sémantiques, contextuelles et pragmatiques d'une langue à l'autre selon les expériences et la compréhension du traducteur. Les caractéristiques du langage traduit dans les œuvres littéraires ne devraient pas être ignorées ou généralisées, confondues avec le langage dans les autres genres de discours (*cf.* Xie, 谢天振, 1998).

Par conséquent, les textes littéraires originaux et les textes traduits constituent un corpus parallèle. Ils présentent une relation plus forte que celle des textes d'un corpus comparable, qui est certes composé de textes d'un même genre, mais sans relation traductionnelle. Nous nous permettons d'appeler notre corpus *Jean-Christophe* un *corpus parallèle souple*, car il comporte des traductions moins fidèles que celles d'autres genres auxquels appartiennent les textes scientifiques et juridiques, par exemple.

3.1.3. Influence des contextes culturels sur les expressions référentielles

La traduction des expressions référentielles dans les œuvres littéraires peut être très compliquée, parce que la langue source peut comprendre des éléments intransmissibles dans la culture et la civilisation de la communauté de la langue cible. Observons par exemple les noms propres de personnages historiques dans l'extrait suivant issu du roman-fleuve *Jean-Christophe*, où le grand-père raconte des histoires de héros à son petit-fils. Les noms propres sont annotés par les crochets et sont mis en gras :

[1] Il parlait de [**Régulus**]$_{npr\text{-}1}$, d'[**Arminius**]$_{npr\text{-}2}$, des chasseurs de [**Lützow**]$_{npr\text{-}3}$, de [**Kœrner**]$_{npr\text{-}4}$ et de [**Frédéric Stabs**]$_{npr\text{-}5}$, celui qui voulait tuer [**l'empereur Napoléon**]$_{npr\text{-}6}$. Sa figure rayonnait, en rapportant des traits d'héroïsme inouïs.

Ces noms propres, présents comme tels dans le roman-fleuve, ne peuvent pas être compris de la même manière et avec la même facilité par un lecteur français et par un lecteur chinois. Si « l'empereur Napoléon » est plus connu du lectorat chinois que les autres personnages en question, il n'en demeure pas moins que la traduction directe de « Régulus », « Arminius », « Lützow », « Koerner » et « Frédéric Stabs », entraînerait probablement des

[3] En chinois 语言流畅，用字丰富，色彩多变.

difficultés de compréhension pour les lecteurs chinois, en raison de l'absence de connaissances culturelles et historiques que les lecteurs français possèdent ou sont censés posséder.

Les référents, ainsi que leurs expressions, peuvent être à la fois familiers pour une communauté, et exotiques pour une autre, ce qui peut provoquer d'éventuelles difficultés de traduction. Ainsi, nous verrons que la traduction de Lei Fu s'efforce de remédier à ce type de problème, dont l'origine nous semble se trouver dans la nature des référents. En d'autres termes, le roman-fleuve original combine des référents du monde réel et des référents du monde fictif. Le monde réel est celui où vivent les lecteurs d'une langue, à savoir les lecteurs francophones qui habitent dans des régions francophones ; le monde fictif est créé par le roman. Une communion est censée s'établir entre Romain Rolland et ses lecteurs francophones. Dès que le roman-fleuve est traduit dans une autre langue, à savoir le chinois, le monde fictif demeure le même, mais le monde réel change : c'est désormais celui où se trouvent les lecteurs sinophones, qui vivent dans les régions sinophones, et avec lesquels Lei Fu doit établir aussi une communion. Comme nous l'avons illustré plus haut dans l'extrait [1], avec les noms de figures historiques, ce changement de communauté de lecteurs peut influencer les expressions référentielles dans le roman-fleuve traduit.

De la même manière pour le français et pour le chinois, nous proposons de classer les expressions en trois groupes en fonction de la familiarité des référents dans les communautés de lecteurs :

- les noms propres issus des connaissances encyclopédiques des lecteurs de chaque langue ;
- les noms communs ;
- les noms propres issus du roman-fleuve.

Le premier groupe concerne les noms propres issus des connaissances encyclopédiques renvoyant aux référents les plus familiers pour les lecteurs de chaque langue, comme « le Rhin » (y compris les dérivés des noms propres, comme « rhénan »), « Cologne », « Mannheim » et « Moselle » dans les extraits [2] et [3] pour les lecteurs francophones. Il s'agit là du groupe de référents le plus « exotique » pour les lecteurs chinois.

[2] Les Krafft étaient sans fortune, mais considérés dans [**la petite ville rhénane**], où le vieux s'était établi, il y avait presque un demi-siècle. Ils étaient musiciens de père en fils et connus des musiciens de tout le pays, entre [**Cologne**] et [**Mannheim**].

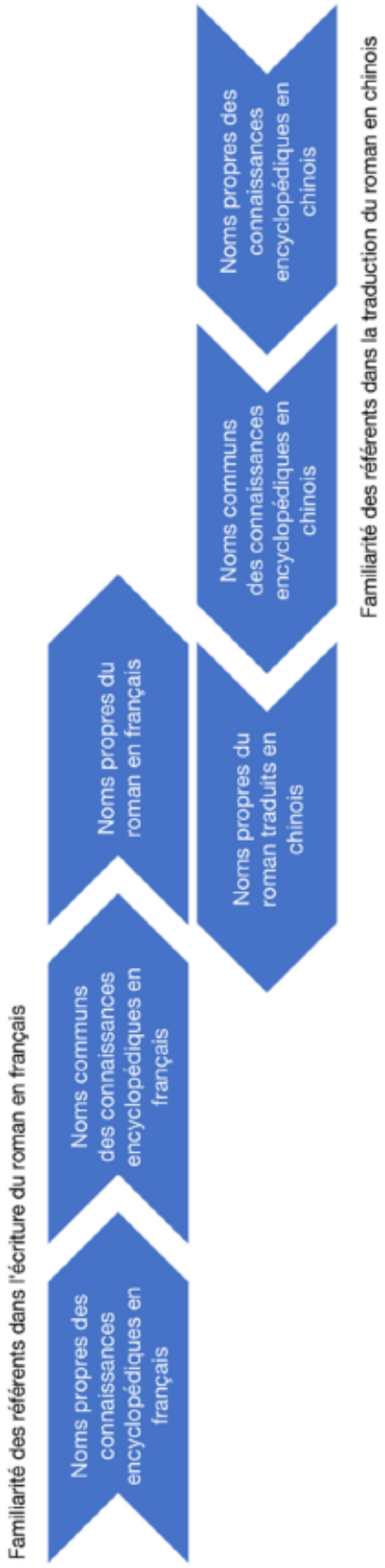
[3] Il avait un robuste appétit ; et s'il reprochait à Melchior de boire, lui-même buvait sec. Il avait une prédilection pour les vins blancs de la [**Moselle**].

Le deuxième groupe se rapporte aux noms communs issus des connaissances encyclopédiques renvoyant aux référents de familiarité intermédiaire pour les lecteurs d'une langue et non pour ceux de l'autre. Ils n'existent souvent que dans une seule communauté ; les lecteurs de l'autre communauté les connaissent peu ou ils ne leur sont pas familiers. Par exemple, dans l'extrait [4], les expressions en gras comme le « grand-duc », les « chérubins joufflus » et les « lauriers fleuris », issues du roman-fleuve *Jean-Christophe*, renvoient à des personnes ou à des objets familiers aux lecteurs français, alors que ces référents semblent exotiques aux lecteurs chinois.

[4] Enfin Christophe fut admis à approcher son héros. C'était le jour du concert. Toute la ville était là. [**Le grand-duc**] et sa cour occupaient la grande loge princière, surmontée d'une couronne, que tenaient dans les airs, avec des ronds de jambes, [**deux chérubins joufflus**]. Le théâtre avait un aspect de gala. La scène était ornée de branches de chêne et de [**lauriers fleuris**].

Le troisième groupe est relatif aux noms propres issus du roman-fleuve renvoyant aux référents les moins familiers pour les lecteurs des deux langues parce que ce sont des référents créés qui n'existent que dans le cadre fictif du roman-fleuve, comme « Christophe » et « Louisa ». Nous proposons de montrer, dans la figure (3.2) plus bas, les écarts de contenus entre les textes originaux et les textes traduits. Les flèches du haut montrent le degré de familiarité des référents dans l'écriture du roman-fleuve pour les lecteurs français et celles du bas pour les lecteurs chinois.

Figure (3.2) : Familiarité des référents dans le roman-fleuve en français et dans sa traduction en chinois



Cette figure nous permet de formuler une question : existe-t-il des corrélations entre la familiarité des référents selon les différentes communautés de lecteurs et la forme des expressions référentielles ? En d'autres termes, il se pourrait que plus les référents sont familiers aux lecteurs, moins les formes d'expressions sont complexes[4]. Cette hypothèse générale se subdivise en trois sous-hypothèses.

Premièrement, les deux premiers groupes de noms, dans les flèches en haut, à savoir « les noms propres / communs aux connaissances encyclopédiques des lecteurs français », qui sont censés être familiers aux lecteurs français, pourraient probablement être traduits par le truchement d'autres noms ou par les mêmes noms avec des modifieurs explicatifs, comme le montre le texte [5] en chinois, traduction de l'exemple [1] : les noms propres français sans modifieur sont repris par des noms propres chinois modifiés :

- « Régulus » est traduit par 'ancien Rome régisseur Régulus' *gu-Luoma-zhizheng-Leigulusi* 古罗马执政雷古卢斯 ;
- « Arminuis » est traduit par 'av-J.C. germain chef Arminuis' *gongyuanqian-de-Riermanzu-shouling-Aminusi* 公元前的日耳曼族首领阿米奴斯 ;
- « Lützow » est traduit par 'allemand général Lützow' *Deguo-dajiang-Lvzuofu* 德国大将吕佐夫 ;
- « Koerner » est traduit par 'poète Koerner' *poète-Kena* 诗人科纳.

[5] 他 讲着 **[古 罗马 执政 雷古卢斯]**，**[公元前 的 日耳曼族 首领 阿米奴斯]**，也
ta jiangzhe gu Luoma zhizheng Leigulusi gongyuanqian de Rierman shouling Aminusi ye
3P parlant ancien Rome régisseur Régulus av-J.C. DE germain chef Arminius aussi
npr-mod *npr-mod*

讲到 **[德国 大将 吕佐夫] 的 轻骑兵** — **[诗人 科纳]**，和 那 个 想 刺死 拿破仑
jiangdao Deguo dajiang Lvzuofu de qingqibing shiren Kena he na ge xiang cisi Napolun
parlant allemand général Lützow DE chasseur poète Koerner et ce-là CL vouloir tuer Napoléon
npr-mod *npr-mod*

皇帝 的 史太勃斯。他 眉飞色舞，讲着 那些 空前绝后的 壮烈的 事迹。
huangdi de Shitaibosi ta meifeisewu jiangzhe naxie kongqianjuehoude zhuangliede shiji
empereur DE Stabs 3P rayonnant rapportant ces-là inouïs héroïque exploit

Deuxièmement, les groupes de noms situés au centre de la figure (3.2), dans les deux flèches en haut et en bas, à savoir « les noms propres du roman-fleuve en français / chinois », sont transcrits selon leur prononciation par les mêmes noms propres créés dans le monde fictif du roman-fleuve, comme « Christophe » et 克利斯朵夫 (transcription en pinyin[5] : ke-li-si-duo-fu), « Louisa » et 鲁意莎 (transcription en pinyin : lu-yi-sha). D'un côté, ces transcriptions phonétiques respectent la prononciation originaire des noms propres en français ; de l'autre, les caractères choisis par le traducteur correspondent aux protocoles de dénomination des personnes en chinois. Autrement dit, la prononciation de « lou » dans « Louisa » peut correspondre à plusieurs caractères, comme *lu* 鲁, *lu* 卤, *lu* 录, *lu* 路 et *lu* 鹿, mais le traducteur choisit *lu* 鲁 qui est utilisé particulièrement comme un nom de famille. Ce type d'adaptation n'existe pas dans la traduction de tous les noms propres parce que toutes les prononciations n'ont pas comme équivalent un nom de famille en chinois.

[4] Cette hypothèse se fonde sur des théories cognitives et pragmatiques, comme celles de Clark & Marshall (1981, cité par Ariel, 1988 : 66) et d'Ariel (1990). Par exemple, les expressions des formes les plus complexes (*i.e.* descriptions définies et noms propres) sont accessibles par des connaissances encyclopédiques.

[5] Le pinyin du chinois est la transcription internationale de la prononciation du chinois mandarin. Il permet de connaître la prononciation des caractères chinois grâce à l'alphabet latin.

Enfin, troisièmement, les deux derniers groupes de noms dans les flèches en bas, à savoir « les noms communs / propres des connaissances encyclopédiques des lecteurs chinois », qui renvoient aux référents familiers des lecteurs chinois, sont en principe peu nombreux voire absents dans les textes traduits, parce que ce type de référents sont absents des textes originaux. Et s'il y en a quelques-uns, on peut se demander à quels moments et sous quelles formes ils pourraient se présenter.

Le travail de traduction et les stratégies du traducteur en fonction des différents types de lecteurs nous amènent à considérer comme probable le fait que les expressions référentielles dans les deux langues peuvent présenter des variations en raison des écarts culturels entre les communautés de locuteurs. Cela peut ainsi entraîner des variations dans les chaînes de référence des deux langues.

En tant que locutrice sinophone, nous n'avons pas la même perception linguistique de la langue française qu'un francophone. Un corpus bilingue de textes écrits par un romancier francophone et un traducteur sinophone fournit un exemple de locuteur natif de chaque langue. Ce sont ces textes authentiques produits par des locuteurs natifs qui constituent les sources de notre observation des chaînes de référence ainsi que des expressions référentielles.

3.2. Caractéristiques de *Jean-Christophe* en termes de chaînes de référence

Parallèlement à notre hypothèse portant sur les différences des expressions référentielles entre les textes originaux français et les textes traduits en chinois provoquées par les écarts culturels des lecteurs, nous nous intéressons aussi aux autres facteurs discursifs qui peuvent influencer la constitution des chaînes de référence dans *Jean-Christophe* :

- au vu du genre discursif, comment les chaînes de référence se réalisent-elles en français et en chinois dans ce roman-fleuve ? Quels sont les points communs et les différences potentiels de ce genre en français et en chinois ?
- au vu du type textuel, comment les chaînes de référence se réalisent-elles dans différents types de séquences textuelles ? Par exemple, par un contraste entre les séquences narratives à strictement parler et les séquences dialogales, ou par un contraste entre les séquences mono-référentielles et les séquences multi-référentielles ?
- au vu de la longueur, comment les chaînes de référence se constituent-elles dans des textes longs en français et en chinois ?

Les études antérieures mettent en lumière certaines propriétés distinctives des expressions référentielles et des chaînes de référence dans différents genres discursifs et dans différents types de séquences textuelles. En nous basant sur ces études, nous tenterons d'étudier la réalisation des chaînes dans le roman-fleuve, *Jean-Christophe*, en français et en chinois.

3.2.1. Définition et caractéristiques d'un *roman-fleuve*

Le genre est l'un des facteurs qui reflètent en quelque sorte les caractéristiques générales d'une classe de discours. Les discours d'un même genre partagent des normes linguistiques communes et des compositions homogènes. Il est ainsi primordial de définir *Jean-Christophe* en termes de genre de discours. Nous constituons un corpus avec *Jean-Christophe* non seulement en raison de sa valeur et de sa notoriété dans le domaine de la littérature française en Chine et de ses caractéristiques dans la traduction, mais également compte tenu de sa place représentative parmi les *romans-fleuve*, terme créé par Romain Rolland, à la faveur d'une phrase présente dans *Jean-Christophe* : c'est ce roman qui marque la naissance du genre du *roman-fleuve*.

> *Jean-Christophe m'est apparu comme un fleuve, je l'ai dit dès les premières pages. (Rolland, « Aux amis de Jean-Christophe », Jean-Christophe, Paris, Albin Michel, 1966 : 1600, cité par Leblond, 2016 : 55)*

La composition des chaînes de référence s'avère déjà dépendre fortement de leur contexte discursif, comme le signalent, entre autres, Chastain (1975), Schnedecker & Longo (2012), Longo & Todirascu (2013) et Schnedecker (2014, 2017, 2021), avec divers genres de textes comme corpus d'études. Dans le *Dictionnaire d'Analyse du Discours*, Charaudeau & Maingueneau (2002) indiquent que la distinction des genres est réalisée « à travers les caractéristiques formelles des textes et en rassemblant les marques les plus récurrentes » (2002 : 279). Les critères qui servent à classer les genres sont nombreux. Le genre romanesque peut se distinguer des autres genres littéraires par sa forme d'écriture, son contenu, son rapport avec la réalité et son organisation énonciative :

> *Dans la tradition littéraire, les genres sont censés permettre de repérer et de classer les différents textes littéraires, qu'ils appartiennent à la prose ou à la poésie. […]*
>
> - *Des critères à la fois de composition, de forme et de contenu qui distinguent les genres : poésie, théâtre, **roman**, essai. […]*
> - *Des critères renvoyant à différentes façons de concevoir la représentation de la réalité, définies à travers des textes ou des manifestes ayant pour fonction de fonder des écoles, et qui ont correspondu à des périodes historiques : les genres romantique, réaliste, naturaliste, surréaliste, etc.*
> - *Des critères renvoyant à la structure des textes et particulièrement à leur organisation énonciative : le fantastique, l'autobiographie, le roman historique, etc.*
>
> *(Charaudeau & Maingueneau, 2002 : 278, nous soulignons)*

En tant que sous-type du genre romanesque, le roman-fleuve partage avec le roman des caractéristiques discursives qui sont relativement plus étudiées.

Selon plusieurs définitions, comme celle de la *Grammaire Méthodique du Français*, celle provenant d'une série d'études d'Adam (1987, 1997, 2001, 2006, 2015 [2011], 2019) et celle de Xu (许余龙, 2010), le roman fait partie des « genres littéraires » (Adam, 2006 : 5). Le langage littéraire se trouve dans la typologie des langages de Wellek (cité par Todorov, 1978 : 20) avec quelques spécificités : il peut tolérer quelques ambiguïtés sémantiques par rapport au langage scientifique ; il peut également se distancier plus ou moins de la situation d'énonciation spécifique par rapport au langage courant. Ce sont donc les particularités

potentielles du langage du roman-fleuve qui font aussi partie du *genre littéraire* (*cf*. Ducrot & Todorov, 1972). Ce dernier présente trois spécificités :

> *(i) Des catégories telles que « l'action » ou « les personnages » se justifient dans la description des textes littéraires ; qu'elles sont présentes obligatoirement ou non ;*
>
> *(ii) Chacune de ces catégories peut être spécifiée par un nombre fini de propriétés qui s'organisent en structure : par exemple que les personnages peuvent être soit « dignes », soit « de basse condition » ;*
>
> *(iii) Les catégories dégagées et leur variété : on examinera tous les types de personnages (ou d'action, etc.) ; la présence de l'un ou de l'autre, dans telle ou telle combinaison, donnera des types littéraires.*
>
> *(Ducrot & Todorov, 1972 : 194)*

Le roman-fleuve est un texte narratif. Adam (1994 : 32) propose de distinguer trois états dans une narration : « avant » (*état initial*), « pendant » (*transformation*) et « après » (*état final*)[6]. Un texte narratif est souvent caractérisé par l'apparition de personnages avec des indications de changements de situation. D'après Yue (乐耀, 2010 : 346), ces personnages jouent le rôle de topique, dont les expressions constituent des chaînes de référents topicaux ; les changements de situation sont marqués par les expressions spatio-temporelles : leurs premières occurrences peuvent marquer le début d'un état ; leurs reprises peuvent garantir la progression d'événements dans le même cadre spatio-temporel. Selon ces caractérisations, il est prévisible qu'un texte narratif contiendra des chaînes de personnages et de repères spatio-temporels, dont la composition peut refléter les différents états d'une narration.

Schnedecker (2005 : 113) qualifie les chaînes de référence dans les romans (et dans les portraits journalistiques) comme hétérogènes pour la variation des maillons, opposées aux chaînes homogènes qui ne contiennent que massivement des maillons en nom propre et en pronom. Baumer (2012b) précise le même fait : les chaînes de référence dans les portraits journalistiques sont caractérisées par les noms propres et les anaphores nominales (par rapport à ceux des chaînes de référence dans les nouvelles).

Les romans-fleuve sont plus longs que les autres textes littéraires (poèmes, essais, contes, nouvelles). Ils sont également plus longs que d'autres textes narratifs, comme les faits divers, les contes de fées et les fables. En tant que genre spécifique, le roman-fleuve ne se distingue pas seulement par sa longueur significative, ; son langage présente d'autres particularités dont certaines peuvent être impliquées et reflétées par les chaînes de référence. Par ailleurs, en raison de leurs dimensions dites *démesurées* (Leblond, 2016), les romans-fleuve sont souvent rédigés et édités en plusieurs volumes dans les publications en papier ; mais cette subdivision en volumes n'empêche pas la progression de la narration et la continuité de l'histoire dans le temps, qui semble s'écouler, justement, comme un fleuve. *Jean-Christophe* est ainsi publié en dix volumes, équivalant à plus de 500 000 mots en français. Il nous paraît donc intéressant et pertinent d'étudier la manière dont les chaînes de référence permettent de garantir la continuité référentielle de ce type de texte excessivement long.

---

[6] Une définition homologue est proposée par Yue (乐耀, 2010 : 343) : le discours narratif peut être considéré comme une « séquence descriptive d'événement » (***the described event group***). Un discours narratif se subdivise en trois sous-séquences : « séquence initiative » (***group-initial***), « séquence intermédiaire » (***group-medial***) et « séquence finale » (***group-final***).

Vu les ressemblances potentielles entre les langages narratif, littéraire, romanesque et du roman-fleuve, nous résumons quelques caractéristiques susceptibles d'appartenir aux chaînes de référence dans *Jean-Christophe* :

- elles sont très longues ;
- elles présentent certaines ambiguïtés référentielles éventuelles ;
- elles sont influencées par des facteurs culturels et linguistiques selon la communauté des lecteurs ;
- elles renvoient à toutes sortes de référents : personnages, objets, événements, actions, etc. ;
- leurs composants expriment des particularités objectives (réelles) ou subjectives (fictives) de référent selon les caractéristiques qualitatives et selon le point de vue du narrateur ;
- leurs composants sont susceptibles d'être très variables pour que ce langage soit plus esthétique et riche dans chaque langue.

Selon la définition d'Adam (1994), les textes sont constitués par des unités élémentaires que sont les séquences. Une *séquence textuelle* est « un réseau relationnel hiérarchique » (194 : 111), soit une unité textuelle plus grande qu'une proposition et plus petite qu'un texte ; une séquence est aussi « une entité relativement autonome », c'est-à-dire que la relation entre deux séquences peut être relativement faible par rapport à celle entre deux propositions. Les séquences textuelles peuvent être, toujours d'après Adam (2015 [2011]), *narratives*, *descriptives*, *dialogales*, voire parfois *argumentatives* et *explicatives*, et leur totalité constitue la *structure compositionnelle des textes*. Xu (许余龙, 2010) utilise le terme *type de texte* (*text type*) dans le même sens que celui d'Adam.

> *(The text type is) the way text is made to respond to its context and to display a particular text type focus such as exposition, argumentation and instruction. (Xu, 许余龙, 2010 : 367)*

En réalité, un discours est rarement composé d'un seul type de séquence textuelle. Le roman-fleuve est un genre de discours complexe et poly-typologique : il est impossible à définir et difficile à caractériser en raison de la grande diversité des types de séquences textuelles[7]. C'est pourquoi Adam (2015 [2011]) propose de décrire un discours par son type de séquence dite *principale* ou *dominante* pour signifier qu'une grande partie de ce discours représente les caractéristiques de ce type de séquence textuelle, en en incluant probablement d'autres en proportion moindre.

> *Dire qu'un texte peut avoir une dominante d'un type ou d'un autre n'a rien à voir avec l'hypothèse trop générale de l'existence des types de textes. Il existe bien des genres qui ont des dominantes préférentielles, c'est tout ce que l'on peut affirmer, en raison de l'extrême hétérogénéité constitutive des textes authentiques. Les trop ambitieuses typologies de textes sont passées à côté de la complexité des agencements séquentiels dont elles ont négligé l'importance au profit de ce qui n'était qu'un effet de dominante. (Adam, 2015 [2011] : 221)*

[7] Une autre difficulté liée à la caractérisation des romans est signalée par Xu (徐赳赳, 2014) et consiste dans le fait que le style d'un auteur peut influencer fortement la manière de déployer la narration ; partant, les chaînes des mêmes référents peuvent avoir des différences selon les auteurs. Ce point de vue de Xu peut être confirmé dans le cas des narrations orales ; mais il nous semble difficile à vérifier à l'écrit dans le cas des romans puisque ceux-ci sont uniques. Aucun corpus ne permet donc de corroborer cette supposition.

On décrit les romans comme un genre narratif parce que ce sont les séquences narratives qui dominent en eux et qui occupent une grande proportion dans l'intégralité d'une œuvre romanesque ; néanmoins, ils peuvent comprendre sans aucun doute plusieurs autres types de séquences textuelles. Cette idée provient de l'« effet de dominante » d'Adam (2015 [2011]) :

> *[…] La catégorisation globale d'un texte résulte d'un* ***effet de dominante*** *: le tout textuel est, dans sa globalité et sous forme de résumé, caractérisable comme plutôt narratif, argumentatif, explicatif, descriptif ou dialogal. (Adam, 2015 [2011] : 219)*

Ainsi, il est assez aisé de comprendre que les romans-fleuve sont définis comme genre narratif parce qu'ils présentent des séquences narratives dominantes ; mais cela n'empêche pas qu'ils contiennent également des séquences dialogales, comme c'est le cas de *Jean-Christophe*, selon les termes d'Adam. Ces séquences sont visiblement marquées par les discours rapportés directs, comme le montre la partie soulignée dans l'extrait [6]. Ces dialogues ont lieu entre les personnages ; ils créent un environnement discursif totalement différent de celui des séquences narratives à proprement parler.

[6] L'Excellence déplia soigneusement le papier, le lut, regarda Christophe, le laissa patauger dans ses explications, puis l'interrompit, et dit, avec un éclair malicieux dans les yeux :
– C'est bien, monsieur Krafft. La demande est accordée.
De la main, il lui donna congé et se replongea dans ses écritures.
Christophe sortit, consterné.
– Sans rancune, Christophe ! lui dit cordialement l'employé, quand l'enfant repassa par le bureau.
Christophe se laissa prendre et secouer la main, sans oser lever les yeux.

Par conséquent, pour analyser les chaînes de référence dans *Jean-Christophe*, notre première approche porte sur la globalité du livre : comment se réalisent-elles dans le genre roman-fleuve ? La seconde concerne les parties du livre : comment se réalisent-elles dans différents types de séquences textuelles ? Pour répondre à ces questions, il nous semble possible et adéquat de découper l'œuvre en plus petits segments qualifiés de « séquences narratives et/ou dialogales ». Ce découpage du roman-fleuve en segments ainsi que leurs définitions seront présentés dans la section *5.1.2. Étape 2 : partition des textes en épisode* du *Chapitre 5*.

3.2.2. Chaînes de référence dans un roman-fleuve en français *vs* en chinois

Les études contrastives sur les chaînes de référence entre différents genres en français se sont beaucoup développées ces dernières années grâce au projet MC4[8] (2012 - 2016) et au projet DEMOCRAT[9] (2016 - 2020). De nombreux linguistes francophones ont effectué des observations sur les chaînes dans des corpus constitués de divers genres discursifs, comme notamment Schnedecker & Longo (2012), Longo & Todirascu (2013) et Schnedecker (2014,

[8] Le projet MC4 : ***Modélisation Contrastive et Computationnelle des Chaînes de Coréférence*** (http://www.lattice.cnrs.fr/projets/projets-passes/projet-peps-mc4-modelisation-contrastive-et-computationnelle-des-chaines-de-coreference/, dernière consultation en juillet 2019).

[9] Le projet DEMOCRAT de l'ANR (ANR-15-CE-38-0008) : ***DEscription et MOdélisation des Chaînes de Référence : outils pour l'Annotation de corpus (en diachronie et en langues comparées) et le Traitement automatique***.

2017, 2021). Cependant, cette question contrastive est peu discutée avec des textes en chinois, comme nous l'avons noté dans le *Chapitre 1*. Nous présenterons cet état de l'art en deux temps : d'une part, nous synthétiserons les études contrastives du français pour examiner les influences du genre discursif sur les chaînes de référence ; d'autre part, nous aborderons quelques études dédiées aux chaînes de référence en chinois, dans lesquelles les auteurs ont pris, intentionnellement ou non, des exemples de textes narratifs pour leurs analyses.

3.2.2.1. Chaînes de référence dans divers genres en français

Avant de comparer les chaînes de plusieurs genres, certains auteurs cherchent à caractériser, tout d'abord, les chaînes de référence dans un seul genre de discours en français. Par exemple, Jenkins (2002) observe les chaînes de référence dans les portraits journalistiques en français : les maillons nominaux présentent des désignations très variables ayant pour but de décrire les caractéristiques qualificatives des personnes, comme le montre la liste suivante :

> ***désignation selon la profession*** (le cinéaste, cet industriel de la confection, madame la maire) ;
>
> ***désignation selon l'origine géographique*** (la belle Scandinave, la Française, la chica madrilène) ;
>
> ***désignation selon l'âge ou le sexe*** (ce jeune homme, cette femme, la jeune femme) ;
>
> ***désignation relationnelle*** (son épouse, la fiancée du Viking, cette mère de deux enfants) ;
>
> ***désignation selon le statut social ou la cause de célébrité*** (la numéro un du tennis français, l'ex-prince du rap, la star) ;
>
> ***désignation selon l'orientation politique ou religieuse*** (cette démocrate-chrétienne, cette catholique qui déteste qu'on mette en avant ses convictions religieuses) ;
>
> ***désignation selon une qualité spécifique*** (la bête politique en devenir, un tel chauffeur de salle, ce colosse sympathique).
>
> *(Jenkins, 2002 : 514 - 515)*

Cette variété est reliée à une autre notion de calcul, celui du *coefficient de stabilité* de la chaîne. La *stabilité* est un indice qui permet de mesurer la variation des têtes lexicales des maillons anaphoriques nominaux. Le coefficient de stabilité des maillons est défini pour la première fois par Perret (2000) pour analyser certaines différences textuelles du français des 14ème et 15ème siècles :

> *Pour s'intéresser, dans un premier temps, à la stabilité désignationnelle on proposera le concept de* ***coefficient de stabilité****, obtenu en divisant, pour un référent donné (un personnage), le nombre total d'anaphores nominales par le nombre de désignations différentes. (Par exemple, dans la* Mélusine *de Jean d'Arras, pour la désignation de l'héroïne on rencontre 164 anaphores nominales, et 17 désignations différentes ; le coefficient de stabilité est donc 164 / 17 = 9,64). (Perret, 2000 : 17)*

Cette notion est ensuite définie par Schnedecker (2021) sur la base de la première définition de Perret, comme suit :

(Le coefficient de stabilité) s'obtient en divisant le nombre total d'anaphores par le nombre de désignations différentes. (Schnedecker, 2021 : 260)

Le calcul de cet indice peut donc être ainsi formulé comme suit :

$$Coefficient\ de\ stabilité = \frac{\text{Nombre de maillons nominaux}}{\text{Nombre de têtes lexicales}}$$

Plus le coefficient est élevé, plus la chaîne est stable. Comme le constate Jenkins (2002), le coefficient de stabilité des chaînes dans les portraits journalistiques est relativement bas, indiquant ainsi le fait que la composition des chaînes n'est pas stable. Il est susceptible que les chaînes dans *Jean-Christophe* ne sont pas stables. Mettons à présent en relief les expressions nominales variables de l'extrait [7] issu de notre corpus : 4 maillons nominaux possèdent 4 têtes lexicales différentes : « enfant », « jésus », « poisson d'or » et « masse » ; le coefficient de stabilité de cet extrait de chaîne est donc 1, ce qui montre que cette chaîne n'est pas stable :

[7] [L'**enfant**] *M-nominal-1* n'a pas cette ressource trompeuse. [Sa] première rencontre avec la douleur est plus tragique et plus vraie. Comme [son] être même, elle [lui] semble sans limites ; [il] la sent installée dans [son] sein, assise dans [son] cœur, maîtresse de [sa] chair. Et cela est ainsi : elle n'en sortira plus qu'après [l']avoir rongé.
La mère [le] presse contre elle, avec de petits mots :
« C'est fini, c'est fini, ne pleurons plus, [mon **jésus**] *M-nominal-2*, [mon petit **poisson d'or**] *M-nominal-3* …

[Il] continue toujours [sa] plainte entrecoupée. On dirait que [cette misérable **masse** *M-nominal-4* inconsciente et informe] a le pressentiment de la vie de peines qui [lui] est réservée. Et rien ne peut [l']apaiser…

Nous inspirant de ces remarques sur la variation de désignation et le coefficient de stabilité, nous nous demandons si les composants nominaux dans les chaînes sont variables ou stables dans le roman-fleuve.

Outre l'étude de Jenkins (2002), Schnedecker & Longo (2012) choisissent également de travailler sur un seul genre spécifique en français : elles récapitulent les caractéristiques les plus remarquables des chaînes de référence dans les faits divers. Trois indices sont choisis pour montrer les caractéristiques des chaînes de ce genre : selon le critère du « nombre de chaînes », les chaînes sont décrites comme « nombreuses » et « denses » ; selon le critère de la « longueur des chaînes », elles sont décrites comme « brèves » ; selon le critère de la « composition des chaînes », elles sont décrites comme « hétérogènes » :

> ***Leur nombre** : les faits divers sont, dans leur grande majorité (60 % des cas), pluri-référentiels et instancient, qui plus est, des référents « obligés » comme les forces de l'ordre, de la sécurité et la magistrature ; corollairement, la **densité** des expressions référentielles dans les textes est assez forte ;*
>
> ***La dimension** : les chaînes de référence sont, dans leur ensemble, brèves et comptent en moyenne 3,42 maillons ; cela tient évidemment à la brièveté même du genre ainsi que, dans une moindre mesure, à la pluri-référentialité[10] ;*
>
> ***La composition** : les maillons des chaînes de référence sont hétérogènes du point de vue de leur catégorie grammaticale qui se manifeste dans des proportions inattendues par rapport aux prédictions de certaines théories (*p.e. *l'accessibilité) puisque les formes de moyenne et haute accessibilité prévalent ; pour ce qui concerne les SN, leur tête lexicale est éminemment variable ; en outre, il est apparu que l'hétérogénéité des chaînes de référence était renforcée, pour ainsi*

[10] Voir la définition de ***pluri-référentiel / multi-référentiel*** et son contraire ***mono-référentiel*** dans la suite de cette section.

> *dire, par l'origine des maillons selon qu'ils s'ancrent dans la situation* vs *dans le faisceau de propriétés du référent : cela a des incidences sur leur empan long* vs *étroit selon leur nature.*
>
> *(Schnedecker & Longo, 2012 : 1968 - 1969, nous soulignons)*

Cependant, les caractéristiques des faits divers, ce dernier étant un genre narratif multi-référentiel de petite longueur, ne peuvent pas représenter celles des autres genres narratifs, ou des autres textes courts, ou des autres textes multi-référentiels. Par le biais de la comparaison entre deux ou plusieurs genres, certains autres auteurs mettent en contraste, dans un second temps, les différences entre les genres narratifs et non narratifs en français, dans lesquels les configurations typiques des chaînes de référence sont radicalement différentes. C'est aussi pour cette raison que Longo & Todirascu (2012) et Schnedecker (2014, 2017, 2021) considèrent les chaînes de référence comme un des indices importants pour distinguer et identifier les genres textuels. Longo & Todirascu (2013) comparent les chaînes de référence dans cinq genres narratifs et non narratifs, *i.e.* les articles journalistiques, les éditoriaux, les romans, les lois et les rapports publics. Certaines conclusions portent sur les contrastes des chaînes entre les genres narratifs et non narratifs. Elles peuvent être résumées par les cinq critères suivants :

i) La longueur moyenne des chaînes de référence

La différence de la longueur moyenne des chaînes est évidente entre les genres narratifs et non narratifs : celle des lois est très courte, comptant seulement 3 maillons, à peine de quoi constituer des chaînes ; celle du roman est beaucoup plus longue, comptant le triple de la première.

> *La longueur moyenne des chaînes de référence varie du simple au triple selon le genre textuel. […] Cette différence significative entre la longueur des chaînes de ces deux genres s'explique en ce que les lois européennes font intervenir de nombreux référents, donc que la compétition référentielle est forte. (Longo & Todirascu, 2013 : 150)*

ii) La distance moyenne entre les maillons[11]

La moyenne de la distance inter-maillonnaire présente une grande différence entre les narratifs et les non narratifs : elle n'excède pas une phrase dans les articles journalistiques, alors qu'elle est supérieure à deux phrases entre le deuxième et le troisième maillon dans les rapports publics.

iii) La fréquence[12] des maillons de différentes catégories grammaticales

Chaque genre présente une catégorie grammaticale privilégiée différente : les noms propres occupent 30,8 % des maillons dans les articles journalistiques ; les SN définis occupent une moitié des maillons dans les éditoriaux ; 40 % des maillons sont SN indéfinis dans les lois ; 28,2 % des maillons sont possessifs dans les romans et 33,3 % des maillons sont SN définis

[11] Voir aussi Ariel (1988), qui prend ce critère pour mesurer le degré d'accessibilité des expressions anaphoriques.

[12] Le terme ***fréquence (d'une unité textuelle)*** est défini en linguistique informatique comme « le nombre de ses occurrences dans le corpus. » (Lebart & Salem, 1994 : 313).

dans les rapports publics. Ces catégories concernent la plus grande proportion de maillons par rapport aux autres catégories dans chaque genre.

iv) La catégorie grammaticale des premiers maillons[13]

Les chaînes de chaque genre ne sont pas initiées de la même manière. Les noms propres prennent essentiellement la place de premier maillon dans les articles journalistiques, les SN définis dans les éditoriaux et les SN indéfinis dans les lois et dans les romans.

v) La correspondance entre la première mention et le thème phrastique

Il est constaté que la coïncidence peut exister entre la première expression référentielle et le thème de la phrase. Cette coïncidence concerne 80 % des cas dans les articles journalistiques mais seulement 40 % dans les rapports publics.

Grâce à cette comparaison de Longo & Todirascu (2013), les différences entre les chaînes dans les genres narratifs et non narratifs sont bien démontrées. Par ailleurs, les chaînes de référence entre différents genres narratifs sont aussi variables. Schnedecker (2017) choisit deux genres narratifs courts, à savoir les faits divers et les *incipit* des contes de fées, comme base d'observation. Ces deux genres sont comparables parce que « dans les deux cas, ce sont des textes écrits dans le but de raconter des histoires, produits à l'intention de destinataires inconnus du producteur du texte, absents de la situation de production » (2017 : 57). Selon l'auteure, les différences de chaînes se manifestent selon cinq aspects : formel, syntaxique, thématique, lexical et fonctionnel. Trois points communs sont constatés au niveau de l'aspect formel :

> *La fonction grammaticale dominante des maillons : sujet ;*
>
> *Les catégories lexicales rares : nom propre ;*
>
> *Les catégories lexicales fréquentes : nom.*
>
> *(Schnedecker, 2017 : 67)*

Au niveau des quatre autres aspects, les chaînes de référence dans les deux genres se distinguent bien, d'après Schnedecker (2017), selon les onze critères suivants :

i) La densité des expressions référentielles

La densité est un indice qui peut refléter le nombre d'expressions référentielles dans une même longueur de texte :

> *(La densité référentielle est) Le ratio du nombre d'expressions référentielles et du nombre de mots d'un texte. (Schnedecker, 2021 : 262)*

Selon cette définition, la densité référentielle de l'extrait suivant est 0,37, à savoir le nombre d'expressions référentielles (37 expressions) divisé par celui des mots de l'extrait (100 mots) :

[13] *Cf.* aussi Manuélian & Schnedecker (2021).

[8] Dans [**le lit**], près de [**la mère**], [**l'enfant**] s'agitait de nouveau. [**Une souffrance inconnue**]
ER-1 ER-2 ER-3 ER-4

montait [**du fond de** [**son**] **être**]]. [**Il**] se raidit contre [**elle**]. [**Il**] tordit [[**son**] **corps**], [**il**] serra
ER-5 ER-6 ER-7 ER-8 ER-9 ER-10 ER-11 ER-12 ER-13

[**les poings**], [**il**] fronça [**les sourcils**]. [**La douleur**] grandissait, tranquille, sûre de [[**sa**] **force**].
ER-14 ER-15 ER-16 ER-17 ER-18 ER-19

[**Il**] ne savait pas [**ce qu'**[**elle**] **était**], ni jusqu'[**où** [**elle**] **allait**]. [**Elle**] [**lui**] paraissait immense, et ne
ER-20 ER-21 ER-22 ER-23 ER-24 ER-25 ER-26

devoir jamais prendre fin. Et [**il**] se mit à crier lamentablement. [[**Sa**] **mère**] [**le**] caressa avec
ER-27 ER-28 ER-29 ER-30

[**de douces mains**]. Déjà [**la souffrance**] devenait moins aiguë. Mais [**il**] continuait de pleurer ;
ER-31 ER-32 ER-33

car [**il**] [**la**] sentait toujours près de [**lui**], en [**lui**].
ER-34 ER-35 ER-36 ER-37

La densité référentielle s'avère plus élevée dans les contes de fées (30 %) que dans les faits divers (9 %), et c'est aussi le cas pour le nombre d'expressions référentielles : 1 / 3 dans les contes de fées, plus élevé que 1 / 11 dans les faits divers.

ii) Le nombre de chaînes de référence

Les chaînes sont beaucoup moins nombreuses dans les contes de fées (23 chaînes) que dans les faits divers (89 chaînes), dans des textes de longueurs similaires.

iii) La longueur moyenne des chaînes de référence

La longueur moyenne des chaînes des contes de fées (7 maillons) atteint presque le double de celle des faits divers (3,42 maillons).

iv) La chaîne de référence la plus longue

En mesure du nombre de maillons, la chaîne la plus longue des contes de fées comporte 15 maillons, équivalant seulement à environ la moitié de celle des faits divers, avec 27 maillons.

v) Le coefficient de stabilité

Ce coefficient est plus élevé dans les contes de fées que dans les faits divers, à savoir que les chaînes sont plus stables et les maillons nominaux moins variables dans les premiers que dans les seconds.

vi) Les composants et les composants privilégiés

Les chaînes tendent à être homogènes dans les contes de fées mais elles sont hétérogènes dans les faits divers.

La composition privilégiée n'est pas similaire dans les deux genres : ce sont les pronoms pour les contes de fées et les SN pour les faits divers.

vii) La coïncidence entre la chaîne et le thème du texte

La coïncidence est toujours positive dans les contes de fées (c'est-à-dire que la chaîne coïncide toujours avec le thème du texte) ; elle ne l'est pas dans les faits divers (à savoir que la chaîne ne coïncide pas tout le temps avec le thème d'un texte de faits divers).

viii) La modalité privilégiée d'introduction du référent principal

Le référent principal est introduit par la seule modalité privilégiée « il était une fois un SN1 qui V SN2 » dans les contes de fées, alors qu'il existe quatre patrons privilégiés dans les faits divers.

ix) Le type d'expansion privilégié

La plus grande proportion d'expansions des SN sont les adjectifs dans les contes de fées ; les expansions les plus privilégiées dans les faits divers sont les adjectifs et les compléments du nom.

x) Le mode de cohabitation des chaînes de référence

Le mode de cohabitation est un critère défini par Schnedecker (1992 : 183) pour démontrer la performance des chaînes *équilibrées*, *i.e.* les chaînes de longueur similaire, dans les textes. Les trois modes les plus typiques sont l'*alternance*, le *parallélisme* et l'*entrecroisement* :

> *Des CR (chaînes de référence)* ***alternent*** *à partir du moment où une CR1 substantielle s'interrompt pour laisser place à une CR2 qui se maintient elle aussi sur une certaine distance.*
>
> *On parlera de CR* ***parallèles*** *quand des référents introduits initialement dans des rôles syntaxico-sémantiques distincts sont réinstanciés dans la suite du texte, en conservant les fonctions qui leur ont été attribuées au départ.*
>
> *Les CR que nous appelons « **entrecroisées** » s'apparentent aux précédentes dans la mesure où elles activent simultanément et successivement deux référents. En revanche, elles s'en distinguent parce que le statut syntaxico-sémantique des référents se modifie d'une phrase à l'autre. Un référent en situation de saillance (*i.e. *dénommé, en première mention, agent) dans une phrase donnée s'affaiblit dans la phrase suivante, en perdant justement (toutes ou certaines) de ces caractéristiques alors qu'un second référent, initialement moins saillant, émerge à son tour. Puis, le premier revient en position dominante et ainsi de suite. (Schnedecker, 1992 : 184 -185, nous soulignons)*

Trois autres modes complètent ce cadre de modalités de cohabitation : la *succession*, la *dérivation*, et la *partition / fusion* :

> *On parlera de* **succession** *tout simplement quand une chaîne s'arrête et qu'il lui en succède une autre, avec, pour effet, la disparition du premier référent.*
>
> *Lorsque le premier maillon de certaines chaîne est lui-même une anaphore, dont la source s'inscrit dans une autre chaîne, on parlera de* **dérivation**.
>
> *La* **partition** *et la* **fusion** *procèdent par deux mouvements inverses. La partition extrait d'un ensemble préalablement posé des référents individués. (…) La fusion*

construit, à partir de référents isolés ou de groupes de référents, un ensemble plus conséquent. (Schnedecker, 2021 : 100-101, nous soulignons)

La succession et l'entrecroisement sont les deux modes *autonomes*, à savoir que les chaînes en question se développent de manière indépendante ; la dérivation, la partition / fusion et le parallélisme sont les trois modes *non autonomes*, à savoir que les chaînes en question sont marquées par des relations avec d'autres chaînes dans le même texte.

D'après ces définitions, dans les contes de fées, le mode de cohabitation est *variable selon les personnages*, alors que les deux modes privilégiés dans les faits divers sont l'entrecroisement et la succession.

En guise de bilan, nous pouvons dire que les deux études de Longo & Todirascu (2013) et de Schnedecker (2017) pour le français montrent, d'un côté, que les différences des chaînes de référence existent bel et bien entre les genres narratifs et non narratifs, et qu'elles sont reflétées par les critères susmentionnés, et, d'un autre côté, que les différences se situent également entre les différents genres narratifs. Nous chercherons donc à adapter ces critères de comparaison pour montrer les spécificités des chaînes de référence dans le roman-fleuve par rapport aux autres genres narratifs.

3.2.2.2. Chaînes pronominales et expressions anaphoriques zéro en chinois

En ce qui concerne le chinois, comme nous l'avons vu dans la section *1.5. Chaîne pronominale de Xu (徐赳赳, 1990)* du *Chapitre 1*, l'étude de Xu (徐赳赳, 1990) porte sur un corpus composé d'articles journalistiques narratifs ; mais aucun auteur n'a comparé les chaînes de référence entre différents genres dans cette langue. D'après son observation basée sur 85 articles journalistiques narratifs, la structure générale des chaînes pronominales dans un texte est synthétisée dans la figure (1.2) du *Chapitre 1* : un texte contient plusieurs noms (ou groupes nominaux), chaque nom est suivi d'un certain nombre de pronoms. Les séquences *nom – pronom – pronom* sont appelées *chaînes pronominales*. Littéralement, les chaînes pronominales sont des séquences ainsi nommées en raison des pronoms utilisés comme composants privilégiés, initiées par un SN ; les maillons zéro et ceux d'autres formes ne sont pas compris dans cette configuration. Outre celle de Xu, de nombreuses autres études s'intéressent aux expressions anaphoriques chinoises dans le cadre phrastique ; elles nous paraissent importantes puisque les maillons des chaînes peuvent également être des expressions anaphoriques.

Par rapport aux SN anaphoriques et aux pronoms anaphoriques, les zéros anaphoriques sont particulièrement nombreux et importants en chinois, comme cela est affirmé notamment par Li & Thompson (1979), Chen (陈平, 1987), Ariel (1990), Huang (2000), Huang & Zhang (黄娴 & 张克亮, 2009) et Jiang & Cheng (蒋云 & 承平, 2014). Par l'approche comparative, Huang (2000 : 59) définit la langue chinoise comme *langue ellipsant le pronom* (*Pronoun-dropping Language*) parce que les pronoms en position de sujet ne sont pas obligatoirement exprimés. Par conséquent, les zéros sont censés être beaucoup plus fréquents en chinois que dans les langues qui n'ellipsent pas le pronom, comme le français et l'anglais.

Les zéros anaphoriques sont aussi beaucoup plus souvent utilisés en chinois qu'en russe (*cf.* Cui, 崔卫, 1998 ; Li, 李春蓉, 2014).

Selon l'approche syntaxique, lorsque les zéros se situent à la place de sujet de phrase, ils peuvent aussi être appelés *sujets zéro* (*cf.* Huang, 2000 : 50 ; Hou & Sun, 侯敏 & 孙建军, 2004 : 15 ; Landragin, 2011 : 72, entre autres), ce pour mettre l'accent sur la fonction syntaxique de ce vide[14] ; selon l'approche grammaticale, ces vides sont dits *pronoms zéro* (*cf.* Li & Thompson, 1979 : 312 ; Ariel, 1990, 2014 ; Chen & Ng, 2013 : 1360), ce qui sous-entend qu'ils remplacent un pronom explicite[15] ; Xu (徐赳赳, 1993 : 332) les appelle tout simplement *zéros*, pour désigner, de manière générale, tous les vides anaphoriques.

Li & Thompson (1979) analysent les zéros en chinois par une approche textuelle et constatent qu'il y a peu de contraintes pour l'usage de zéros dans les textes narratifs. Selon leur observation, les zéros en chinois sont régulièrement utilisés aux endroits équivalents aux pronoms explicites en anglais. Il leur paraît plus important d'observer les conditions d'utilisation des pronoms explicites en chinois au lieu de celles des zéros, puisque ces derniers sont systématiquement utilisés comme sujet des propositions dans tout le texte :

> *Zero-anaphora is widespread in Chinese discours, so much so, in fact, that the non-occurrence of anaphoric arguments in discourse must be regarded as the normal, unmarked situation? Thus, it is the OCCURRENCE of pronouns in Chinese discourse that must be explained. (Li & Thompson, 1979 : 322)*

Le résultat de leur enquête[16] montre que les zéros sont privilégiés par rapport aux pronoms explicites dans la plupart des propositions de l'extrait [9] issu du roman chinois *Chronique indiscrète des mandarins* (*Rulin-waishi* « 儒林外史 »). Le pourcentage d'enquêtés qui choisissent le pronom explicite est marqué à chaque place vide. Par exemple, pour la « place-1 », 6 % des enquêtés choisissent de mettre un pronom explicite au lieu d'y laisser un vide, cela étant le choix pour la plupart des places, sauf deux, « 3 » et « 18 », qui peuvent valablement être remplies par un prénom explicite pour plus d'une moitié des enquêtés (respectivement 76 % et 56 %) ; mais ces pourcentages sont encore loin de former un consensus. Toutes les autres places montrent un pourcentage assez bas pour les pronoms explicites : seulement entre 14 % (places « 7 » et « 19 ») et 0 % (places « 5 », « 9 », « 12 », « 13 », « 14 » et « 15 ») des enquêtés choisissent d'utiliser un pronom explicite plutôt que de laisser vides les places.

[9]	这	**[王冕]**	天性	聪明，	**[Ø]**	年纪	不	满	二十岁，	**[Ø]**	就	把	那
	zhe	*Wangmian*	*tianxing*	*congming*		*nianji*	*bu*	*man*	*ershisui*		*jiu*	*ba*	*na*
	this	Wang-Mian	nature	smart		age	not	exceed	twenty-year		already	obj. marker	that
		premier-maillon			*place-1-6%*					*place-2-2%*			

[14] La ***Grammaire méthodique du français*** (Riegel ***et al.***, 2016 [1994] : 249) classe le cas du ***zéro*** comme ***ellipse du sujet*** : « une condition nécessaire mais non suffisante pour que le sujet d'une phrase ne soit pas exprimé est que le contexte linguistique ou la situation de communication permettent de le restituer. » À l'écrit, les principaux cas concernent les ***phrases impératives***, les ***constructions infinitives subordonnées*** et les ***phrases juxtaposées ou coordonnées***.

[15] Chen & Ng (2013 : 1360) définissent le ***pronom zéro*** comme « a gap in a sentence that is found when a phonetically null form is used to refer to a real-world entity ».

[16] L'enquête a été réalisée avec 50 sinophones. Les enquêtés avaient le choix entre un pronom explicite et un zéro pour remplir ou non les endroits vides dans l'extrait [8].

天文， 地理， 经史 上的 大 学问 无 一 不 贯通， 但 [Ø] 性情
tianwen *dili* *jingshi* *shangde* *da* *xuewen* *wu* *yi* *bu* *guantong* *dan* *xingqing*
astronomy geography classics in great knowledge not a-bit not master however personality
place-3-76%

不 同， [Ø] 既不 求 官爵， [Ø] 又 不 交那 朋友， [Ø] 终日 闭 户
bu *tong* *jibu* *qiu* *guanjue* *you* *bu* *jiaona* *pengyou* *zhongri* *bi* *hu*
not similar only-not seek officialdom also not make friend all-day close door
place-4-2% *place-5-0%* *place-6-2%*

读书， [Ø] 又 在 楚辞图 上 看见 画 的 屈原 衣冠， [Ø] 便 自 造
dushu *you* *zai* *Chucitu* *shang* *kanjian* *hua* *de* *Quyuan* *yiguan* *bian* *zi* *zao*
study also at Chu-Ci-Tu in see paint rel. cl. marker Qu-Yuan attire then self make
place-7-14% *place-8-8%*

一顶 极 高的 帽子， 一件 极 阔的 衣服， [Ø] 遇着 花明柳媚的 时节， [Ø]
yiding *ji* *gaode* *maozi* *yijian* *ji* *kuode* *yifu* *yudao* *huamingliumeide* *shijie*
one very tall hat one very loose garment chance-upon balmy days
place-9-0% *place-10-12%*

把 一乘 牛车， 载了 母亲， [Ø] 便 带了 高 帽， [Ø] 穿了
ba *yisheng* *niuche* *zaile* *muqin* *bian* *daile* *gao* *mao* *chuanle*
take a cattle-cart bring-aspect mother then wear-aspect tall hat put on-aspect
place-11-10% *place-12-0%*

阔 衣， [Ø] 支着 鞭子， [Ø] 口里 唱着 歌曲， [Ø] 在 乡村 镇
kuo *yi* *zhizhe* *bianzi* *kouli* *changzhe* *gequ* *zai* *xiangcun* *zhen*
loose garment hold-aspect whip month-in sing-aspect song at village town
place-13-0% *place-14-0%* *place-15-0%*

上 以及 湖边， 到处 玩耍， [Ø] 惹得 乡下 孩子们 三 五 成 群 跟着
shang *yiji* *hubian* *daochu* *wanshua* *rede* *xiangxia* *haizimen* *san* *wu* *cheng* *qun* *genzhe*
in and lake-side everywhere cruised about make country children three five from group follow
place-16-4%

[Ø] 笑， [Ø] 也 不 放在意下， [Ø] 如此 就 生活 安定 快乐 健康。
xiao *ye* *bu* *fangzaiyixia* *ruci* *jiu* *shenghuo* *anding* *kuaile* *jiankang*
laugh even so not mind thus then live stable happy healthy
place-17-4% *place-18-56%* *place-19-14%*

(Wu, 吴敬梓, *Chronique indiscrète des mandarins*, « 儒林外史 », 1803, cité par Li & Thompson, 1979 : 323-324, nous annotons)

La comparaison avec l'extrait parallèle [10] en anglais montre que le privilège du zéro a disparu et est remplacé par les pronoms explicites anglais. Pour la plupart des cas du même type de cet extrait, les lecteurs natifs sinophones préfèrent laisser les places vides et les lecteurs natifs anglophones préfèrent les remplir par des pronoms explicites.

[10]This [Wang-Mian] *premier-maillon* was gifted. [He] *place-1* was not more than twenty years of age. [He] *place-2* had already mastered everything in astronomy, geography, and classics. However, [he] *place-3* had a different personality. Not only did [he] *place-4* not seek officialdom, [he] *place-5* also did not make friends. [He] *place-6* studied at home all day. [He] *place-7* also saw the attire of Qu-Yuan painted in Chu-Ci-Tu. [He] *place-8* then made a very tall hat, and a very loose garment. When [he] *place-9* chanced upon balmy days, [he] *place-10* took a cattle cart and brought his mother. [He] *place-11* then wore a tall hat. [He] *place-12* put on a loose garment. [He] *place-13* held a whip. [He] *place-14* sang songs. [He] *place-15* cruised about the villages, towns, lake-side and everywhere. [He] *place-16* made the children in the country follow him about and laugh in groups of three and five. [He] *place-17* did not even mind. Thus, [he] *place-18* lived a stable, happy and healthy life.

(Li & Thompson, 1979 : 323-324, nous annotons)

Le même type d'enquête étant réalisé avec d'autres extraits du même roman, Li & Thompson (1979) résument trois situations dans les textes en chinois où les sinophones

préfèrent utiliser un pronom explicite plutôt qu'un zéro ; il s'agit de trois moments où la continuité référentielle est interrompue par certains facteurs textuels :

> *I. Conjoinability appears to be greatly impaired when the clauses involved contain a switch from background information to foreground information or vice versa.*
>
> *II. Conjoinability between two clauses is greatly impaired when the second clause is marked with adverbial expressions such as time phrases or contrastive morphemes like* however, but *etc.*
>
> *III. Conjoinability appears to be impaired when the two clauses in question constitute different turns, that is, are spoken by different participants, as opposed to being part of the same turn, spoken by one participant.*
>
> *(Li & Thompson, 1979 : 331-333)*

En ce qui concerne le premier principe, le zéro est privilégié en chinois si les phrases suivantes tendent à mettre le référent comme information de contexte et à mettre le focus sur d'autres référents, comme le montre l'extrait [11] : aucun enquêté ne choisit de remplir la place par un pronom explicite. Dans le cas contraire, le pronom explicite chinois est privilégié, c'est-à-dire si la phrase suivante continue toujours à parler du même référent à la place du focus de texte, comme le montre l'extrait [12] : 75 % des enquêtés considèrent plus pertinent de remplir la place par un pronom explicite au lieu de la laisser vide. Contrairement à l'anglais, c'est toujours le pronom explicite qui est privilégié dans les deux cas.

[11]外边 走进 一个 人 来，[…] [Ø] 和 [众人] 拱一拱手 。
waibian zoujin yige ren lai *he zhongren gongyigongshou*
outside walk-in a person come toward everyone salute with hand
place-1-0% *autre-rét*

'From outside entered a man, [(he)] saluted everyone with his hands.'
place-1

[12]外边 走进 一个 人 来，[…] [Ø] 姓 夏。
waibian zoujin yige ren lai *xing Xia*
outside walk-in a person come surname Xia
place-2-75%

'From outside entered a man, [(he)] was surnamed Xia.'
place-2

(Li & Thompson, 1979 : 325-326, nous annotons)

Le deuxième principe peut servir à expliquer le choix du pronom explicite à la « place-3 » de l'exemple [9], en l'occurrence après le connecteur contrastif 'mais' *dan* 但 qui s'avère avoir le plus haut pourcentage d'enquêtés pour l'usage du pronom explicite parmi toutes les places marquées dans l'extrait.

À propos du troisième principe, le pronom explicite tend à être privilégié dans le dialogue, et le zéro dans le monologue et la narration. Dans le cas du dialogue, 58 % des enquêtés sinophones considèrent qu'il est pertinent d'employer un pronom à la « place-1 » de l'exemple [13], alors qu'aucun enquêté n'utilise de pronom dans le cas de la narration à la « place-2 » de l'exemple [14].

[13]A : 张三 油头粉脸。
Zhangsan youtoufenlian
Zhang-San slick-looking
B : [Ø] 是 一个 小丑。
Is a clown
place-1-58%

'A : Zhang-San is slick-looking. B : [He] is a clown.'
place-1

[14]张三　　油头粉脸，　[Ø]　　是　一个　小丑。
Zhangsan　*youtoufenlian*　　*shi*　*yige*　*xiaochou*
Zhang-San　slick-looking　　is　a　clown
place-2-0%

'Zhang-San is slick-looking, [(he)] is a clown.'
place-2

(Li & Thompson, 1979 : 333)

En ce qui concerne la comparaison des genres, Jiang & Cheng (蒋云 & 承平, 2014) constatent que les zéros ne sont pas utilisés de la même manière dans les textes scientifiques, journalistiques et narratifs : les zéros sont plus récurrents dans les narratifs que dans les scientifiques et les journalistiques ; pour les premiers, ces zéros occupent le plus souvent la fonction de sujet de phrase et le rôle de topique du discours.

Sur la base de ces études et de ces remarques sur les conditions des zéros en chinois, nous formulons l'hypothèse selon laquelle on peut prévoir qu'il y aura un écart important de zéros entre le français et le chinois. Il nous semble intéressant et pertinent de voir non seulement quels écarts pourraient être constatés de manière quantitative, mais également si les conditions similaires d'utilisation des zéros en chinois existent aussi pour le français. Existe-t-il d'autres conditions qui favorisent l'emploi du pronom au lieu de laisser une place vide dans le discours ? Nous aborderons cette question par le truchement de chiffres et d'exemples issus de notre corpus dans la *Partie III* du présent travail.

3.2.3. Chaînes de référence au vu des séquences textuelles

Les textes narratifs longs contiennent souvent des structures textuelles plus complexes que les courts, comme c'est le cas du roman-fleuve en termes de séquences textuelles comme composants du genre. Les séquences textuelles du roman-fleuve ont des spécificités bien distinctes. Nous constatons ces séquences dans *Jean-Christophe* par le moyen de deux approches. La première s'inscrit dans la typologie textuelle d'Adam (1987, 2015 [2011]). Selon les analyses textuelles de cet auteur, un texte peut contenir diverses séquences, dont les sept principales sont les *narratives*, les *injonctives-instructionnelles*, les *descriptives*, les *argumentatives*, les *explicatives-expositives*, les *dialogales-conversationnelles* et les *poétiques-autotéliques*. Tous les types de séquences textuelles n'existent pas dans les romans-fleuve et tous les types existants n'occupent pas des proportions identiques. D'après notre observation, deux types de séquences sont les principales dans *Jean-Christophe* : les *narratives* et les *dialogales*. Les chaînes de référence seront examinées selon ces deux types de séquences. La seconde approche est proposée par Schnedecker (1992), qui prend en compte le nombre de référents saillants comme critère de distinction des situations textuelles. Nous verrons que l'auteure a constaté que les chaînes de référence ne se réalisent pas de la même manière dans une situation mono-référentielle et multi-référentielle.

3.2.3.1. Séquences narratives et dialogales dans le genre narratif

Selon la définition d'Adam (1987), les séquences narratives présentent des caractéristiques qui ressemblent beaucoup à celles du genre narratif : les séquences narratives

se fondent sur un *ordre chronologique*, à savoir que le texte doit comporter les différentes étapes d'un événement : début, pendant, et fin.

> *La séquentialité narrative correspond à une structure hiérarchique globale qui confère aux différents événements (même si leur ordre chronologique est déconstruit en surface) une certaine valeur différentielle : orientation situation initiale (avant le procès) – complication (début du procès) – action ou évaluation (pendant) – résolution (fin du procès) – moral ou situation finale (après le procès). (Adam, 1987 : 61)*

Le roman-fleuve fait partie des textes appartenant au genre narratif ; cependant, malgré cette étiquette de narratif au sens large, tous les segments d'un roman-fleuve ne correspondent pas aux caractéristiques des séquences narratives au sens strict. En ce qui concerne *Jean-Christophe*, la majeure partie de ce roman-fleuve est constituée de séquences narratives ; le reste concerne des séquences dialogales. Cette constitution composite permet, d'une part, de distinguer les caractéristiques des chaînes de référence dans le genre narratif au sens large et dans les séquences narratives à proprement parler, et, d'autre part, d'observer les chaînes de référence dans les trois types de séquences textuelles, à savoir les narratives, les dialogales et les autotéliques. Il ne fait aucun doute, d'après les définitions des deux séquences textuelles élémentaires, que le roman-fleuve contient des chaînes de référence potentiellement différentes dans chaque type de séquence. Il est donc important de ne pas prendre ces chaînes de séquences narratives et de séquences dialogales pour celles du genre narratif au sens large.

Contrairement au contexte linguistique des référents repérés dans les textes narratifs, les référents mentionnés dans les conversations sont souvent *présents*, perceptibles directement par les participants. Ce type de référents sont facilement exprimés par des formes pronominales comme *il* et *ça*, avec un geste de pointer ou tout simplement un regard (*cf.* Kleiber, 1990). Pourtant, une conversation dans un roman-fleuve se distingue largement d'une vraie conversation orale. Faute de situation d'énonciation proprement dite, le contexte est purement linguistique. Le narrateur est censé fournir un contexte imitant celui de l'oral et les informations nécessaires pour que les référents, dans la situation linguistique des personnages du roman-fleuve, soient perceptibles (*cf.* Schnedecker, 2021 : 157). Un énoncé du type [15] se passe probablement à l'oral avec un référent facile à identifier dans la situation d'énonciation, sans nécessité de s'exprimer avec des formes pleines.

[15]Mais [**il**] est fou !

(Kleiber, 1990 : 29)

Il nous semble que ce type de formes pronominales existent également dans les dialogues du roman-fleuve, mais sans doute sous certaines conditions : pour désigner les référents par des formes pronominales, ceux-ci doivent être déjà mentionnés dans le contexte précédant le dialogue.

À notre connaissance, les études antérieures s'intéressent aux expressions coréférentielles et anaphoriques des deux langues, le français et le chinois, dans les conversations, et non dans les textes sous forme de dialogues. Pour les séquences dialogales en français, Corblin (2005) expose, de manière comparative, les caractéristiques des chaînes de référence dans la narration et les dialogues. L'auteur remarque, malgré l'absence d'observation sur un corpus, que l'usage des noms complets et des pronoms dans les chaînes sont radicalement différents dans les deux types de textes : les conversations favorisent

l'emploi des noms propres, alors que les désignateurs dans les autres types de discours s'appuient sur les relations de familiarité.

> *La conversation est régie par des maximes sur l'usage des désignateurs dont l'effet est de maintenir l'identification des communautés épistémiques concrètes, réelles, qui rassemblent les interlocuteurs.*
>
> *On a opposé ici à la* conversation *la notion de* discours à interlocuteur générique *(le plus souvent monologique, souvent écrit, fictionnel ou non). Par définition, ce type de discours n'est pas contraint par les relations de familiarité effectives entre locuteur récepteur et objet. Il ne relie pas un locuteur et un récepteur ancrés et définis par un ensemble spécifié de communautés épistémiques. Il s'agit d'un discours dans lequel un locuteur exprime ses connaissances, croyances et fictions pour un interlocuteur générique. Ce type de discours n'est pas contraint par l'existence de communautés épistémiques réelles associant un « nom propre » à chaque triplet locuteur-récepteur-référent. Il repose sur la simulation de relations de familiarité, ce qui ouvre la possibilité de jouer des relations différentes au cours d'un seul et même discours et d'une même chaîne de référence. (Corblin, 2005 : 15)*

L'étude de Debras & Beaupoil-Hourdel (2019) porte également sur la construction des chaînes de référence pendant les interactions orales. D'après leur observation, ces dernières sont souvent accompagnées des gestes des locuteurs qui servent à « impliquer une réduction ou au contraire une expansion des formes réitérées » (2019 : 20).

L'établissement de la continuité référentielle dans les conversations en chinois est en particulier étudié par Huang (1994), qui examine l'introduction et la maintenance d'une troisième personne au singulier par les expressions anaphoriques dans les conversations chinoises, et obtient certaines conclusions sur l'usage des noms propres, des SN et des pronoms. Au moment d'introduire un référent, les noms propres chinois sont préférés par rapport aux SN :

> *First, a proper name is often chosen over a definite description even in cases where the role of the referent is obviously relevant and thereby a definite description might have usefully been employed. (…)*
>
> *Secondly, a proper name is often used immediately following a definite description. (…)*
>
> *Thirdly, in cases where the speaker has a 'word-search' problem, and cannot find a referential form for the person he has in mind, it is normally the proper name that is solicited.*
>
> *(Huang, 1994 : 212)*

Au moment du changement de référent, ce sont les SN qui sont relativement préférés :

> *The shift of reference in Chinese conversation is done through the use of a lexical NP. (Huang, 1994 : 222)*

Les pronoms explicites ainsi que les zéros peuvent tous les deux favoriser la maintenance du référent :

> *The maintenance of reference in Chinese conversation tends to receive first a pronoun and then a zero anaphor encoding. (Huang, 1994 : 224)*

Comme en français, ces remarques de Huang sur les expressions anaphoriques chinoises utilisées dans les conversations peuvent avoir des points similaires avec celles des séquences dialogales ; mais ces dernières ne sont pas accompagnées de situation d'énonciation immédiate.

Les séquences dialogales n'équivalent pas à d'authentiques dialogues à l'oral, malgré quelques points communs entre eux, comme la présence des interlocuteurs et des expressions déictiques. Elles peuvent comporter des descriptions gestuelles qui servent de contexte énonciatif aux discours rapportés directs. Notre observation portera surtout sur ces facettes de notre corpus pour vérifier quelles sont les différences reflétées par les expressions anaphoriques entre les conversations et les séquences dialogales en chinois. Notre hypothèse est donc relative aux différences et aux points communs éventuels entre ces remarques sur les conversations et celles qui pourraient être présentes dans les séquences dialogales dans le roman-fleuve.

3.2.3.2. Séquences mono-référentielles et multi-référentielles

Avant de définir ce qu'est un contexte mono-référentiel et multi-référentiel, il nous semble important de reprendre tout d'abord la notion de *saillance* [17] sur laquelle se fonde la définition des deux types de contextes relativement au nombre de référents. Conditionné par le genre discursif, la saillance d'un référent peut être reflétée principalement par quatre facteurs discursifs, selon Schnedecker (2011) : la forme morpho-syntaxique de l'expression, sa place dans le cadre phrastique, le nombre de reprises et la catégorie sémantique du référent (*cf*. aussi Landragin, 2004 ; Schnedecker, 2014 ; Hou, 2021, entre autres). Lorsqu'il y a un seul référent d'une saillance bien distincte des autres, le contexte est dit *mono-référentiel* ; lorsqu'il y a plusieurs référents aussi saillants les uns que les autres, il est possible que les lecteurs ne retiennent pas tout le temps les référents déjà mentionnés ou qu'ils confondent les référents avec les expressions variables, d'où les problèmes de continuité référentielle dans le contexte dit *multi-référentiel*.

> *Un texte ne parle pas seulement d'un objet mais mentionne une pluralité d'entités, sous la forme d'expressions référentielles dont le nombre est en proportion. C'est dans ce sens qu'il est multi-référentiel. (Schnedecker, 1992 : 161)*
>
> *Par situations de saillance mono-référentielle, nous entendons ici des textes ou séquences textuelles focalisées sur un personnage principalement – ce qui peut être explicitement indicié par le fait que le titre mentionne le nom propre de ce personnage –, sans qu'il y ait, par ailleurs, référence à d'autres entités humaines animées, ou sinon, sans qu'il y ait de challengers, tels qu'ils ont été caractérisés précédemment. (Schnedecker, 1992 : 218)*

En conséquence, le contexte, *mono-référentiel* ou *multi-référentiel*, conditionne la composition des chaînes de référence.

Un texte mono-référentiel ne signifie pas qu'il contient un seul référent ou une seule chaîne de référence, mais un seul référent saillant, reflété par son grand nombre d'occurrences, ou sa densité élevée, ou encore la longueur significative de sa chaîne, comme la chaîne « Jean-Christophe » et les autres référents dans l'extrait [16] :

[16]Trois années ont passé. [**Christophe**] va avoir onze ans. [**Il**] continue [**son**] éducation musicale. [**Il**] apprend l'harmonie avec Florian Holzer, l'organiste de Saint-Martin, un ami de grand-père, un homme très savant. Le maître [**lui**] enseigne que les accords qu'[**il**] aime le mieux, des harmonies qui [**lui**] caressent si doucement l'oreille et le cœur qu'[**il**] ne peut les entendre sans un petit frisson tout le long de l'échine, sont mauvais et défendus. Quand [**l'enfant**] demande pourquoi, il n'est pas

[17] Voir aussi la section ***1.6.3. Configuration selon la saillance du référent*** du ***Chapitre 1***.

d'autre réponse, sinon que c'est ainsi : la règle les défend. Comme [il] est naturellement indiscipliné, [il] ne les en aime que mieux. [Sa] joie est d'en trouver des exemples chez les grands musiciens qu'on admire, et de les apporter à grand-père, ou à [son] maître.

En revanche, un texte multi-référentiel contient deux ou plusieurs référents saillants, dont les nombres d'occurrences sont similaires ou aussi élevés les uns que les autres ; ces référents saillants constituent donc une relation *leader-challenger* entre eux. Les interférences potentielles et réelles des *challengers* sont synthétisées dans le tableau (3.1) de Schnedecker (1997).

Tableau (3.1) : Interférence réelle ou potentielle des référents *challengers* (Schnedecker, 1997 : 80)

	Interférence potentielle	Interférence réelle
Contexte d'occurrences du « challenger »	> 3 phrases	< 3 phrases
Nbre de « challengers »	< 5	> 5
Profil du « challenger »	Nom d'inanimé Non sujet Non agent Dépendant Nouveau	Npr Sujet Agent Autonome Pré-activé

Par exemple, les chaînes « grand-père » et « Jean-Christophe » dans l'extrait [17] forment une relation leader-challenger entre les deux référents :

[17][**Grand-père**]$_{CR1}$ dormait trop longtemps quelquefois ; [**son**]$_{CR1}$ visage devenait rigide, [**son**]$_{CR1}$ long nez se tirait, [**sa**]$_{CR3}$ bouche s'ouvrait en long. [**Christophe**]$_{CR2}$ [**le**]$_{CR1}$ regardait avec inquiétude et [**Ø**]$_{CR2}$ craignait de voir [**sa**]$_{CR1}$ tête se changer en une forme fantastique. [**Il**]$_{CR2}$ chantait plus fort pour [**le**]$_{CR1}$ réveiller, ou [**il**]$_{CR2}$ se laissait dégringoler à grand fracas de [**son**]$_{CR2}$ talus de pierres. Un jour, [**il**]$_{CR2}$ inventa de [**lui**]$_{CR1}$ jeter à la figure quelques aiguilles de pin, et de [**lui**]$_{CR1}$ dire qu'elles étaient tombées de l'arbre. [**Le vieux**]$_{CR1}$ le crut : cela fit bien rire [**Christophe**]$_{CR2}$. Mais [**il**]$_{CR2}$ eut la mauvaise idée de recommencer ; et, juste au moment où [**il**]$_{CR2}$ levait la main, [**il**]$_{CR2}$ vit les yeux de [**grand-père**]$_{CR1}$ qui [**le**]$_{CR2}$ regardaient. Ce fut une méchante affaire : [**le vieux**]$_{CR1}$ était solennel et [**Ø**]$_{CR1}$ n'admettait point la raillerie sur le respect qu'on [**lui**]$_{CR1}$ devait ; ils restèrent en froid pendant plus d'une semaine.

Comme Schnedecker & Longo (2012) l'affirment, les faits divers sont un genre de discours narratif *multi-référentiel*. Il est donc question du nombre de référents centraux portés par le texte : Schnedecker (1992) définit le contexte *mono-référentiel* par la présence d'un seul référent central, et le contexte *multi-référentiel* par le fait que plusieurs référents sont aussi saillants les uns que les autres.

D'autres auteurs utilisent le même cadre de définition, comme Masseron & Schnedecker (1995), pour montrer que le nombre de référents peut sans conteste influencer les chaînes de référence. En ce qui concerne celles des textes mono-référentiels, Masseron & Schnedecker (1995) observent qu'elles sont composées de deux modes d'expression en alternance : la répétition par un SN et le pronom personnel de troisième personne. Certains

textes sont caractérisés comme mono-référentiels, comme les portraits journalistiques (Jenkins, 2002) : à chaque fois une seule personne particulière comme thème central ; et comme les textes encyclopédiques (Oberlé *et al*., 2018 ; Guo, 2021) : chaque texte un objet ou un thème porté dans le titre et qui se situe à la place du thème de texte ; d'autres textes sont multi-référentiels, comme les romans. Les textes que nous choisissons comme corpus permettront d'observer les caractéristiques des chaînes dans un contexte multi-référentiel dans les deux langues.

3.2.4. Chaînes de référence au vu de la longueur du texte

Le roman-fleuve est un genre narratif plus long que d'autres genres littéraires, comme la poésie et le théâtre ; il est également plus long que d'autres genres narratifs, comme les contes de fées et les faits divers. Dans toutes les études que nous avons présentées, les auteurs choisissent plutôt des textes courts comme corpus : il s'agit, pour certaines, de l'intégralité de textes courts, et pour d'autres, d'extraits de textes longs.

Lorsqu'il s'agit de l'intégralité de textes courts, les auteurs, comme Schnedecker (2005) et Baumer (2012b), constatent des différences de chaînes de référence dans les nouvelles courtes et les portraits journalistiques, deux genres courts narratifs et non narratifs ; ils concluent que les textes journalistiques sont représentatifs de l'alternance éventuelle des pronoms et des SN dans les chaînes.

> *[Dans les textes journalistiques] Le pronom n'est pas nécessairement la seule expression référentielle en vigueur et qu'il est susceptible d'alterner avec des SN pleins à certains endroits-clés des textes. Mais elles ne prédisent pour autant ni la catégorie des SN pleins à ces endroits (nom propre, SN défini ou démonstratif) ni,* a fortiori, *la tête lexicale de ces SN. […] Les frontières des unités considérées sont marquées par des SN pleins alors qu'au sein de ces unités, le maintien de la référence est assuré par des pronoms. (Schnedecker, 2005 : 8)*

Lorsqu'il est question d'extraits courts issus de textes longs, Baumer *et al*. (2020) choisissent des débuts de romans réalistes ; la longueur de ces extraits ne dépasse pas 10 000 mots. Hou (2021) constitue son corpus de thèse avec des extraits d'environ 3000 tokens de textes narratifs, à savoir un roman d'Émile Zola (1840-1902) avec sa traduction en chinois et une science-fiction romanesque de l'écrivain chinois Cixin Liu (刘慈欣, 1963-) avec sa traduction en français. Les corpus de ces deux dernières études sont potentiellement élargissables à des textes plus longs. Hou (2021) choisit les critères suivants pour comparer les chaînes de référence en français et en chinois dans les extraits appartenant à des textes longs ; il nous paraît donc important de savoir si ces critères sont également pertinents pour comparer les chaînes dans les textes longs des deux langues :

i) Le nombre d'expressions référentielles et la densité référentielle

Les expressions référentielles sont plus nombreuses et la densité référentielle plus haute dans les textes chinois de science-fiction ; le roman narratif montre la tendance inverse.

ii) Le nombre de référents

Les référents sont plus nombreux dans les textes chinois dans les deux œuvres.

iii) La catégorie grammaticale privilégiée

Dans le même contexte, ce sont les SN définis qui occupent la plus grande proportion de maillons des chaînes de référence en français, alors que ce sont les SN sans déterminant qui dominent dans les extraits en chinois.

iv) La longueur moyenne des chaînes et la longueur de la chaîne la plus longue

Dans les deux langues, les chaînes de référence sont en moyenne plus longues dans le roman narratif que dans la science-fiction. Les chaînes les plus longues montrent la même tendance dans chaque texte. Ces deux critères relatifs à la longueur des chaînes sont dépendants des genres et non des langues.

v) Le coefficient de stabilité

Le coefficient de stabilité est plus élevé en français qu'en chinois dans les deux œuvres. Les désignations nominales sont moins nombreuses en français qu'en chinois.

vi) La distance inter-maillonnaire

Les maillons dans les chaînes les plus longues présentent une distance plus importante dans le roman d'Émile Zola que dans le récit de science-fiction de Cixin Liu. Le changement de langue n'est pas le facteur primordial qui influence cet indice.

vii) La catégorie grammaticale des premiers maillons

Les SN définis servent davantage à initier une chaîne de référence en français ; ce sont les SN sans déterminant qui jouent ce rôle en chinois.

viii) La modalité d'introduction d'un nouveau référent

Une observation qualitative permet d'obtenir la construction canonique « SN indéfini – circonstanciel – verbe » pour le français, et la construction existentielle « circonstanciel – SN indéfini » pour le chinois.

> *Pour introduire un nouveau référent dans le discours, l'auteur du texte original en français penche pour une construction canonique, […], dans lesquelles les GN indéfinis se trouvent au début de la phrase (ou à la position préverbale), et les syntagmes locatifs se trouvent entre les GN indéfinis et les verbes. Lorsque ces deux phrases sont traduites en chinois, […], les traducteurs choisissent d'utiliser les constructions existentielles, en introduisant en première position les syntagmes locatifs et en postposant les GN indéfinis. (Hou, 2021 : 162)*

D'après tous ces travaux et toutes ces remarques, les chaînes de référence dans les romans-fleuve ne sont pas vraiment étudiées dans une seule langue ou par l'approche comparative franco-chinoise. Le roman-fleuve est un sous genre narratif du roman, caractérisé par sa longueur importante. D'une part, les chaînes de référence dans les romans-fleuve ne présentent pas les mêmes caractéristiques que celles des courts textes littéraires et narratifs, et, d'autre part, elles ne se constituent pas non plus de la même manière en français et en chinois.

Certains aspects nous semblent intéressants en ce qui concerne les textes longs. Premièrement, les chaînes des référents saillants d'un roman-fleuve doivent présenter des longueurs plus significatives que celles des textes narratifs courts ; les chaînes des référents non saillants peuvent être aussi longues ou moins longues que celles des textes narratifs courts lorsqu'elles couvrent une petite partie d'un roman-fleuve. Deuxièmement, les moyens de maintenance dans les textes longs peuvent être similaires à ceux des textes courts ou différents d'eux, ce qui est également une question capitale quant aux composants privilégiés des chaînes longues. Troisièmement, les textes longs permettent de voir les modalités éventuelles de redémarrage des chaînes après une longue distance inter-maillonnaire.

3.3. Bilan et hypothèses

Les études sur les chaînes de référence dans divers genres et diverses séquences textuelles pour le français nous incitent à observer les chaînes en chinois selon la même optique : les chaînes de référence manifestent-elles les mêmes caractéristiques dans les mêmes types de séquences textuelles en français et en chinois ? À partir des résultats obtenus selon plusieurs approches définitoires et comparatives, *Jean-Christophe* est défini comme un roman-fleuve de longueur démesurée faisant partie du genre narratif, constitué par des séquences narratives et dialogales, ainsi que par des séquences mono-référentielles et multi-référentielles. Nous chercherons à formuler, de manière globale, les caractéristiques hypothétiques des chaînes de référence dans *Jean-Christophe* en français et en chinois selon divers critères.

Relativement aux aspects sémantiques, les types de référents sont censés être aussi variables dans les textes parallèles des deux langues ; mais, d'après la comparaison réalisée par Hou (2021), le nombre de référents peut être plus élevé en chinois qu'en français. Les personnages sont théoriquement au centre de la narration ; ils peuvent donc être référents privilégiés des chaînes dans les deux langues. Il se peut qu'il y ait des référents de lieux et de temps qui marquent les changements d'état dans la narration. Et vu que le langage littéraire tolère une certaine ambiguïté, il se peut aussi que des expressions renvoient à des référents de manière équivoque, ce qui demande l'interprétation des lecteurs.

En ce qui concerne les aspects formels, les catégories grammaticales privilégiées en français peuvent être les noms propres, les SN (dont les définis peuvent être plus fréquents que les autres sous-types) et les pronoms explicites, alors que celles du chinois sont les SN sans déterminant, les pronoms et les zéros. Le premier maillon français peut être un nom propre, un SN défini ou un SN indéfini, alors que celui du chinois tend à être un SN sans déterminant ou un SN indéfini. La variation des têtes lexicales et les expansions de SN peuvent refléter les états ou la progression d'un personnage topical dans les séquences narratives dans

les deux langues. Les expansions des maillons peuvent se réaliser en adjectifs, en groupes prépositionnels et en d'autres formes possibles dans les deux langues. Le coefficient de stabilité est censé être similaire en français et en chinois puisque les textes sont parallèles : il est prévisible que les têtes lexicales seront traduites de la même manière du français vers le chinois.

En ce qui concerne les aspects syntaxiques, les modalités d'introduction des référents topicaux sont presque imprévisibles : il peut y avoir une seule modalité privilégiée, comme « Il était une fois un SN1 qui verbe SN2 » dans les contes de fées, ou y en avoir plusieurs, comme dans les faits divers (*cf.* Schnedecker, 2017). On peut également trouver la construction existentielle « il y a – SN indéfini » en français et 'il y a – SN sans déterminant' *you* 有 – SN sans déterminant en chinois. Nous supposons donc qu'il existe des constructions syntaxiques privilégiées qui conditionnent l'introduction de référents saillants et non saillants dans notre roman-fleuve en français et en chinois.

On peut d'ores et déjà prévoir que de grandes différences apparaîtront entre les deux langues : celles-ci sont relatives aux structures syntaxiques et aux variations morphologiques, en raison des caractéristiques inhérentes à chaque langue, d'une part, et, de l'autre, à cause des résultats confirmés des études antérieures :

- composition générale des chaînes
- composant(s) privilégié(s) des chaînes
- catégorie sémantique privilégiée des SN
- coefficient de stabilité
- modalité privilégiée d'introduction du référent central
- type d'expansion privilégié des SN

Concernant les aspects textuels, les chaînes de référence dans les deux langues peuvent présenter des points similaires en termes de critères au niveau global du roman :

- nombre de chaînes
- longueur moyenne des chaînes
- nombre de référents centraux
- distance moyenne entre les maillons
- densité des expressions référentielles
- longueur de la chaîne la plus longue
- coïncidence entre la chaîne et le thème du texte
- coïncidence entre la première mention et le thème de la phrase
- mode privilégié de cohabitation

Ce sont donc là des pistes d'observation que nous suivrons pour examiner les chaînes de référence de manière quantitative et qualitative dans la *Partie III*.

Chapitre 4 : Outils pour la constitution d'un corpus bilingue annoté en expressions référentielles et en chaînes de référence

Les logiciels linguistiques sont indispensables quand il s'agit de réaliser et de faciliter des opérations sur un corpus de grande taille. Ces outils sont déjà conçus pour le traitement des expressions référentielles dans certaines langues spécifiques : Referee (De Cristofaro *et al.*, 1999), PALinkA (Orasan, 2003) pour l'angleis, et, pour le français, Glozz (Widlöcher & Mathet, 2009), RefGen (Longo & Todrascu, 2010), TXM (Heiden, 2010), SACR (Oberlé, 2018) et CoFr (Wilkens *et al.*, 2020). La spécificité de notre corpus bilingue franco-chinois nous conduit à choisir des outils adéquats et compatibles pour opérer dans les deux langues en vue de l'annotation des expressions référentielles et coréférentielles. Le traitement des textes en deux langues par les mêmes outils facilitera énormément le travail long et lourd de l'annotation que nous détaillerons dans le *Chapitre 5*. Il convient de choisir les plus appropriés, les plus efficaces et les plus ergonomiques pour le traitement de notre corpus bilingue, afin d'optimiser ce travail de collection des données les plus représentatives possibles avec une quantité la plus importante possible. Après avoir testé plusieurs outils, nous avons sélectionné AiHanYu, SACR et CRViewer pour le traitement de notre corpus. Nous présenterons dans cette partie les opérations de notre corpus par le biais de ces trois outils.

4.1. Outil de tokenisation et de catégorisation automatique du chinois : AiHanYu

La présentation de l'outil AiHanYu est divisée en trois parties. Nous expliquerons tout d'abord la notion de *tokenisation* pour le chinois ; nous aborderons ensuite les fonctions de l'outil en question ; nous terminerons enfin par deux tests relatifs au taux de bonne annotation automatique de cet outil.

4.1.1. Particularité du traitement des textes en chinois : tokenisation

Une caractéristique de la langue chinoise consiste dans le fait que les mots s'écrivent en continuum, c'est-à-dire qu'ils ne sont pas séparés par des espaces. À cette particularité ajoutent des difficultés liées au traitement automatique du chinois (*cf.* Xia *et al.*, 2000). Les mots chinois se présentent en tant que suites de *caractères*[1] depuis le début le début d'un paragraphe et jusqu'à sa fin. Par conséquent, les outils conçus spécifiquement dans le but de traiter les langues en alphabet latin pourraient traiter automatiquement un paragraphe en chinois comme un seul mot. Il est donc nécessaire d'effectuer au préalable un traitement spécifique de segmentation des unités lexicales dans les textes en chinois avant toute autre

[1] Le terme *caractère* est utilisé ici dans son sens informatique, qui signifie « signe typographique utilisé pour l'encodage du texte sur un support lisible par l'ordinateur. » (Lebart & Salem, 1994 : 311 ; Zimina-Poirot, 2005 : 234).

annotation. Dans le domaine de la linguistique informatique, ce processus s'appelle *tokenisation* : il consiste à séparer les caractères chinois en *tokens* en insérant systématiquement des espaces ou des marqueurs selon les outils, afin de pouvoir obtenir les unités lexicales chinoises.

4.1.2. AiHanYu : fonctions et opérations

En raison de la spécificité de l'écriture chinoise, le premier outil que nous avons retenu est AiHanYu, une plateforme linguistique multi-fonctionnelle ouverte et gratuite en ligne, conçue par l'*Institut de Linguistique appliquée du Ministère de l'Éducation chinois*[2]. Parmi les multiples fonctions proposées, nous utilisons celles de tokenisation et de catégorisation automatique des unités lexicales chinoises.

Comme le montre la figure (4.1), la plateforme relative à ces deux opérations est divisée en deux fenêtres : une en haut pour l'import du texte brut, et une en bas pour l'export du texte annoté en tokens et en catégories grammaticales. Il suffit de taper ou de copier-coller le texte dans la fenêtre du haut et ensuite de cliquer sur le bouton marqué par une flèche rouge dans la figure (4.1) au milieu 'tokenisation & catégorisation automatiques' 自动分词 & 标注词性 : le texte sera automatiquement tokenisé par des slash et annoté en catégories grammaticales dans la fenêtre du bas. Les unités lexicales chinoises, à savoir les mots, sont séparés par des *caractères délimiteurs*[3], *i.e.* les slash, et leur catégorie grammaticale est annotée en signes abrégés[4].

[2] 教育部语言文字应用研究所计算语言学研究室.

[3] La définition informatique de *caractère (non) délimiteur* est la suivante : « distinction opérée sur l'ensemble des caractères, qui entrent dans la composition du texte permettant aux procédures informatisées de segmenter le texte en occurrences (suite de caractères non délimiteurs bornés à ses extrémités par des caractères délimiteurs). » (Zimina-Poirot, 2005 : 234).

[4] *Cf. Annexe 1 : Signes abrégés des catégories grammaticales dans AiHanYu.*

Figure (4.1) : Plateforme de tokenisation et de catégorisation du chinois par AiHanYu

AiHanYu est donc choisi comme tokeniseur automatique de notre corpus ; l'outil tokenise et catégorise automatiquement les textes en chinois importés en les exportant sous forme tokenisée avec les étiquettes des catégories grammaticales sur chaque token.

Après cette opération, le texte brut [1] en chinois sera segmenté en unités lexicales et sera annoté en catégories grammaticales comme le montre l'extrait [2] ci-dessous. Par exemple, les noms sont marqués par « /n » ('nouveau-né' *yinger* 婴儿, 'berceau' *yaolan* 摇篮, 'vieil homme' *laoren* 老人), et les verbes sont marqués par « /v » ('s'agiter' *niudong* 扭动, 'entrer' *jinlai* 进来, 'laisser' *tuo* 脱).

[1] 初生的婴儿在摇篮里扭动。老人进来虽然把木靴脱在门外，走路的时候地板还是格格的响：孩子哼啊嗜的哭了。母亲从床上探出身子抚慰他；祖父摸索着点起灯来，免得孩子在黑夜里害怕。灯光照出老约翰·米希尔红红的脸，粗硬的白须，忧郁易怒的表情，炯炯有神的眼睛。

'Le nouveau-né s'agite dans son berceau. Bien que le vieux ait laissé, pour entrer, ses sabots à la porte, son pas a fait craquer le plancher : l'enfant commence à geindre. La mère se penche hors de son lit, afin de le rassurer ; et le grand-père allume la lampe en tâtonnant, pour que le petit n'ait pas peur de la nuit. La flamme éclaire la figure rouge du vieux Jean-Michel, sa barbe blanche et rude, son air bourru et ses yeux vifs.'

[2] 初生/v 的/u 婴儿/n 在/p 摇篮/n 里/nd 扭动/v 。/w 老人/n 进来/v 虽然/c 把/p 木/n 靴/x 脱/v
yinger *yaolan* *niudong* *laoren* *tuo*
在/p 门外/nl ，/w 走路/v 的/u 时候/n 地板/n 还是/d 格格/n 的/u 响/v ：/w 孩子/n 哼/v 啊/e 嗜/x 的/u 哭/v 了/u 。/w 母亲/n 从/p 床上/nl 探出/v 身子/n 抚慰/v 他/r ；/w 祖父/n 摸索/v 着/u 点起/v 灯/n 来/vd ，/w 免得/c 孩子/n 在/p 黑夜/nt 里/nd 害怕/v 。/w 灯光/n 照出/v 老/a 约/v 翰/x ·/w 米希/nh 尔/x 红/a 红/a 的/u 脸/n ，/w 粗硬/a 的/u 白/a 须/d ，/w 忧郁/a 易怒/v 的/u 表情/n ，/w 炯炯有神/a 的/u 眼睛/n 。/w

4.1.3. Deux tests sur les erreurs systématiques dans les annotations automatiques

À côté des avantages des opérations d'annotations automatiques que permet de réaliser AiHanYu, nous avons constaté des erreurs systématiques. Un premier examen visant à vérifier la fiabilité de l'annotation sur un échantillon de 100 tokens au début du roman[5] a montré que le taux de bonne tokenisation atteint 81 %. Autrement dit, sur 100 tokens, 81 sont correctement délimités et 19 ne le sont pas. En outre, le taux de bonne catégorisation est de 75 %, ce qui signifie que 75 tokens sont correctement catégorisés et 25 le sont mal. Certaines erreurs systématiques et particulières dans la tokenisation et dans la catégorisation sont mises en exergue dans les trois extraits suivants [3] – [5].

Dans l'extrait [3], 'fleuve-grondement' *jiang-sheng* 江声 (« le grondement du fleuve ») est un « nom commun[6] », censé être catégorisé par « n » selon les codes d'AiHanYu, qui l'annote cependant par « nh » pour signifier « nom propre d'humain » parce que 'fleuve' *jiang* 江 est un caractère polysémique qui peut à la fois être utilisé pour un nom de famille et désigner « le fleuve ». Les deux usages ne se distinguent pas dans leur forme d'écriture. Cette erreur de polysémie n'est pas la seule dans cette phrase : 'à-partir-de' *zi* 自 et 'maison' *wu* 屋 sont en effet respectivement une « préposition » (annotée par « p ») et un « nom commun » (annoté par « n »), alors qu'AiHanYu les annote comme un seul « nom propre de lieu » (« ns ») *ziwu* 自屋/ns, donc avec une tokenisation incorrecte ainsi qu'une mauvaise catégorisation.

[3] **江声/nh** 浩荡/a 。/w **自** **屋/ns** 后/nd 上升/v 。/w
jiangsheng haodang *zi* *wu* *hou* *shangsheng*
→ **江声/n** **自/p** **屋/n**
grondement-du-fleuve à-partir-de maison
'Le grondement du fleuve monte derrière la maison.'

Dans l'extrait [4], 'jaunâtre' *hunhuangde* 昏黄的 est un adjectif, censé être annoté par « a » selon les codes d'AiHanYu, qui l'annote néanmoins comme deux mots *hunhuang* 昏黄/a *de* 的/u : un « adjectif » et un « auxiliaire ». Les annotations sont donc ici erronées au niveau de la tokenisation et de la catégorisation. Le même type de rajout dans la tokenisation se rencontre dans la suite de la même phrase : 's'étendre' *heixialai* 黑下来 est un verbe, censé être annoté par « v », alors qu'AiHanYu l'annote comme deux mots : un adjectif *hei* 黑/a et un verbe *xialai* 下来/v.

[4] **昏黄/a** **的/u** 天色/n **黑/a** **下来/v** 了/u 。/w
hunhuang de *tianse* *hei* *xialai* *le*
→ **昏黄的/u** **黑下来/v**
jaunâtre s'éteindre
'Le jour jaunâtre s'éteint.'

Dans l'extrait [5], 'sabot' *muxue* 木靴 est un « nom commun », censé être annoté par « n », alors qu'AiHanYu le sépare faussement en un « nom commun » *mu* 木/n et en un « caractère non morphème » *xue* 靴/x. Dans la suite de la même phrase, 'temps' *shihou* 时候

[5] *Cf. Annexe 2 : Examen du taux de bonne annotation par AiHanYu.*

[6] La définition de « nom commun » ici n'est pas la même de celle de la grammaire du français. Ces noms communs excluent les catégories comme « nom de temps », « nom de direction » et « nom de lieu commun » (voir l'*Annexe 1* et les exemples dans la section *5.2.2.4. Fonction syntaxique* du *Chapitre 5*).

est un « nom de temps », censé être annoté par « nt » d'après les codes de la catégorisation d'AiHanYu, alors qu'il est annoté comme un « nom commun » *shihou* 时候/n.

[5] 老人/n 进来/v 虽然/c 把/p **木/n** **靴/x** 脱/v 在/p 门外/nl ，/w 走路/v 的/u **时候/n** 地板/n
laoren *jinlai* *suiran* *ba* *mu* *xue* *tuo* *zai* *menwai* *zoulu* *de* *shihou* *diban*
→ **木靴/n** **时候/nt**
sabot temps
还是/d **格格/n** **的/u** **响/v** 。/w
haishi *gege* *de* *xiang*
→ **格格的响/v**
craquer
'Bien que le vieux ait laissé, pour entrer, ses sabots à la porte, son pas a fait craquer le plancher.'

D'après les analyses de ces trois exemples, les annotations erronées se produisent pour diverses raisons : les noms propres et les noms communs peuvent se confondre ; certaines tokenisations peuvent être superflues ; certaines catégorisations sont incorrectes. D'autres types d'erreurs systématiques et particulières existent probablement dans les autres fragments du corpus. ; pourtant, toutes les ponctuations sont systématiquement bien annotées. ; donc, il est fort probable que le taux de bonne annotation se réduira encore si on enlève le taux des bonnes annotations des ponctuations dans le texte.

Par ailleurs, il est difficile de savoir si l'annotation des expressions référentielles et coréférentielles est meilleure que celle des expressions non référentielles. En effet, si la plupart des erreurs se trouvent dans les parties non référentielles, il pourrait être plus pertinent de choisir cette annotation automatique d'expressions référentielles au lieu de les annoter manuellement. Nous avons donc effectué une deuxième observation des premiers 100 tokens concernant les expressions référentielles. Parmi les 100 tokens échantillonnés (*cf. Annexe 2*), 47 sont composants des expressions référentielles et coréférentielles, à savoir les déterminants, les classificateurs, les noms, les adjectifs, les pronoms et les auxiliaires ; 37 sont correctement tokenisés et catégorisés et 10 sont erronés. Le taux de bonne annotation diminue légèrement et tombe à 74 % par rapport au taux de bonne annotation de 75 % sur la totalité du texte.

Pour confirmer cette tendance, nous avons poursuivi l'analyse dans la suite du roman en retenant les 100 tokens suivants concernés par les expressions référentielles[7], où le taux de bonne tokenisation tombe à 59 % ; celui de bonne catégorisation chute à 55 %. Certaines erreurs systématiques et particulières des premiers 100 tokens sont illustrées dans les extraits [6] – [8].

Dans l'extrait [6], 'Jean-Michel' *Yuehan-Mixier* 约翰·米希尔 est un « nom propre d'humain », censé être annoté par « nh », alors qu'AiHanYu l'annote comme plusieurs mots avec des catégories grammaticales erronées, comme *yue* 约/v *han* 翰/x ·/w *mixi* 米希/nh *er* 尔/x, ce qui signifie qu'AiHanYu ne comprend pas du tout ce mot comme un nom propre. Cette erreur est récurrente et systématique dans l'annotation de tout le reste du corpus. Ensuite, 'barbe' *xu* 须 doit être annoté comme un « nom commun », par « n », et non comme un adverbe, « d ». Les trois adjectifs sont systématiquement mal tokénisés et catégorisés : 'rouge' *honghongde* 红红的 doit être annoté par « a » au lieu de *hong* 红/a *hong* 红/a *de* 的/u ; 'rude' *cuyingde* 粗硬的 par « a » au lieu de *cuying* 粗硬/a *de* 的/u ; 'bourru' *youyuyinude* 忧郁易怒

[7] *Cf. Annexe 3 : Examen du taux de bonne annotation des 101 – 200 tokens concernant les expressions référentielles et coréférentielles par AiHanYu.*

的 par « a » au lieu de *youyu* 忧郁/a *yinu* 易怒/v *de* 的/u ; ‘vif’ *jiongjiongyoushende* 炯炯有神 的 par « a » au lieu de *jiongjiongyoushen* 炯炯有神/a *de* 的/u.

[6] 灯光/n 照出/v 老/a **约/v 翰/x ·/w 米希/nh 尔/x 红/a 红/a 的/u** 脸/n ，/w
dengguang zhaochu lao yue han mixi er hong hong de lian
→ **约翰·米希尔/nh** **红红的/a**
Jean-Michel rouge
粗硬/a 的/u 白/a **须/d** ，/w **忧郁/a 易怒/v 的/u** 表情/n ，/w **炯炯有神/a 的/u** 眼睛/n 。/w
cuying de bai xu youyu yinu de biaoqing jiongjiongyoushen de yanjing
→ **粗硬的/a** **须/n** **忧郁易怒的/a** **炯炯有神的/a**
rude barbe bourru vif
‘La flamme éclaire la figure rouge du vieux Jean-Michel, sa barbe blanche et rude, son air bourru et ses yeux vifs.’

Dans l’extrait [7], le « classificateur[8] » *gu* 股/q (catégorie sans équivalent en français) est fautivement annoté comme un « nom commun » (« n ») en raison de la polysémie de ce caractère signifiant également « fesse » ; ‘bleu-chausson’ *lan-buxie* 蓝布鞋, constitué d’un « adjectif » ‘bleu’ *lan* 蓝 et d’un « nom commun » ‘chausson’ *buxie* 布鞋, est mal tokenisé et mal annoté comme « nom commun » *lanbu* 蓝布/n ‘bleu-tissu’ et un « nom commun » *xie* 鞋 /n ‘chaussure’.

[7] 他/r 走近/v 摇篮/n ，/w 外套/n 发出/v **股/n** 潮气/n ，/w 脚下/nl 拖/v 着/u 双/q
ta zoujin yaolan waitao fachu gu chaoqi jiaoxia tuo zhe shuang
→ **股/q**
CL
大/a **蓝布/n 鞋/n** 。/w
da lanbu xie
→ **蓝/a 布鞋/n** 。/w
bleu chausson
‘Il vient près du berceau. Son manteau sent le mouillé ; il traîne en marchant ses gros chaussons bleus.’

Dans l’extrait [8], l’adverbe *ban* 般 ‘comme’ (« d ») est annoté par AiHanYu comme « nom commun » (« n ») ; l’adjectif *heshande* 和善的 ‘doux’ (« a ») est incorrectement tokenisé et catégorisé comme un « adjectif » *heshan* 和善/a et un « auxiliaire » *de* 的/u ; le « verbe » *you* 有/v ‘avoir’ et le « descripteur » *xie* 些/f ‘des’ sont mal annotés comme un seul « numéral » (« m »).

[8] 她/r 的/u 淡黄/a 头发/n 差不多/a 象/p 白/a 的/u ；/w 绵羊/n **般/n 和善/a 的/u** 脸/n
ta de danhuang toufa chabuduo xiang bai de mianyang ban heshan de lian
→ **般/d** **和善的/a**
comme doux
都/d 打/v 皱/n 了/u ，/w 颇/d **有些/m** 雀斑/n ；/w
dou da zhou le po youxie queban
→ **有/ v 些/f**
avoir des
‘Elle est d’un blond presque blanc ; ses traits sont tirés ; sa douce figure mouton est marquée de taches de rousseur.’

[8] En chinois, certaines catégories ne sont pas considérées comme grammaticales en français. Nous citons ces catégories dans le *CNCorpus : Tagged Corpus Specification, Institute of Applied Linguistice Ministry of Education* (国家语委现代汉语通用平衡语料库 : 标注语料库数据及使用说明) de Xiao (肖航, 2009) en chinois avec notre traduction en français : ‘classificateur’ 量词, ‘descripteur’ 区别词, ‘préfixe’ 前接成分, ‘suffixe’ 后接成分, ‘idiome’ 习用语, ‘caractère morphème’ 语素词 et ‘caractère non morphème’ 非语素词.

Le résultat de cette dernière observation confirme notre hypothèse selon laquelle le pourcentage de bonne annotation d'AiHanYu diminue. Quand l'annotation se fait sur tous les tokens, le taux d'erreurs nous semble plus acceptable que celui des composants des expressions référentielles ; l'outil semble commettre encore plus d'erreurs sur ces derniers.

L'observation de la fiabilité relative des annotations automatiques nous a permis de souscrire à l'idée qu'il existe encore de grands écarts entre les analyses linguistiques humaines fines et les résolutions automatiques des machines. Des erreurs systématiques, et sans doute des erreurs particulières, sont inévitablement produites par les machines pendant les annotations automatiques. Néanmoins, nous considérons la tokenisation et la catégorisation automatiques d'AiHanYu comme référence des textes en chinois dans notre corpus ; il conviendra seulement de rectifier manuellement ces annotations inexactes à l'étape suivante.

4.2. Outil d'annotation manuelle : SACR

Nous avons choisi SACR (*Scripts d'Annotation des Chaînes de Référence*, Oberlé, 2018) pour l'annotation manuelle de notre corpus. Avant de présenter cet outil dans la seconde partie, nous expliquerons d'abord, dans la première, les raisons pour lesquelles nous avons choisi d'annoter manuellement au lieu d'opter pour une annotation automatique.

4.2.1. Problèmes relatifs à l'annotation automatique

En ce qui concerne la détection automatique des expressions anaphoriques et coréférentielles, celle des chaînes de référence est moins développée en raison du nombre plus élevé des expressions concernées. De nombreux auteurs ont remarqué la difficulté de mettre en relation automatiquement des expressions anaphoriques et coréférentielles, comme notamment Hobbs (1979), Corblin (1985), David (1990), Mitkov (2000), Mitkov *et al.* (2000), Doddington *et al.* (2004), Boudreau & Kittredge (2006), Potau (2010), Schnedecker & Landragin (2014), Lassalle (2015), Grobol *et al.* (2017), Landragin (2021) et Grobol (2021). Plusieurs raisons expliquant cette difficulté sont proposées : Boudreau & Kittredge (2006) constatent que les expressions coréférentielles peuvent concerner les expressions d'une grande diversité de catégories grammaticales, comme principalement les noms propres, les SN et les pronoms, et d'une grande variation de relations inter-maillonnaires, comme la coréférence, l'anaphore nominale et pronominale.

En termes de forme variable des maillons dans les chaînes de référence, deux types d'expressions sont particulièrement difficiles à capter en anaphore et en coréférence pour les machines. Le premier concerne les SN avec des têtes lexicales variables dans les anaphores nominales infidèles, comme le signalent Corblin (1987), Schnedecker (2006) et Schnedecker & Landragin (2014). En d'autres termes, quand un référent est repris par un SN avec un autre nom à la place de tête lexicale initiale, les outils n'arrivent en général pas à capter automatiquement la relation anaphorique entre les deux SN. Citons les exemples [9] – [12] donnés par Corblin (1987). « Un livre » dans le début de la phrase peut être repris par un SN de la même classe « ce livre » comme [9], ou des SN d'autres classes comme « objet »,

« roman » et « torchon » comme [10] – [12]. Dans les trois derniers énoncés, les humains arrivent à trouver intuitivement la relation entre « un livre » et le second SN, alors que les noms de diverses classes semblent plus complexes pour les machines, qui risquent d'échouer à constituer une relation coréférentielle avec un antécédent, surtout lorsqu'il y a des changements de tête lexicale, comme dans les énoncés [10] – [12] :

[9] Il y avait sur la table [un livre]. [**Ce livre**] …
[10]Il y avait sur la table [un livre]. [**Cet objet**] …
[11]Il y avait sur la table [un livre]. [**Ce roman**] …
[12]Il y avait sur la table [un livre]. [**Ce torchon**] …

(Corblin, 1987 : 210)

Ce type de variation des têtes lexicales dans les expressions nominales coréférentielles apparaît également dans l'extrait [13] de notre corpus, où la chaîne Christophe est composée de SN avec différents noms : « Christophe », « enfant », « petit-fils » et public ».

[13]Grand-père toussotait. [**Christophe**] savait bien ce que cela voulait dire. Le vieux brûlait d'envie de raconter une histoire ; mais il voulait que [**l'enfant**] la lui demandât. [**Christophe**] n'y manquait pas. Ils s'entendaient ensemble. Le vieux avait une immense affection pour [**son petit-fils**] ; et ce lui était une joie de trouver en lui [**un public complaisant**]. Il aimait à conter des épisodes de sa vie, ou l'histoire des grands hommes antiques et modernes.

Le second type d'expressions difficiles à détecter par les machines concerne les entités nommées[9] par les noms propres. Pour Poibeau (2003), celles-ci représentent une grande variation, différente dans chaque langue. Certaines ressources lexicales existent déjà dans les dictionnaires informatiques pour des langues européennes, mais elles sont largement insuffisantes pour beaucoup d'autres langues. De plus, Poibeau (2005) indique que les entités nommées sont difficiles à délimiter également en raison de la polysémie : un nom propre peut être classé de différentes manières : « l'ONU » peut être catégorisé comme une institution dans [14], comme une communauté d'individus dans [15] ou encore comme un bâtiment dans [16].

[14][**L'ONU**] a fêté ses 50 ans.
[15][**L'ONU**] était en grève hier.
[16][**L'ONU**] n'acceptera pas une attaque frontale de l'Irak (forum du Monde).

(Poibeau, 2005 : 175)

4.2.2. Problèmes soulevés spécifiquement en français et en chinois

Les questions de la détection automatique des expressions coréférentielles sont multiples et spécifiques à chaque langue. Toutes ne sont pas abordées de manière égale selon les systèmes linguistiques ; l'anglais est sans doute le plus étudié. Les nombreuses solutions apportées fournissent des orientations de travail pour le français et le chinois (Passonneau, 2004 ; Zhao & Ng, 2007 ; Raghunathan *et al.*, 2010). Pour le français, la reprise anaphorique pronominale s'accorde au nombre et au genre de l'antécédent, ce qui permet en quelque sorte aux machines de retrouver les entités en question. Comme le montre l'énoncé [17], « elle » correspond au nombre singulier et au genre féminin de son antécédent « la valise ».

[17][La valise] est jolie mais [**elle**] est trop petite.

[9] Poibeau (2005) définit les entités nommées comme suit : « Les entités nommées désignent les noms de personnes, de lieux, d'organisations mais aussi les dates ou les unités monétaires » (2005 : 173).

Pourtant, la machine capte difficilement les relations coréférentielles exprimées par les pronoms dans les contextes où plusieurs sont possibles, comme le montrent les exemples [18] et [19] cités par Schnedecker *et al.* (2017 : 7). Certains pronoms peuvent être affectés à des référents non pertinents.

[18][La poussette] n'entre pas dans [la valise] car [**elle**] est trop grande.
réf-1 *réf-2* *réf-1*

[19][La poussette] n'entre pas dans [la valise] car [**elle**] est trop petite.
réf-1 *réf-2* *réf-2*

(Schnedecker *et al.*, 2017 : 7)

Deux entités féminines au singulier sont mentionnées dans la première partie des deux phrases : « la poussette » et « la valise ». Toutes deux peuvent être reprises dans la suite de la phrase par un pronom singulier féminin à la troisième personne « elle ». « La poussette » est repris par « elle » dans la phrase [18] et « la valise » est repris par « elle » dans la phrase [19]. Aucune différence ne se trouve ni au niveau syntaxique ni au niveau morphologique. Les humains distinguent spontanément et intuitivement les deux relations coréférentielles entre « elle » et son antécédent selon le sens des deux phrases ; mais ni les analyses morphologiques ni les analyses syntaxiques des outils ne permettent, à notre connaissance, de détecter le référent de « elle ».

Pour le chinois, mis à part des difficultés similaires à celles du français, certaines sont causées par la grande quantité des zéros, qui a fait l'objet de nombreuses discussions de la part d'auteurs comme Hou & Sun (侯敏 & 孙建军, 2004), Zhao & Ng (2007), Yeh & Chen (2007) et Huang & Zhang (黄娴 & 张克亮, 2009), entre autres. Alors que les accords en genre et en nombre des SN et des pronoms en français proposent des traces complémentaires pour identifier la coréférence dans les situations où se trouve un seul référent (comme le montre l'énoncé [17]), le chinois, en revanche, n'en dispose pas, comme le montre la traduction chinoise [20].

[20][行李箱] 很 漂亮， 但是 [**Ø**] 太 小了。
xinglixiang *hen* *piaoliang* *danshi* *tai* *xiaole*
valise très joli mais trop petit
réf-1 *réf-1-zéro*

Les énoncés [18] et [19] en français sont traduits en chinois par nous comme [21] et [22], où les reprises pronominales en français sont équivalentes à des zéros sans aucune information sur le nombre ou le genre de l'antécédent.

[21][小推车] 放不进 [行李箱]， 因为 [**Ø**] 太 大了。
xiaotuiche *fangbujin* *xinglixiang* *yinwei* *tai* *dale*
poussette entrer-NEG valise car trop grand
réf-1 *réf-2* *réf-1-zéro*

[22][小推车] 放不进 [行李箱]， 因为 [**Ø**] 太 小了。
xiaotuiche *fangbujin* *xinglixiang* *yinwei* *tai* *xiaole*
poussette entrer-NEG valise car trop petit
réf-1 *réf-2* *réf-2-zéro*

Par ailleurs, à la différence des pronoms personnels chinois qui sont toujours référentiels, le pronom personnel « il » en français peut également être utilisé dans des structures impersonnelles dans lesquelles « il » ne réfère pas, comme « il pleut » et la structure « il faut ».

À notre connaissance, pour le français, certains outils de détection automatique des chaînes de référence sont déjà réalisés, comme *RefGen* (Todirascu & Longo, 2011), ANCOR (Désoyer *et al.*, 2016) et CoFr (Wilkens *et al.*, 2020). Ces outils fournissent cependant des

résultats d'annotation moyens qui ne satisfont pas pleinement les attentes des linguistes (*cf.* Landragin *et al.*, 2017 ; Landragin, 2021) : certains peuvent traiter seulement des textes spécifiques, comme ANCOR pour le français oral ; d'autres produisent des erreurs systématiques, comme CoFr, sur l'identification d'expressions référentielles et / ou sur la constitution de la coréférence entre les expressions coréférentielles. En effet, d'une part, les phénomènes des chaînes de référence détectés automatiquement par les outils sont assez restreints, et, d'autre part, les capacités sont toujours insuffisantes pour des analyses linguistiques méticuleuses. Ainsi, à notre connaissance, aucun outil ne dispose de détection automatique des expressions coréférentielles à la fois pour le français et pour le chinois ; il est donc impossible d'annoter automatiquement un corpus bilingue franco-chinois avec un seul outil.

Après avoir examiné la fiabilité des annotations automatiques, nous souscrivons à l'idée consensuelle qu'il existe encore de grands écarts entre les analyses linguistiques humaines fines et les résolutions automatiques des machines. Des erreurs systématiques, et sans doute des erreurs particulières, sont produites par les machines. En raison de ces points faibles, il nous semble que les annotations manuelles pourraient être plus pertinentes pour des observations linguistiques fouillées dans les deux langues. Nous avons donc choisi d'effectuer une annotation manuelle des expressions référentielles et des chaînes de référence en français et en chinois.

4.2.3. Opérations en SACR

Certains logiciels d'annotation sont utilisés pour annoter les expressions référentielles et coréférentielles en français, comme TXM (Heiden, 2010) et ANALEC (Landragin, 2016b) ; ils annotent de manière automatique ou semi-automatique. Or, nous avons choisi finalement SACR, outil qui permet seulement d'annoter de manière manuelle. Plusieurs raisons président à ce choix. Tout d'abord, SACR est une plateforme d'opération gratuite en ligne spécifiquement destinée, dès sa conception et sa création, à annoter les expressions référentielles et coréférentielles. Il s'agit d'une plateforme de dépouillement de texte ayant pour objectif de créer des corpus ouverts[10], sur laquelle les utilisateurs peuvent importer et opérer tout type de texte. Pour importer un texte sur la plateforme, comme l'illustre l'interface d'importation des textes dans la figure (4.2), il suffit de cliquer sur « choisir un fichier » et choisir un texte au format texte brut ; il sera importé dans la case en bas de la fenêtre. La figure (4.2) montre la fin de l'opération de l'import du texte initié par « le grondement du fleuve ».

[10] Cette fonction est similaire à celle de *TXM* (*cf.* Heiden, 2010 ; Heiden *et al.*, 2010) et *ANALEC* (http://www.lattice.cnrs.fr/ressources/logiciels/analec/, http://explorationdecorpus.corpusecrits.huma-num.fr/analec-2/, *cf.* Landragin *et al.*, 2012 ; Landragin, 2016b).

Figure (4.2) : Interface de l'importation de texte brut dans SACR

L'annotation manuelle des chaînes de référence est un travail assez lourd. Chaque étape d'annotation et de rectification nous a pris énormément de temps en raison des multiples étapes de traitement du corpus (ce que nous préciserons dans le *Chapitre 5*). Le fait que SACR présente en particulier une ergonomie par rapport aux autres outils nous paraît donc important car cela permet un traitement de texte plus facile et plus rapide. Cette ergonomie est appréciable étant donné que l'annotation sur SACR permet de rassembler les trois étapes principales de l'annotation que nous souhaitons effectuer :

(i) délimiter les expressions référentielles ;

(ii) relier et nommer les expressions coréférentielles afin de constituer les chaînes de référence ;

(iii) ajouter les étiquettes des propriétés linguistiques de chaque expression référentielle.

Ces trois étapes d'annotation dans les deux langues sont combinées dans SACR. Présentons-les avec plus de précision.

- Étape (i) : délimiter et nommer les expressions référentielles

La délimitation des expressions référentielles se réalise en cliquant sur le premier et le dernier mot inclus dans les expressions. Par exemple, pour construire l'expression « le grondement du fleuve » comme *mention*[11], il suffit de cliquer sur le premier mot « le » et le dernier mot « fleuve » ; ensuite, la mention « le grondement du fleuve » apparaît dans un encadré gris. Si la mention ne contient qu'un seul mot, il suffit de double-cliquer sur le mot en question. De cette manière, toutes les expressions référentielles sont délimitées

[11] Voir la définition de *mention* dans la section *5.2.1. Annotation des mentions : singletons et maillons* du *Chapitre 5*.

manuellement. L'illustration de cette étape dans la fenêtre d'opération de SACR se trouve dans la figure (4.3).

Figure (4.3) : Interface de délimitation des expressions référentielles en français dans SACR

Si on ne nomme pas en particulier les mentions annotées, elles seront nommées automatiquement par un numéro dans SACR, comme « M1 » (« le grondement du fleuve ») signifiant « mention 1 », et ainsi de suite, comme « M2 » (« du fleuve ») et « M3 » (« la maison »).

Ensuite, la figure (4.4) illustre l'étape de la dénomination des mentions. Pour attribuer un nom de référent à chaque mention, il suffit de cliquer sur la mention concernée en tapant sur la touche « n ». Une fenêtre apparaît et permet d'entrer le nouveau nom de la mention. Par exemple, « fleuve » pour nommer la mention « du fleuve » ; la numérotation automatique « M2 » sera remplacée par cette dénomination.

Figure (4.4) : Dénomination des mentions par le nom du référent dans SACR

L'inconvénient de SACR pour les textes en chinois est que l'on ne peut pas attribuer un nom de référent dans l'écriture chinoise. En revanche, ces étiquettes alphabétiques, routines de travail dans le domaine du traitement automatique des langues, faciliteront la conversion ultérieure d'annotation de SACR vers d'autres outils. Comme le montre la figure (4.5), nous avons nommé les mentions en chinois par des noms en lettres, soit selon leur prononciation, soit selon leur sens en français.

Figure (4.5) : Dénomination des mentions en chinois par des noms en lettre dans SACR

- Étape (ii) : nommer et constituer les chaînes de référence

La figure (4.6) montre la plateforme de SACR au moment de la constitution des chaînes de référence à partir des mentions annotées lors de la première étape. Pour réaliser cette opération, les chaînes sont construites en déplaçant une mention vers les autres, qui appartiennent à la même chaîne de référence. Par exemple, pour créer la chaîne du personnage « Jean-Michel », dans la figure (4.6), on peut nommer d'abord la première mention de la chaîne « le vieux » par l'abréviation « JM ». Ensuite, pour constituer le lien coréférentiel entre les mentions à la suite, comme « le vieux », « ses » et « son », il suffit de cliquer sur la première mention déjà nommée en tapant *Ctrl* et en cliquant sur la deuxième mention « ses » en même temps, ou en faisant glisser les cartouches les unes dans les autres, et ainsi de suite pour « son » et les autres mentions de la même chaîne.

Figure (4.6) : Dénomination et construction des chaînes de référence dans SACR

À partir de la troisième mention renvoyant au même référent, toutes les mentions coréférentielles sont mises en relief automatiquement par une couleur ; chaque chaîne est alors marquée par une couleur aléatoire et différente. Par exemple, dans la figure (4.7), la chaîne « JM » désignant le personnage « Jean-Michel » apparaît en rose et contient les mentions comme « le vieux », « ses » et « son » ; la chaîne « Jean-Christophe » désignant le personnage principal « Jean-Christophe » apparaît en orange clair et se compose des

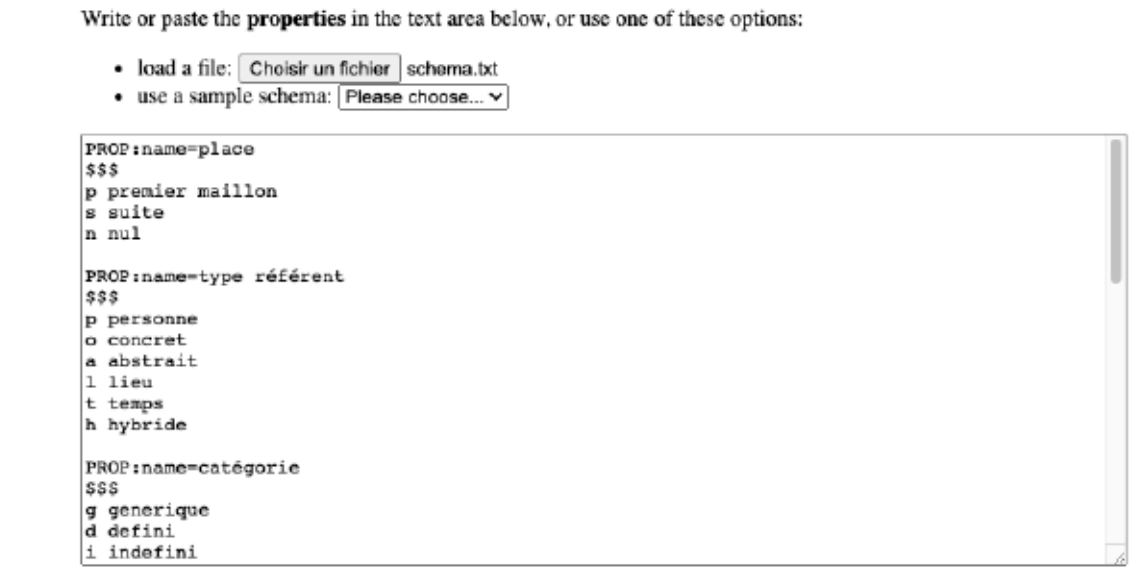
mentions comme « le nouveau-né », « son » et « l'enfant ». Les mentions qui ne réalisent pas de chaîne, comme « le grondement du fleuve », « du fleuve » et « la maison », restent marquées par défaut par les encadrés gris.

Figure (4.7) : Mise en couleur des chaînes de référence dans SACR

- Étape (iii) : ajouter les propriétés linguistiques des mentions

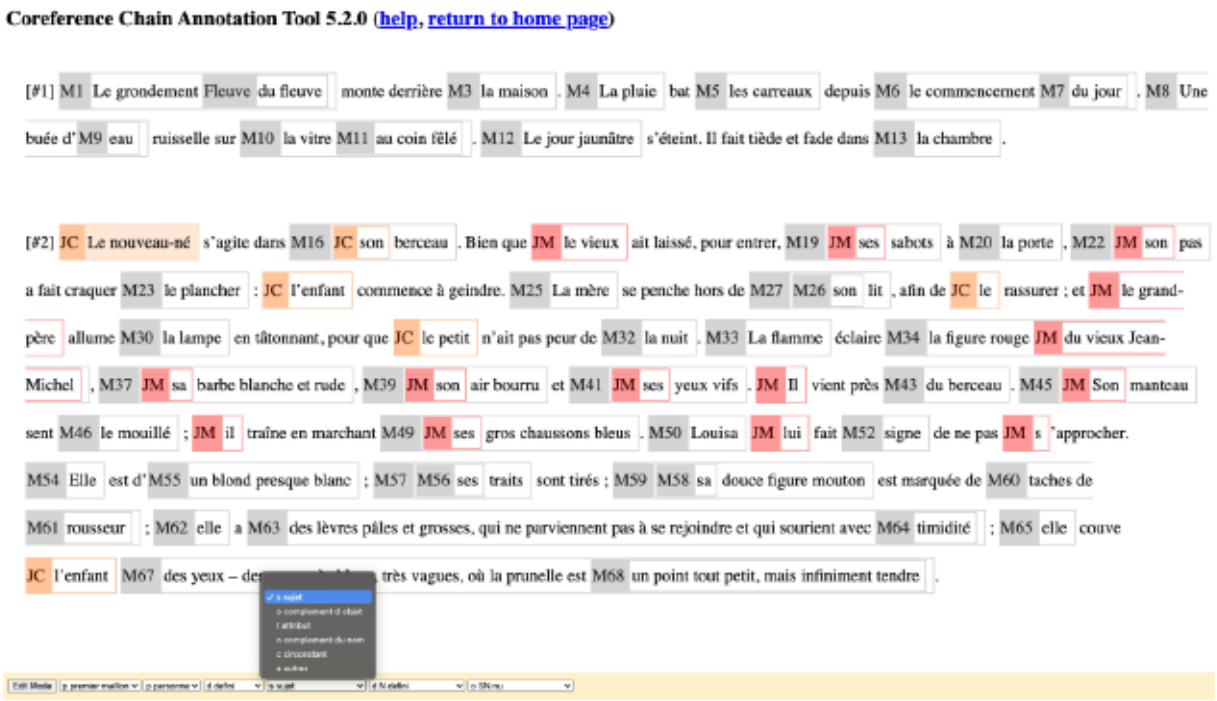
Après avoir délimité les expressions référentielles et constitué les liens entre les expressions coréférentielles, on peut ajouter les étiquettes des propriétés linguistiques à chaque mention annotée selon le schéma des propriétés importées, comme le montre la figure (4.8). Nous préciserons les propriétés choisies et les motivations dans la section *5.2.2. Annotation des caractéristiques des mentions* du *Chapitre 5*.

Figure (4.8) : Importation du schéma des étiquettes dans SACR



Après avoir ajouté les étiquettes d'annotation, la liste des propriétés s'affiche en bas de la fenêtre dans un menu déroulant quand on clique sur chaque mention, comme l'illustre la figure (4.9).

Figure (4.9) : Ajout des étiquettes des propriétés linguistiques des mentions annotées dans SACR



Après avoir expliqué les trois étapes principales de l'annotation dans SACR, nous présentons les autres avantages de cet outil d'annotation.

Le troisième avantage concerne sa capacité à traiter les corpus multilingues composés de différents alphabets, à savoir l'écriture française et l'écriture chinoise. Notre corpus franco-chinois exige en effet que l'outil soit compatible avec une annotation en français *et* en chinois ; or, SACR permet de traiter les textes en chinois suivant les mêmes principes d'annotation, comme le montre la figure (4.10).

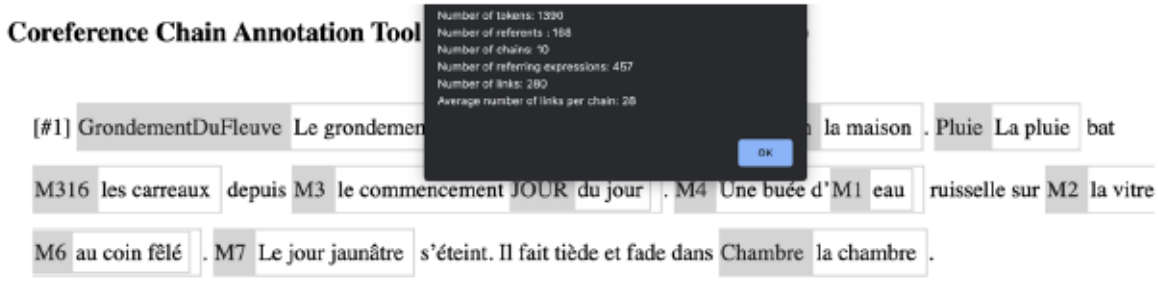
Figure (4.10) : Interface d'annotation des expressions référentielles en chinois dans SACR

Le quatrième avantage concerne la possibilité d'effectuer des calculs sur les textes annotés. En effet, SACR permet de directement calculer les chiffres issus du corpus annoté (ce qui, autrement, serait très fastidieux).

Après avoir annoté les textes en trois étapes, SACR peut calculer directement les paramètres suivants, que montre la figure (4.11) et la liste suivante :

- le nombre de mots du texte (« number of tokens ») ;
- le nombre de référents (« number if referents ») ;
- le nombre de chaînes de référence (« number of chains ») ;
- le nombre d'expressions référentielles (« number of referring expressions ») ;
- le nombre de maillons dans les chaînes de référence (« number of links ») ;
- le nombre de maillons en moyenne par chaîne de référence (« average number of links per chain »).

Figure (4.11) : Calculs directs sur les textes annotés dans SACR



Les calculs en haut concernent les paramètres dans l'intégralité du texte annoté. Sur une chaîne spécifique, SACR permet également de trier et de visualiser les maillons concernés. Par exemple, si l'on veut obtenir les maillons de la chaîne « Jean-Christophe » renvoyant au personnage « Jean-Christophe », il convient de cliquer sur une mention de la chaîne en appuyant sur *Alt+o* : tous les maillons de la chaîne « Jean-Christophe » sont alors surlignés avec leur couleur dans le texte, comme le montre la figure (4.12). Pour visualiser d'autres chaînes, il suffit de changer la mention sélectionnée au début.

Figure (4.12) : Visualisation d'une chaîne spécifique sélectionnée dans SACR

Pour visualiser les mentions d'une certaine étiquette, il suffit d'appuyer sur *Alt+s*, et une fenêtre s'ouvre, comme le montre la figure (4.13). Dans le menu déroulant, les étiquettes du schéma d'annotation sont affichées ; il suffit de sélectionner une des étiquettes souhaitées afin de visualiser toutes les mentions annotées de cette étiquette dans le texte.

Figure (4.13) : Interface de la recherche des mentions d'une même étiquette dans SACR

Après avoir sélectionné l'étiquette souhaitée, toutes les mentions annotées avec la même étiquette sont affichées. Par exemple, si on lance une recherche de « sujet », toutes les mentions occupant la fonction de « sujet » sont affichées, comme le montre la figure (4.14).

Figure (4.14) : Recherche des mentions d'une étiquette spécifique

Le dernier avantage de SACR est le fait que le texte annoté peut facilement être converti en d'autres formats et outils comme ceux de Glozz (Widlöcher & Mathet, 2009), de TXM (Heiden, 2010 ; Heiden *et al.*, 2010), d'ANALEC (Landragin *et al.*, 2012 ; Landragin, 2016b), et de CRViewer (Oberlé, 2017), ce qui permet d'envisager éventuellement des traitements ultérieurs.

4.3. Outil de calcul et de visualisation : CRViewer

Après les annotations manuelles dans SACR, il nous reste à obtenir les chiffres des chaînes de référence à partir des textes annotés. Pour ce faire, nous avons choisi CRViewer[12] (Oberlé, 2017) qui calcule automatiquement les chiffres des chaînes de référence et ensuite les affiche sous trois types de présentation : les tableaux de chiffres, les histogrammes et les diagrammes circulaires.

Pour les tableaux de chiffres, comme le montre la figure (4.15), CRViewer peut calculer les paramètres de chaque texte annoté du corpus en nombre absolu :

- la première colonne correspond à la « distance moyenne inter-maillonnaire » (*i.e.* le nombre de mots du texte divisé par le nombre total de maillons d'une chaîne), marqué par « AvgL2LDist », abrégé de « average link to link distance » ;
- la deuxième colonne présente la « longueur moyenne des maillons des chaînes de référence » (*i.e.* le nombre total de maillons divisé par le nombre de chaînes), marqué par « AvgLinkLength », *i.e.* « average link length » ;
- la troisième colonne montre la « longueur moyenne des chaînes de référence » (*i.e.* le nombre moyen de maillons par chaîne), marqué par « AvgChainSize », signifiant « average chain size » ;
- la quatrième colonne correspond au « nombre de chaînes de référence », marqué par « ChainCount » ;
- la cinquième colonne calcule le « nombre de maillons dans les chaînes de référence », marqué par « LinkCount ».

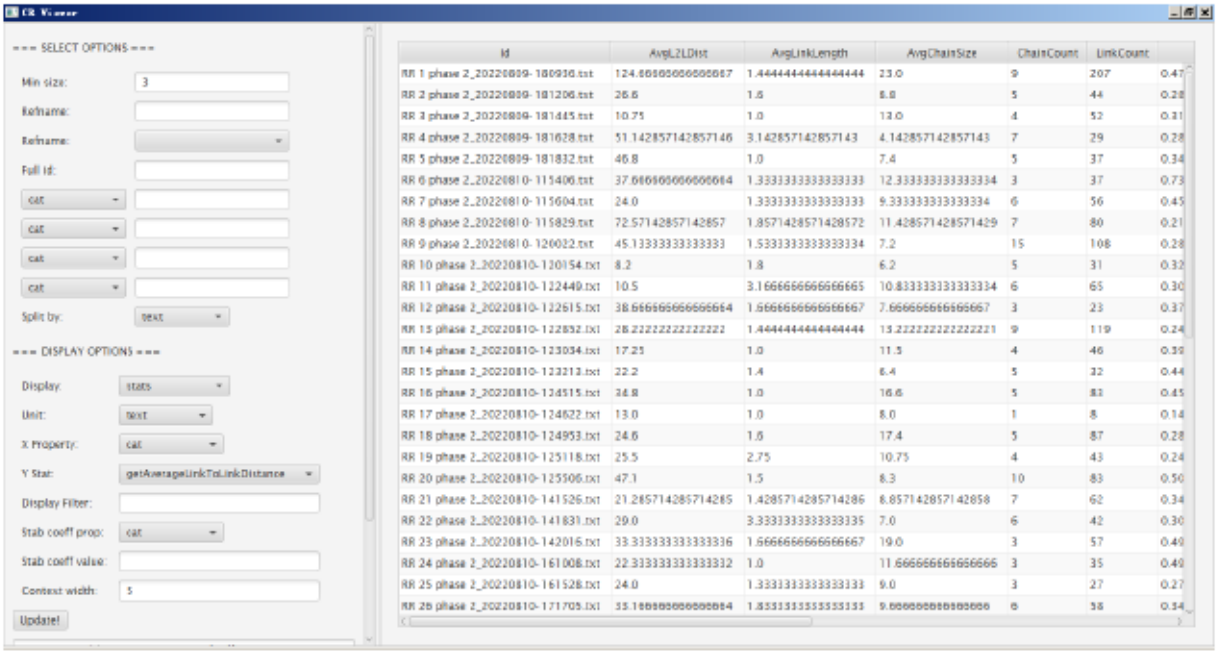
Les autres paramètres sont calculés dans la suite du tableau, que la capture d'écran, étant donné sa taille limitée, ne peut montrer dans son intégralité. Ces paramètres sont les suivants :

- le coefficient normalisé de stabilité (le nombre d'une tête lexicale d'une chaîne divisé par le nombre total des maillons nominaux de la chaîne[13]) ;
- la densité de chaîne (le nombre de chaînes divisé par le nombre de mots du texte) ;
- la densité de maillons (le nombre total de maillons divisé par le nombre de mots du texte) ;
- la densité référentielle (le nombre de maillons divisé par le nombre de mots du texte) ;
- le nombre de paragraphes.

[12] http://boberle.com/projects/crviewer/index.html, dernière consultation en juillet 2019.

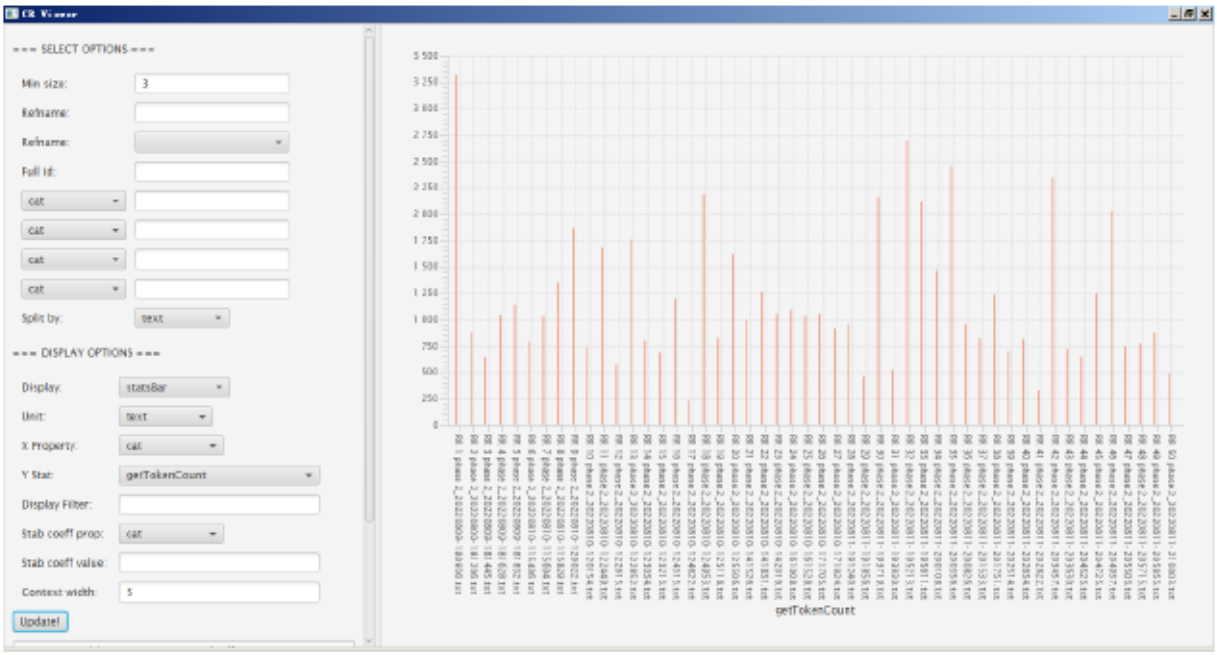
[13] Le coefficient d'Oberlé est un chiffre entre 0 et 1 : plus il est élevé, moins il y a de variation : 1 correspond à des reprises nominales identiques (*un chat … le chat*), 0 à des reprises toutes différentes (*un chat … l'animal*). (Oberlé *et al.*, 2018). Pourtant, ce calcul est l'inverse de celui de Perret (2000) et de Schnedecker (2021) que nous avons présenté dans la section *3.2.1.1. Caractérisation des chaînes de référence dans divers genres en français* du *Chapitre 3* : ce dernier concerne le nombre de maillons nominaux divisé par le nombre de têtes lexicales ; donc le chiffre obtenu est toujours supérieur à 1 qui correspond à la moindre stabilité et au fait que toutes les têtes lexicales sont différentes dans la chaîne. Nous utilisons cette dernière définition de Perret et de Schnedecker dans notre thèse.

Figure (4.15) : Calculs sur la base des textes annotés en tableau de chiffres par CRViewer



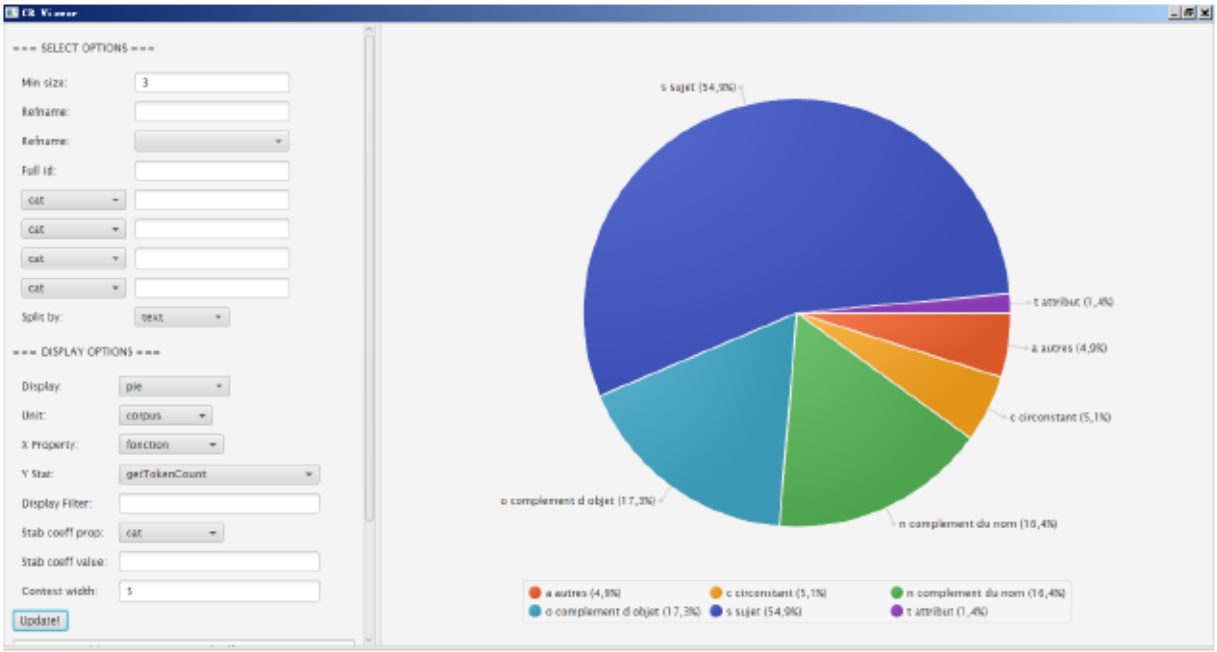
Les histogrammes permettent d'illustrer des paramètres en chiffres absolus qui servent à comparer un paramètre dans les différents textes du corpus, comme, par exemple, dans la figure (4.16), la longueur des textes, par rapport à laquelle on peut calculer le ratio des chaînes de référence dans chaque texte de différente longueur.

Figure (4.16) : Visualisation du nombre de mots des épisodes en français en histogramme par CRViewer



À la différence des tableaux et des histogrammes qui présentent les chiffres en absolu, les diagrammes circulaires peuvent clairement montrer les pourcentages, comme les types de référents, les catégories grammaticales des maillons et leurs fonctions syntaxiques. Par exemple, la figure (4.17) révèle la proportion des maillons selon leur fonction syntaxique dans le texte.

Figure (4.17) : Visualisation des paramètres en diagramme circulaire par CRViewer



Par le biais de ces trois types de présentation, nous arrivons à exposer les chiffres que nous avons obtenus dans le corpus. Les tableaux montrent tous les chiffres du texte et du corpus de manière spécifique et systématique. Les histogrammes illustrent plutôt la variation d'un seul paramètre entre les différents blocs de texte du corpus ; les tableaux et les histogrammes font voir les chiffres absolus, alors que les diagrammes circulaires révèlent les chiffres proportionnés par rapport à l'intégralité des phénomènes en question.

L'emploi de CRViewer est étroitement lié aux textes annotés par SACR. L'outil permet d'importer un corpus constitué d'un seul ou de plusieurs fichiers en format texte brut. Par conséquent, grâce au traitement rendu possible par ces deux outils, notre corpus peut être soit entièrement, soit partiellement importé selon les besoins d'obtention des chiffres.

4.4. Bilan

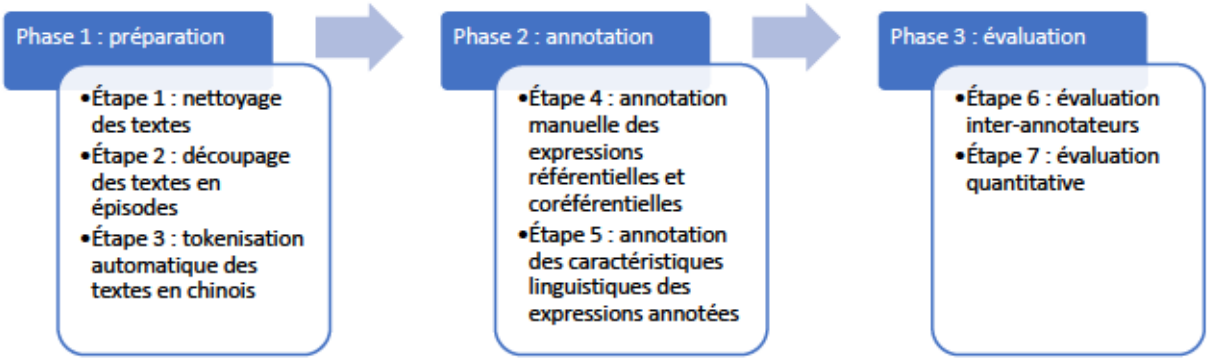
Récapitulons. Nous pouvons affirmer que AiHanYu, SACR et CRViewer répondent à nos attentes relativement au traitement de notre corpus bilingue. Ayant testé le taux d'erreurs de l'annotation automatique de AiHanYu, nous avons choisi de faire référence aux résultats de cette annotation en nous méfiant du découpage des tokens et des étiquettes de la catégorie grammaticale dans les textes en chinois. À partir de cette annotation automatique, nous avons

effectué une annotation manuelle pour garantir la fiabilité de nos textes chinois annotés. SACR nous sert pour la majorité du traitement du corpus. Il permet d'annoter manuellement les expressions référentielles ainsi que les chaînes de référence en français et en chinois ; ses opérations efficaces et ergonomiques facilitent le regroupement de plusieurs étapes d'annotation dans un seul outil. CRViewer permet ensuite de calculer les paramètres relatifs aux expressions référentielles et aux chaînes de référence à partir des textes annotés par SACR en diverses formes de présentation en chiffres absolus et en proportions relatives. Dans la partie suivante, nous présenterons la démarche de traitement de texte et de constitution de notre corpus.

Chapitre 5 : Processus du traitement outillé des expressions référentielles et des chaînes de référence en français et en chinois

Notre objectif est de constituer un *corpus annoté*[1] en expressions référentielles et en chaînes de référence, qui pourra servir de ressource pour les travaux ultérieurs relatifs au même domaine et à la méthode de constitution d'un corpus dans deux langues. La démarche comprend sept étapes divisées en trois phases, comme l'illustre la figure (5.1). Tous les travaux préparatoires sont inclus dans la phase 1, ce qui comprend : 1) le nettoyage des textes, 2) leur découpage en épisodes et 3) leur tokenisation automatique en chinois. La phase 2 est composée de tous les travaux d'annotations manuelles, ce qui inclut 4) celle des expressions référentielles et coréférentielles et 5) celle de leurs caractéristiques linguistiques. La phase 3 concerne les travaux évaluatifs, à savoir 6) une évaluation échantillonnée inter-annotateurs et 7) tous les travaux de l'évaluation quantitative à partir des annotations dans les textes annotés.

Figure (5.1) : Processus de constitution du corpus



5.1. Phase 1 – préparation

La première phase de la constitution de notre corpus comprend trois étapes qui concernent trois termes dans le domaine du traitement automatique des langues :

- le *nettoyage* des textes ;
- la *partition* des textes en épisodes ;
- la *tokenisation* automatique des textes en chinois.

[1] Un « corpus annoté » propose des données linguistiques plus enrichies par rapport à un « corpus brut », *i.e.* corpus sans annotation, *cf.* Hu & Peng (胡壮麟 & 彭宣维, 2013) et Zheng & Chang (郑茗元 & 常霜林, 2015).

5.1.1. Étape 1 : nettoyage des textes

Les linguistes informaticiens parlent souvent du *nettoyage* des textes dans le domaine du traitement automatique des langues pour désigner en général la préparation des textes qui précède leur importation dans les outils de traitement informatique (*cf*. Zheng & Chang, 郑茗元 & 常霜林, 2015 ; Oberlé, 2016). Après avoir choisi SACR comme outil d'annotation manuelle pour diverses raisons présentées dans le *Chapitre 4*, le nettoyage se concrétise dans le traitement de notre corpus. Cette étape sert, d'une part, à convertir les textes au format adapté à l'outil, à savoir les textes bruts en format txt, et, d'autre part, pour SACR, à supprimer les parties périphériques qui ne concernent pas les expressions référentielles et coréférentielles, comme les avant-propos, les numéros et les notes de bas de page dans l'édition du roman-fleuve.

5.1.2. Étape 2 : *partition* des textes en *épisodes*

Le premier principe qui régit la constitution de notre corpus est de faire en sorte que sa taille soit le plus grande possible, et ce dans le but de collecter un nombre maximum de données et d'exemples ; les résultats quantitatifs et qualitatifs que nous en retirerons seront ainsi censés être plus représentatifs. Les textes sélectionnés concernent les quatre premières sections du début du roman-fleuve qui comptent au total 62 000 mots français et 89 000 caractères chinois ; la taille totale de notre corpus s'élève donc à 151 000 mots et caractères.

Si on le compare au corpus DEMOCRAT[2], défini comme « grand corpus » qui totalise 560 000 mots français et dont la constitution a mobilisé 48 participants (*cf*. Landragin, 2021), notre corpus atteint un quart de ce travail collectif, et sa réalisation est l'œuvre d'une seule personne. Nous considérons que cette taille suffit pour collecter les données souhaitées et les exemples représentatifs dans le but d'observer les chaînes de référence d'un roman-fleuve en français et en chinois. Nous empruntons le titre de l'œuvre pour nommer notre corpus. Les 20 000 premiers mots français du corpus *Jean-Christophe*, avec l'annotation en expressions référentielles et en chaînes de référence, font partie du corpus DEMOCRAT.

Les textes français sont déjà découpés par le romancier en quatre sections numérotées, dont la longueur est environ de 15 000 mots en moyenne. Parallèlement, les textes chinois sont divisés en quatre sections, dont la longueur moyenne est à peu près de 22 000 caractères. Cette taille paraît énorme pour SACR même si l'outil permet des opérations sur des textes longs ; néanmoins l'enregistrement des annotations dure quelques secondes à chaque opération manuelle, et, étant donné le nombre d'opérations à effectuer sur les textes, ce temps de saisie rendrait le travail inefficace et non ergonomique. Ce type de questions techniques ne viennent pas embarrasser les auteurs qui travaillent sur des textes relativement courts, comme les faits divers, les contes de fées et les recettes de cuisine. Ainsi, nous avons dû subdiviser le texte en morceaux plus courts, dont la taille est plus pertinente pour le traitement dans l'outil.

[2] https://www.ortolang.fr/market/corpora/democrat.

5.1.2.1. Trois principes de la partition des textes en épisodes

Le découpage de textes longs en courts morceaux pendant le traitement du corpus est appelé *partition* par Zimina-Poirot (2005 : 242), qu'il définit ainsi :

> *Division d'un corpus en « parties » constituées par des fragments de texte consécutifs, n'ayant pas d'intersection commune et dont la réunion est égale au corpus. (Zimina-Poirot, 2005 : 242)*

Cette opération est également utilisée dans de nombreuses autres études. Par exemple, les longs textes, dans le corpus DEMOCRAT, sont découpés en texte de *tailles homogènes* d'environ 10 000 mots (Landragin, 2021). En tant que méthodologie générale de la constitution des corpus, cette opération est suggérée par Pincemin (2012) dans le but de *neutraliser la taille des textes*, c'est-à-dire de découper les textes longs en taille plus petite et de combiner les textes courts en taille plus grande, ce afin d'obtenir des textes de longueurs similaires et de faciliter les comparaisons ultérieures des indices entre ceux-ci. Ayant les mêmes objectifs, nous avons choisi trois critères linguistiques pour effectuer la partition de nos textes.

- Critère typo-dispositionnel

La partition s'est faite en premier lieu selon les marqueurs typo-dispositionnels propres au roman-fleuve, comme le découpage en chapitres, les numéros des sections, les paragraphes et les changements de ligne. Selon Charolles (1988, 1995), ces découpages naturels peuvent être considérés comme des marqueurs importants dans la structure textuelle qui permettent aux lecteurs de mieux comprendre les textes. Il s'agit en effet de :

> *Les marques configurationnelles (alinéas, organisateurs métadiscursifs) qui délimitent au sein du continuum textuel des ensembles présentés par le locuteur comme constituant une ou plusieurs unités en regard d'un certain critère dispositionnel. (Charolles, 1995 : 128)*

Les termes sont différents chez les auteurs. Par exemple, Adam (2001) parle de la *segmentation typographique* ; Zimina-Poirot (2004) utilise le terme *partition naturelle* ; Schnedecker & Landragin (2014) proposent de nommer l'opération *découpage typographique*. Ces termes désignent en réalité le découpage des textes selon le même principe et en fonction des marqueurs typo-dispositionnels. Pour cette raison, la partition des textes de notre corpus ne se fait jamais au milieu d'un paragraphe.

- Critère sémantique

Si, d'un côté et selon le critère typo-dispositionnel, le roman-fleuve est découpé seulement en fonction des titres, des sous-titres et des numéros de son édition, il se peut qu'il demeure toujours très long, ce qui est le cas de *Jean-Christophe* ; mais, d'un autre côté, si le découpage s'effectue suivant la séparation en paragraphes, à savoir que chaque paragraphe devient un morceau du corpus, ces « morceaux de texte » seraient trop nombreux et les chaînes de référence difficiles à obtenir dans chaque paragraphe. Il convient donc de compléter le critère typo-dispositionnel par la mise en œuvre d'un deuxième principe, dit

sémantique, dans le but de découper les sections du roman-fleuve en « morceaux » plus petits et pour rassembler les paragraphes en plus grands « morceaux ». Ces morceaux de taille intermédiaire sont appelés *épisodes*. Cette notion d'*épisode* est utilisée dans le système des termes de Leblond (2016), qui définit *Jean-Christophe* en 10 *cycles romanesques*, correspondant à 10 volumes du roman-fleuve. Chaque cycle romanesque comporte quelques sections numérotées par le romancier. Par exemple, le volume 1 de l'œuvre est numéroté en trois sections dans l'édition de sa version originale. Ensuite, chaque section peut être découpée en quelques épisodes selon les marqueurs sémantiques situés au début des paragraphes. Ces marqueurs sont souvent des expressions spatio-temporelles. Ils sont définis par Charolles (1995) comme « les expressions introductrices de cadres de discours qui délimitent des domaines ou cadres (temporels, spatiaux, modaux, etc.) s'étendant parfois sur de vastes séquences » (1995 : 128) (*cf.* aussi Adam, 2001 ; Schnedecker & Landragin, 2014). Nous avons repéré des expressions spatio-temporelles et des marqueurs d'intégration linéaires au début des paragraphes, comme « ce jour-là », « enfin », « il y avait des moments … » et « assez longtemps ». Comme le montrent les extraits [1] – [4], ce type d'expressions constituent ensemble le critère sémantique du découpage.

[1] **Ce jour-là**, sa mère lui avait mis ses habits les plus propres, de vieux habits donnés, dont l'ingénieuse patience de Louisa avait su tirer parti.
[début de l'épisode 13]

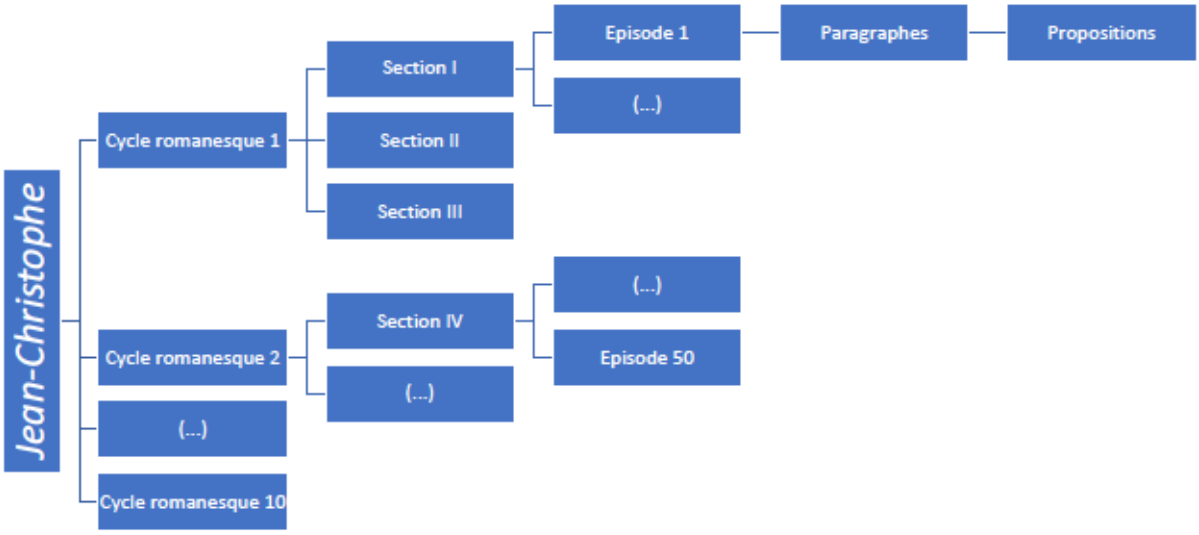
[2] **Enfin**, il arriva ; il monta en courant le vieil escalier noir, jusqu'à sa niche habituelle dans l'embrasure d'une fenêtre, au-dessus du fleuve.
[début de l'épisode 14]

[3] **Il y avait des moments de gêne très étroite à la maison**. Ils étaient de plus en plus fréquents.
[début de l'épisode 15]

[4] Christophe fut **assez longtemps** avant de s'apercevoir que son père buvait. L'intempérance de Melchior ne passait pas certaines limites, au moins dans les commencements.
[début de l'épisode 16]

De cette manière, les épisodes sont obtenus avec des marqueurs comme « frontières spatio-temporelles des événements » (Duranti, 1985 : 193). Cette hiérarchie des unités textuelles est illustrée dans la figure (5.2).

Figure (5.2) : Unités textuelles dans *Jean-Christophe* et dans notre corpus



- Critère de longueur

Selon le principe de neutralisation de la longueur des textes de Pincemin (2012), le dernier critère de l'opération de découpage est plus superficiel que les deux premiers puisqu'il concerne la longueur des textes. Nous avons fait en sorte que le découpage des textes aboutisse à des tailles qui soient similaires, sans pour autant omettre le caractère prioritaire des deux premiers critères.

5.1.2.2. Application des trois principes dans le corpus

Pour les textes en français, la partition de notre corpus utilise un des trois principes ou une combinaison de deux d'entre eux, voire des trois. Les textes en chinois sont découpés en parallèlement aux textes français. Le début et la fin des épisodes 3, 4 et 5 du corpus *Jean-Christophe* sont montrés dans les extraits [5] – [7]. Le début de chaque extrait est marqué par une expression concernant un temps (« la nuit » dans les trois extraits [5] – [7]) ou un endroit (« Saint-Martin » dans l'extrait [6]). Un épisode correspond à une scène dans le roman-fleuve : ce peut être une nouvelle histoire du personnage principal, l'introduction d'un nouveau personnage, un commentaire de l'écrivain, la description d'un paysage, entre autres. Par exemple, avec le début et la fin des trois extraits successifs dans les extraits [5] – [7], nous constatons clairement le changement de scène qui s'opère entre les épisodes : l'épisode 3 de l'extrait [5] commence par une scène de « la nuit » et finit par « rien ne peut l'apaiser » ; l'histoire se concentre toujours sur le personnage principal de l'épisode ; l'épisode 4 de l'extrait [6] reprend par « les cloches de Saint-Martin … », qui n'a plus rien à voir ni avec le personnage, ni avec l'action de l'épisode précédent. Le même type de changement de scène est facile à remarquer entre la fin de l'épisode 4 de l'extrait [6] et le début de l'épisode 5 de l'extrait [7] : le premier se termine par « leurs voix profondes et familières chantent » et le dernier continue sur la description d'un environnement (« la nuit … une pâle lueur blanchit la vitre »).

[5] **La nuit** *début* était tout à fait venue. […] // Il continue toujours sa plainte entrecoupée. On dirait que cette *fin*
misérable masse inconsciente et informe a le pressentiment de la vie de peines qui lui est réservée. Et **rien ne peut l'apaiser**…

[début et fin de l'épisode 3]

[6] **Les cloches de Saint-Martin** *début* chantèrent dans **la nuit**. Leur voix était grave et lente. […] // Si loin *fin*
qu'il se souvienne, – dans les lointains du temps, à quelque heure de sa vie que ce soit, – toujours **leurs voix profondes et familières chantent…**

[début et fin de l'épisode 4]

[7] **La nuit** *début* – à demi endormi… **Une pâle lueur blanchit la vitre**… […] // Pour l'ouvrir, il monte sur une *fin*
chaise : tout s'écroule, il se fait très mal, il hurle ; et par-dessus le marché, il est encore fouetté. Il est toujours fouetté !…

[début et fin de l'épisode 5]

La partition des textes a pour objectif de les découper en sections ou épisodes de longueur plus ou moins égale. Après cette opération, les textes sont découpés en 50 épisodes dans chaque langue ; la longueur de chaque épisode est limitée entre 200 à 2400 mots français.

Les épisodes équivalents en chinois sont découpés de la même manière, comme le montrent les extraits parallèles chinois [8] – [10] des épisodes 3, 4 et 5 :

[8] 天色 全 黑 了。[…] // 他 老是 断断续续的 悲啼。 仿佛 这 一 堆 无意识的
tianse quan hei le ta laoshi duanduanxuxude beiming fangfu zhe yi dui wuyishide
ciel tout noir LE 3P toujours continu se-plaindre CMP ce un CL inconscient
début *fin*

尚未成形的 肉， 对 他 命中注定的 痛苦的 生涯 已经 有 了 预感。 他
shangweixingchengde rou dui ta mingzhongzhudingde tongkude shengya yijing you le yugan ta
informe masse à son réservé misérable vie déjà avoir LE pressentiment 3P

怎么也 静不下来 ……
zenmeye jingbuxialai
comment s'apaiser-NEG

[début et fin de l'épisode 3]

[9] 黑夜 里 传来 圣·马丁 寺 的 钟声。 严肃 迟缓的 音调， […] // 不论 你
Heiye li chuanlai Shengmading si de zhongsheng yansu chihuande yindiao bulun ni
nuit pendant passer Saint-Martin église DE cloche-voix grave lent voix peu-importe 2P
début *fin*

回溯 到 如何 久远， — 不论 你 在 辽远的 时间 中 想到 你 一 生 的
huisu dao ruhe jiuyuan bulun ni zai liaoyuande shijian zhong xiangdao ni yi sheng de
se-souvenir à comment loin n'importe 2P PREP lointain temps pendant penser ton un vie DE

哪 一 刻， — 永远 是 它们 深沉 而 熟悉的 声音 在 歌唱 ……
na yi ke yongyuan shi tamen shenchen er shuxide shengyin zai changge
quel un moment toujours être leur profond et familier voix en-train-de chanter

[début et fin de l'épisode 4]

[10] 夜 里， — 半睡半醒的 时候 …… 一 线 苍白的 微光 照 在 窗 上 …… […] // 他
ye li banshuibanxingde shihou yi xian cangbaide weiguang zhao zai chuang shang ta
nuit pendant demi-endormi temps un CL pâle lueur blanchir PREP vitre dessus 3P
début *fin*

爬上 椅子 开 门， 连 人 带 椅 的 滚了下来， 跌得 很 痛， 哇 的
pashang yizi kai men lian ren dai yi de gunlexialai diede hen tong wa de
monter chaise ouvrir porte avec personne avec chaise DE s'écrouler-LE tomber très mal ONOMA DE

一 声 叫起来； 结果 还 挨 了 一 顿 打。 他 老是 挨打的！ ……
yi sheng jiaoqilai jieguo hai ai le yi dun da ta laoshi aidade
un voix crier finalement encore recevoir LE un CL fouetter 3P toujours fouetté

[début et fin de l'épisode 5]

De cette manière, les textes sont découpés en épisodes parallèles dans les deux langues. Les expressions référentielles ainsi que les chaînes de référence sont censées être similaires dans chaque épisode de texte. La partition du roman-fleuve en épisodes peut faciliter l'observation de la répartition des chaînes de référence dans l'ensemble de l'œuvre et dans chaque épisode du corpus.

Prenons un épisode entier comme exemple. Le début de l'extrait fixe tout d'abord le cadre spatio-temporel de l'événement (« à quelque temps de là et à l'école »), et ensuite le personnage apparaît sous la forme pleine d'un nom propre (« Christophe »). Ces trois éléments établissent ensemble l'état initial d'une intrigue. En passant à l'état de transformation, la chaîne « Jean-Christophe » est maintenue par quelques déterminants possessifs qui introduisent des référents associatifs secondaires (« son temps » et « ses voisins »). La saillance du personnage principal est maintenue par la forme pronominale. L'état final est marqué par un changement de cadre spatio-temporel (« au cabinet noir » et « quelques instants après »), qui n'entraîne aucun changement notable dans la continuité de la chaîne principale.

[11] **À quelque temps de là**, **à l'école**, où [**Christophe**] passait [**son**] temps à regarder les mouches au plafond et à donner des coups de poing à [**ses**] voisins, pour les faire tomber du banc, le maître qui [**l'**]avait pris en grippe, parce qu'[**il**] remuait toujours, parce qu'on [**l'**]entendait toujours rire, et parce qu'[**il**] n'apprenait jamais rien, [**Ø**] fit une allusion inconvenante, un jour que [**Christophe**] s'était [**lui-même**] laissé choir, à certain personnage bien connu dont [**il**] semblait vouloir suivre brillamment les traces. Tous les enfants éclatèrent de rire ; et certains se chargèrent de préciser l'allusion, en

des commentaires aussi clairs qu'énergiques. [**Christophe**] se releva, rouge de honte, [**Ø**] saisit [**son**] encrier, et [**Ø**] le lança à toute volée à la tête du premier qu'[**il**] vit rire. Le maître tomba sur [**lui**] à coups de poing ; [**il**] fut fustigé, mis à genoux, et condamné à un pensum énorme. [**Il**] rentra chez [**lui**], blême, rageant en silence ; et [**il**] déclara froidement qu'[**il**] n'irait plus à l'école. On ne fit pas attention à [**ses**] paroles. Le lendemain matin, quand [**sa**] mère [**lui**] rappela qu'[**il**] était l'heure de partir, [**il**] répondit avec tranquillité qu'[**il**] avait dit qu'[**il**] n'irait plus. Louisa eut beau prier, crier, menacer : rien n'y fit. [**Il**] restait assis dans [**son**] coin, le front obstiné. Melchior [**le**] roua de coups : [**il**] hurla ; mais à toutes les sommations qu'on [**lui**] faisait après chaque correction, [**il**] répondait rageusement : « Non ! » On [**lui**] demanda au moins de dire pourquoi ; [**il**] serra les dents et [**Ø**] ne voulut rien dire. Melchior [**l'**]empoigna, [**le**] porta à l'école et [**le**] remit au maître. Revenu à [**son**] banc, [**il**] commença par casser méthodiquement tout ce qui se trouvait à [**sa**] portée : [**son**] encrier, [**sa**] plume, [**il**] déchira [**son**] cahier et [**son**] livre, – le tout d'une façon bien visible, en regardant le maître d'un air provocant. On [**l'**]enferma **au cabinet noir**. – **Quelques instants après**, le maître [**le**] trouva, [**son**] mouchoir noué autour du cou, tirant de toutes [**ses**] forces sur les deux coins : [**il**] tâchait de s'étrangler. Il fallut [**le**] renvoyer.

Considérons à présent l'extrait parallèle comme exemple d'un épisode en chinois. Le déroulement des chaînes manifeste quelques nuances par rapport à celles du français. Le cadre temporel est donné en premier lieu 'après-un-temps' *guo-le-yi-shang* 过了一晌. Le cadre spatial est exprimé dans la locution verbale intransitive 'aller-école' *shang-xue* 上学. La maintenance de la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫 est constituée par des pronoms 'il' *ta* 他 et des déterminants possessifs 'son' *ta(de)* 他 (的) , qui sont ensuite suivis par des zéros :

[12]过了 一 晌， [克利斯朵夫] 上学 了； [他] 老 望着 天花板 上 的 苍蝇， [Ø]
guole yi shang Kelisiduofu shangxue le ta lao wangzhe tianhuaban shang de cangying
après un temps Christophe aller-école LE 3P toujours regardant plafond dessus DE mouche
npr pro zéro

把 拳头 捶着 旁边 的 孩子， 推 在 地 下； [他] 动个 不停， [Ø] 笑个
ba quantou chuizhe pangbian de haizi tui zai di xia ta dongge buting xiaoge
BA poing battant voisin DE enfant pousser PREP terre dessous 3P se-remuant sans-cesse riant
pro zéro

不停， [Ø] 从来不 念书。有 一 天， [克利斯朵夫 自己] 摔 在 了 地 下，
buting conglaibu nianshu you yi tian Kelisiduofu ziji shuai zai le di xia
sans-cesse ne-jamais lire avoir un jour Christophe soi-même tomber PREP LE terre dessous
zéro npr

讨厌 [他] 的 老师 便 说 了 句 难听的 话 隐射 某 个 大家 知道 的
taoyan ta de laoshi bian shuo le ju nantingde hua yinshe mou ge dajia zhidao de
détester 3P DE maître donc dire LE CL inconvenant phrase faire-allusion certain CL on connaître DE
pro

人， 说 [他] 大概 要 青出于蓝的 走上 那 条 路 了。所有的 孩子 听着
Ren shuo ta dagai yao qingchuyulande zoushang na tiao lu le suoyoude haizi tingzhe
personne dire 3P peut-être vouloir brillamment allant ce-là CL trace LE tous enfant entendant
pro

都 哈哈大笑； 有些 同学 还 揭穿 隐喻， 加上 一些 又 明白 又 有分量的
dou hahadaxiao youxie tongxue hai jiechuan yinyu jiashang yixie you mingbai you youfenliangde
tous éclater-rire certain camarade encore révéler allusion chargeant des et claire et énergique

注解。 [克利斯朵夫] 爬起来， [Ø] 羞得 满脸通红， [Ø] 拿起 墨水瓶 对准 一 个
zhujie Kelisiduofu paqilai xiude manliantonghong naqi moshuiping duizhun yi ge
commentaire Christophe se-lever honteux rougissant saisissant encrier viser un CL
npr zéro zéro

正在笑的 人 扔 过去。老师 冲上来 就是 一 顿 拳头， 用 鞭子 抽 [他]，
zhengzaixiaode ren reng guoqu laoshi chongshanglai jiushi yi dun quantou yong bianzi chou ta
riant personne jeter AUX maître tomber ainsi un coup poing avec fouet fustiger 3P
pro

要 [他] 跪 在 地 下， 再加上 极重的 罚课。[他] 脸色 发了 青， [Ø]
yao ta gui zai di xia zaijiashang jizhongde fake ta lianse fale qing
demander 3P s'agenouiller PREP terre dessous ajoutant énorme pensum son visage devenant noir
pro dét-poss zéro

憋着 一 肚子 怨气 回家， [Ø] 冷冷的 说 [他] 再也不 上学 了。家里人 并 没 把
biezhe yi duzi yuanqi huijia *lenglengde shuo ta zaiyebu shangxue le jialiren bing mei ba*
tolérant un CL rage rentrer (zéro) froidement dire 3P (pro) ne-plus aller-école LE famille même NEG BA

[他的] 话 放 在 心 上。明天 早上， 母亲 提醒 [他] 该 上学 了， [他]
tade hua fang zai xin shang mingtian zaoshang muqin tixing ta gai shangxue le ta
son (dét-poss) parole poser PREP cœur dessus demain matin mère rappeler 3P (pro) devoir aller-école LE 3P (pro)

却 安安静静的 回答， [他] 早 说过 不 去的 了。鲁意莎 对 [他] 软骗 硬吓 都
Que ananjingjingde huida ta zao shuoguo bu qude le Luyisha dui ta ruanpian yingxia dou
mais tranquillement répondre 3P (pro) déjà dire NEG allant LE Louisa PREP 3P (pro) prier menacer tout

没用。 [他] 坐 在 一 角， [Ø] 死赖 在 那里。曼希沃 揍 [他]， [他] 就 直嚷；
Meiyong ta zuo zai yi jiao silai zai nali Manxiwo zou ta ta jiu zhirang
avoir-beau 3P (pro) s'asseoir PREP un coin (zéro) obstiné PREP là-bas Melchior rouer 3P (pro) 3P (pro) donc hurler

每 次 揍过了 叫 [他] 上学， [他] 总是 火气 更大的 回答 一 声“ 不 去！”
mei ci zouguole jiao ta shangxue ta zongshi huoqi gengdade huida yi sheng bu qu
chaque fois rouer-LE demander 3P (pro) aller-école 3P (pro) toujours rage plus répondre un coup NEG aller

人家 要 [他] 至少 说出 理由 来， [他] 却 咬紧 牙关， 死 不 开口。曼希沃
renjia yao ta zhishao shuochu liyou lai ta que yaojin yaguan si bu kaikou Manxiwo
on vouloir 3P (pro) au-moins dire raison AUX 3P (pro) mais serrer dent mort NEG parler Melchior

抓着 [他] 硬 到 学校 交给 老师。可是 [他] 一 到 座位 上， 就 有计划的 毁坏
zhuazhe ta ying dao xuexiao jiaogei laoshi keshi ta yi dao zuowei shang jiu youjihuade huihuai
rattrapant 3P (pro) forcer à école laisser maître mais 3P (pro) dès arriver place dessus donc planifié détruire

手头 所有的 东西： 墨水瓶， 笔， 练习簿， 书本， 而且 [Ø] 故意 做得 教
shoutou suoyoude dongxi moshuiping bi lianxice shuben erqie guyi zuode jiao
à-portée tout chose encrier stylo cahier livre et (zéro) intentionnellement faisant laisser

人 看见， [Ø] 带着 挑战的 意味 望着 老师。结果 [他] 被 关进 黑 房。— 过了 一会，
ren kanjian daizhe tiaozhande yiwei wangzhe laoshi jieguo ta bei guanjin hei fang guole yihui
on voir (zéro) avec provocant air regardant maître enfin 3P (pro) par fermer noir salle après un-temps

老师 发见 [他] 用 手帕 缚着 脖子， [Ø] 拼命 往 两 头 拉， [他] 要 把
laoshi faxian ta yong shoupa fuzhe bozi pinming wang liang tou la ta yao ba
maître trouver 3P (pro) avec mouchoir nouant cou (zéro) s'efforcer vers deux bout tirer 3P (pro) vouloir BA

[自己] 勒死。 人家 只得 打发 [他] 回去。
ziji leisi renjia zhide dafa ta huiqu
soi-même (pro) s'étrangler on falloir renvoyer 3P (pro) rentrer

5.1.3. Étape 3 : tokenisation des textes en chinois

La *tokenisation* est un terme informatique que nous avons déjà présenté dans la section *4.1.1. Particularité du traitement des textes en chinois : tokenisation* du *Chapitre 4*. Cette étape est effectuée pour traiter spécifiquement les textes chinois dans notre corpus. Ces derniers sont importés dans AiHanYu ; l'outil les tokénise et les catégorise automatiquement, en exportant les textes tokenisés avec une étiquette des catégories grammaticales sur chaque token. Ayant évalué le taux d'imprécision de l'annotation automatique d'AiHanYu, nous avons constaté que la tokenisation et la catégorisation donnent des erreurs systématiques et particulières qui doivent ensuite être corrigées par l'annotation manuelle. Malgré ces erreurs, cette première tokenisation peut livrer des indications relatives

à la délimitation manuelle des expressions référentielles dans les étapes suivantes. Donc, à partir de cette annotation automatique, nous effectuerons une annotation manuelle de tout le corpus pour en garantir la fiabilité.

5.2. Phase 2 – annotation manuelle d'expressions référentielles et de chaînes de référence

L'*annotation* est une opération qui permet d'enrichir les informations relatives à des entités linguistiques spécifiques dans les corpus. Selon la définition de Habert (2005), elle consiste à attacher des étiquettes morpho-syntaxiques à chaque mot ou à chaque entité linguistique afin de classer et de détailler les informations linguistiques dans les corpus. Dans le présent travail, l'annotation consiste spécifiquement à délimiter les expressions référentielles, à constituer des liens coréférentiels entre elles et à les étiqueter par des propriétés linguistiques qui nous intéressent. Ainsi, nous allons présenter dans cette partie les opérations d'annotation de manière détaillée ainsi que les définitions et les arguments linguistiques qui leur sont liés. Ces choix d'annotation seront illustrés par des exemples en français et en chinois issus de notre corpus *Jean-Christophe*, accompagnés de nos remarques. Nous optons pour une méthode d'annotation majoritairement manuelle à l'aide d'outils semi-automatiques afin d'obtenir des résultats fiables, exploitables et réutilisables potentiellement pour des études ultérieures. Même si l'annotation manuelle est chronophage, les annotateurs humains disposent d'un bon niveau de connaissances linguistiques et ils sont en principe capables d'éviter les erreurs systématiques caractéristiques de l'annotation automatique des machines (*cf.* Mathet & Widlöcher, 2016). De ce fait, les conventions d'annotation nous paraissent cruciales en ce qui concerne les annotations manuelles. Elles aident à éviter les variations, les interprétations biaisées des annotateurs et les incertitudes pendant l'annotation (*cf.* Mélanie-Becquet & Landragin, 2014). Les annotateurs sont censés respecter les mêmes conventions du début d'une annotation manuelle jusqu'à sa fin. À notre connaissance, les conventions d'annotation des expressions référentielles et des chaînes de référence dans les corpus franco-chinois n'existent que dans peu d'études récentes, comme la thèse de Hou (2021) qui a été soutenue après la fin de l'annotation de notre corpus. Quatre conventions d'annotation nous servent de sources principales. Les deux que nous utilisons pour le chinois sont le manuel d'annotation du programme *Automatic Content Extraction*[3] (Doddington *et al.*, 2004) et *OntoNotes Chineses Coreference Guidelines*[4] (2008) ; pour le français, il s'agit du *Manuel d'annotation du projet MC4*[5] (2012) et du *Manuel DEMOCRAT* (2018). Nous présenterons d'abord les principes de délimitation des expressions référentielles et coréférentielles, puis nous détaillerons les conventions pour les étiqueter par les propriétés linguistiques selon différentes approches.

[3] Ci-après *ACE*.

[4] Ci-après *Guidelines*.

[5] Ci-après *Manuel MC4*.

5.2.1. Annotation des *mentions* : *singletons* et *maillons*

Une expression référentielle est appelée une *mention* (*cf*. Charolles, 2002 ; *Guidelines*, 2008). Ses catégories grammaticales principales en français et en chinois sont les noms propres, les SN, les pronoms et les zéros. Toutes les mentions coréférentielles sont nommées par le nom du référent, qui est ainsi le nom de la chaîne.

L'objectif de notre annotation est de repérer les expressions coréférentielles qui constituent des chaînes. Il est probable que certaines expressions référentielles ne coréfèrent à aucune autre expression et que certaines autres coréfèrent à une ou plusieurs autres expressions. Les premières renvoient à des *référents isolés* (Chastain, 1975 : 214) et sont appelés *singletons* (entre autres, Manuélian & Schnedecker, 2021) dans l'annotation ; les secondes sont appelées *maillons* des chaînes de référence. Nous avons choisi d'annoter tous les singletons et tous les maillons parce que toute expression référentielle peut éventuellement devenir le composant d'une chaîne dans la suite du texte. Ce choix d'annotation nous paraît surtout important pour notre corpus issu d'un roman-fleuve, puisqu'il peut contenir une proportion moins importante de singletons par rapport aux autres genres de textes. Comme le montrent Manuélian & Schnedecker (2021) dans le tableau (5.1), parmi les romans, les articles journalistiques, les articles scientifiques et les textes juridiques, les expressions référentielles dans les romans présentent une proportion plus élevée de « premiers maillons » (35,1 %), ce qui implique que la probabilité que les expressions référentielles aient des suites d'expressions coréférentielles est plus grande dans les romans que dans les autres genres.

Tableau (5.1) : Proportion entre singletons / premiers maillons de suite coréférentielle dans le corpus de Manuélian & Schnedecker (2021 : 76) (nous soulignons)

Texte	Singletons	Premiers maillons
Proportion roman	**64,9 % (198)**	**35,1 % (107)**
Proportion presse	76,9 % (160)	23,1 % (48)
Proportion science	77,8 % (193)	22,2 % (55)
Proportion juridique	80,9 % (161)	19,1 % (38)
Proportion globale	74,2 % (712)	25,8 % (248)

Comme le montrent les extraits parallèles [13] et [14] en français et en chinois de *Jean-Christophe*, les expressions entre crochets en gras sont référentielles. Elles sont annotées dans notre corpus comme mentions : certaines sont des singletons et d'autres des maillons de chaîne de référence :

[13][**Le nouveau-né**] *CR-1-M-1* s'agite dans [[**son**] *CR-1-M-2* **berceau**] *singleton-1*. Bien que [**le vieux**] *CR-2-M-1* ait laissé, pour entrer, [[**ses**] **sabots**] *CR-2-M-2* *singleton-2* à [**la porte**], *singleton-3* [[**son**] *CR-2-M-3* **pas**] *singleton-4* a fait craquer [**le plancher**] *singleton-5* : [**l'enfant**] *CR-1-M-3* commence à geindre. [**La mère**] *singleton-6* se penche hors de [[**son**] *singleton-7* **lit**] *singleton-8*, afin de [**le**] *CR-1-M-4* rassurer ; et [**le grand-père**] *CR-2-M-4* allume [**la lampe**] *singleton-9* en tâtonnant, pour que [**le petit**] *CR-1-M-5* n'ait pas peur de [**la nuit**] *singleton-10*.

[14][初生的 婴儿] 在 [摇篮 里] 扭动。[老人] 进来 虽然 把 [木靴] 脱 在
chushengde yinger zai yaolan li niudong laoren jinlai suiran ba muxue tuo zai
nouveau-né bébé PREP berceau intérieur s'agiter vieux-homme entrer malgré BA sabot laisser PREP
CR-3-M-1 *singleton-11* *singleton-12* *singleton-13* *singleton-14*

[门 外]，[走路 的 时候] [地板] 还是 格格的响：[孩子] 哼啊嗐啊 的 哭 了。
men wai zoulu de shihou diban haishi gegedexiang haizi hengahaia de ku le
porte extérieur marcher DE temps plancher encore craquer enfant ONOMA DE pleurer LE
singleton-15 *singleton-16* *singleton-17* *singleton-18* *CR-3-M-2*

[母亲] 从 [床 上] 探出 [身子] 抚慰 [他]；[祖父] 摸索着 点起 [灯]
muqin cong chuang shang tanchu shenzi fuwei ta zufu tansuozhe dianqi deng
mère par lit dessus sortir corps assurer 3P grand-père tâtonnant allumer lampe
singleton-19 *singleton-20* *singleton-21* *singleton-22* *CR-3-M-3* *singleton-23* *singleton-24*

来，免得 [孩子] 在 [[黑夜] 里] 害怕。
lai miande haizi zai heiye li haipa
AUX évitant enfant PREP nuit dedans avoir-peur
CR-3-M-4 *singleton-25* *singleton-26*

Toutes les annotations sont réalisées selon les mêmes conventions d'annotation dans les deux langues.

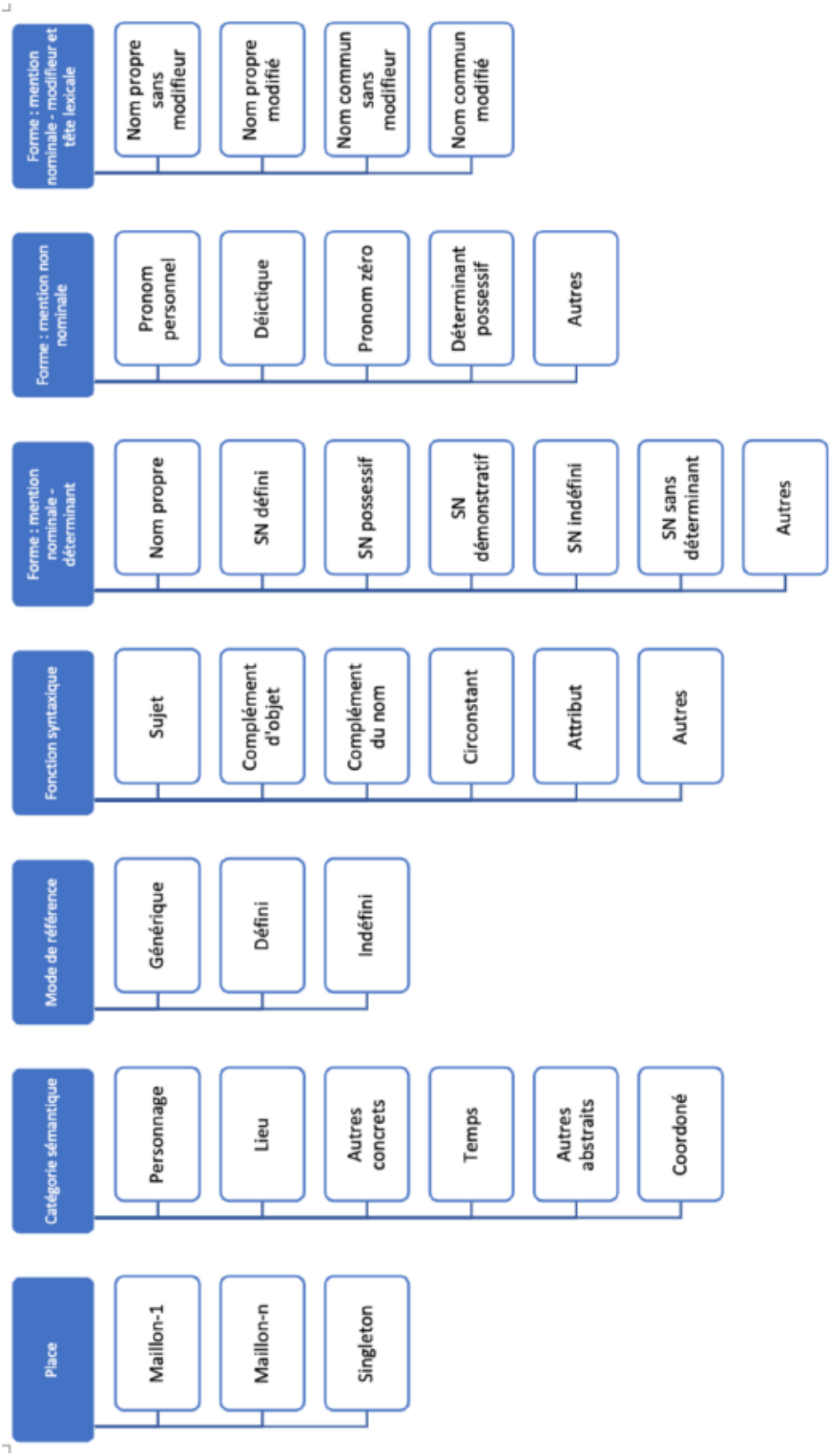
5.2.2. Annotation des caractéristiques des mentions

Nous avons annoté toutes les mentions selon sept critères :

- la place de la mention dans la chaîne
- la catégorie sémantique du référent
- le mode de la référence : générique, défini et indéfini
- la fonction syntaxique
- la forme : le déterminant de la mention nominale
- la forme : la mention non nominale
- la forme : le modifieur et la tête lexicale de la mention nominale

Le schéma d'annotation est visible dans la figure (5.3). Dans cette partie, nous présentons l'annotation de chaque propriété annotée et les raisons pour lesquelles nous avons décidé de les annoter.

Figure (5.3) : Schéma d'annotation des propriétés linguistiques des mentions



Toutes les caractéristiques et toutes les étiquettes concernant les mentions sont discutées dans cette partie, mais elles ne sont pas argumentées de la même manière : celles qui sont censées être fréquentes et nombreuses dans le corpus et dans les chaînes de référence sont présentées le plus précisément possible avec des exemples issus du corpus *Jean-Christophe*, comme « nom propre » et « zéro » ; en revanche, celles qui sont présumées être moins nombreuses sont abordées de manière brève, comme les « pronoms indéfinis » et les « pronoms possessifs ».

5.2.2.1. Place de la mention dans la chaîne

Les mentions annotées concernent toutes les expressions référentielles, dont certaines sont supposées constituer des chaînes et certaines autres n'en constituent pas.

Les expressions des référents isolés sont annotées par « singleton ».

Les maillons des chaînes se distinguent, selon Chastain (1975), Corblin (1995) et Cornish (1998), par leur position « initiale » et « non initiale » en raison des caractéristiques sémantiques et discursives particulières des premiers maillons, qui servent à introduire un nouveau référent dans le discours et de la source référentielle (*cf.* Corblin, 1995 ; Cornish, 1998), alors que les maillons suivants servent à maintenir ce référent dans le discours.

> *Le concept de chaîne autorise à distinguer des positions selon leur place, minimalement des positions initiales (positions-I) et des positions non-initiales (positions-N). Les deux types de chaînes6 pourraient être distingués selon les positions que différentes expressions peuvent y occuper. (Corblin, 1995 : 153)*

Ainsi, nous avons annoté « maillon-1 » pour les premiers maillons des chaînes et « maillon-n » pour les maillons suivants, afin de distinguer les maillons des deux positions. Comme le montre l'extrait [15], les mentions comme « le grondement du fleuve », « du fleuve » et « la maison » sont annotées comme « singleton » ; nous n'avons pas repéré d'expressions coréférentielles dans le même épisode. Les mentions dans les chaînes sont annotées par « maillon-1 » et « maillon-n », comme « le nouveau-né » (« maillon-1 »), « son » (« maillon-n »), « le vieux » (« maillon-1 ») et « ses » (« maillon-n »). L'extrait [16] contient les singletons, les premiers maillons et les maillons de suites parallèles en chinois.

[15][**Le grondement** [**du fleuve**]] monte derrière [**la maison**]. [**La pluie**] bat [**les carreaux**] depuis
singleton *singleton* *singleton* *singleton* *singleton*

[**le commencement** [**du jour**]]. [**Une buée d'**[**eau**]] ruisselle sur [**la vitre** [**au coin fêlé**]]. [**Le jour**
singleton *singleton* *singleton* *singleton* *singleton* *singleton* *singleton*

jaunâtre] s'éteint. Il fait tiède et fade dans [**la chambre**].
singleton

[**Le nouveau-né**] s'agite dans [[**son**] **berceau**]. Bien que [**le vieux**] ait laissé, pour entrer, [[**ses**]
maillon-1 *maillon-n* *singleton* *maillon-1* *maillon-n*

sabots] à [**la porte**], [[**son**] **pas**] a fait craquer [**le plancher**] : [**l'enfant**] commence à geindre.
singleton *singleton* *maillon-n* *singleton* *singleton* *maillon-n*

[16][江声]	浩荡，	自	[[屋]	后]	上升。	[雨水]	[整天]	的	打	在	[[窗]	上]。	[一
jiangsheng	*haodang*	*zi*	*wu*	*hou*	*shangsheng*	*yushui*	*zhengtian*	*de*	*da*	*zai*	*chuang*	*shang*	*yi*
fleuve-voix	gronder	par	maison	derrière	monter	pluie	tout-jour	DE	battre	PREP	fenêtre	dessus	un
singleton			*singleton*	*singleton*		*singleton*	*singleton*				*singleton*	*singleton*	*singleton*

[6] Les deux types de chaînes sont *chaînes anaphoriques* et *chaînes de référence*.

层 水雾] 沿着 [[玻璃] 的 裂痕] 蜿蜒流下。[昏黄的 天色] 黑下来 了。[[室] 内]
ceng *shuiwu* *yanzhe* *boli* *de* *liehen* *wanyanliuxia* *hunhuangde* *tianse* *heixialai* *le* *shi* *nei*
couche buée suivant vitre DE coin-fêlé ruisseler jaunâtre jour s'éteindre LE chambre intérieur
singleton *singleton* *singleton* *singleton* *singleton* *singleton*

有 [股 闷热 之 气]。
you *gu* *menre* *zhi* *qi*
avoir CL tiède-chaud DE air
singleton

[初生的 婴儿] 在 [[摇篮] 里] 扭动。[老人] 进来 虽然 把 [木靴] 脱 在 [门]
chushengde *yinger* *zai* *yaolan* *li* *niudong* *laoren* *jinlai* *suiran* *ba* *muxue* *tuo* *zai* *men*
nouveau-né bébé PREP berceau dedans s'agir vieux-homme entrer malgré BA sabot laisser PREP porte
maillon-1 *singleton* *singleton* *singleton* *singleton* *singleton*

外]，[走路 的 时候] [地板] 还是 格格的响：[孩子] 哼啊嗐啊 的 哭 了。
wai *zoulu* *de* *shihou* *diban* *haishi* *gegedexiang* *haizi* *hengahaia* *de* *ku* *le*
extérieur marcher DE temps plancher encore craquer enfant ONOMA DE pleurer LE
singleton *singleton* *singleton* *maillon-n*

5.2.2.2. Catégories sémantiques du référent

Les typologies des catégories ontologiques des référents sont nombreuses : les concrets et les abstraits, les animés et les inanimés, les humains et les non humains, les naturels et les fabriqués. Les référents sont des entités perçues par les humains comme *objet de pensée* (Dubois *et al.*, 2012 : 404). Cet objet peut manifester tout trait physique et imaginaire d'une entité perçue et reçue directement ou indirectement dans l'esprit, correspondant à une représentation concrète ou abstraite (*cf.* Charolles, 1988 : 8 ; 2002 : 19), comme les humains, les animés, les objets, les événements et les notions abstraites (Dubois & Lagane, 1973 : 39).

La catégorie sémantique du référent est annotée pour deux raisons. Premièrement, le degré de référenciation n'est pas le même selon les catégories sémantiques des entités : les objets et les humains sont les référents classiques consensuels que l'on rencontre le plus souvent ; néanmoins, il peut exister des entités moins typiques dans les romans, comme les notions, les repères temporels ou les quantités, et nous nous demandons si ces types de référents peuvent être candidats potentiels des chaînes. Deuxièmement, nous supposons que toutes les catégories de référent ne se présentent pas en chaîne de la même manière dans notre corpus. Par exemple, selon les résultats exposés dans le tableau (5.2) de l'étude de Manuélian & Schnedecker (2021), les « humains » sont censés se situer au centre des romans qui font plus souvent des chaînes (30,7 %) que des singletons ; les « objets » montrent aussi cette tendance mais de manière beaucoup moins forte (15,9 %), tandis que les référents, comme les « notions abstraites » et les « parties du corps », ont justement une tendance inverse (39,6 % et 13,9 %).

Tableau (5.2) : Catégories ontologiques des référents des singletons et des premiers maillons des suites coréférentielles dans les romans (Manuélian & Schnedecker, 2021 : 79)

	Singletons	Suites coréférentielles
Abstrait	**39,6 %**	11,4 %
Action / événement	5,9 %	3,4 %
Lieu	6,9 %	6,8 %
Objet	8,6 %	**15,9 %**
Partie du corps	**13,9 %**	11,7 %
Temps	10,7 %	11,4 %
Idéalité	4,3 %	4 %
Humain	5,3 %	**30,7 %**

Une observation primaire de *Jean-Christophe* révèle que cinq étiquettes de catégories sémantiques sont fixées comme suit, en combinant le cadre sémantique des noms de Flaux & Van de Velde (2000), de Gross (2012) et de la *Grande Grammaire du Français* (Abeillé & Godard, 2022) :

- les référents concrets sont subdivisés en trois catégories : « personnages » (« la mère »), « lieu » (« au bord du fleuve ») et « autres » (« la pluie ») ;
- les référents abstraits sont subdivisés en deux catégories : « temps » (« la nuit ») et « autres » (« sa peur ») ;
- les référents combinés hybrides forment le troisième groupe de catégories (« l'enfant et son piano »).

5.2.2.2.1. Référents concrets

Les référents concrets sont désignés par des noms concrets. Ce sont les référents que l'on peut percevoir directement par des gestes de compréhension reliée à la couleur, la forme, l'odeur, etc., des entités ; par exemple *voir* (« une pièce de théâtre »), *regarder* (« une peinture »), *toucher* (« un costume »), *sentir* (« un parfum ») et *écouter* (« une symphonie »), selon la définition de « concret » dans la *Grande Grammaire du Français* :

> *Les **noms concrets** décrivent des entités situées dans l'espace et directement perceptibles, c'est-à-dire que l'on peut voir, entendre, toucher, etc. Ils incluent les noms humains et sont le complément de verbes comme voir, entendre, ou toucher, sentir. D'autres noms comme idée, hypothèse, théorie, sont des **noms abstraits**.*
> *(Abeillé & Godard (éds), 2022 : 399, nous soulignons)*

Les référents concrets sont annotés par trois étiquettes : « personnage », « lieu » et « autres ».

- « Personnage »

Parmi les référents concrets, les personnages sont annotés par une étiquette à part en raison de leur grand nombre dans le corpus ainsi que de leur importance dans les romans par rapport aux autres genres (*cf.* Manuélian & Schnedecker, 2021). Il peut s'agir de personnages fictifs créés dans le cadre du roman-fleuve et / ou de personnages réels de l'histoire. Comme le montrent les extraits suivants, la narration romanesque est un mélange de trois

personnages du roman-fleuve : « grand-père » (« personnage-1 » de [17] et [18]), « le » (« personnage-3 » de [17]) et un personnage historique (« personnage-2 » de [17] et [18]).

[17][**Grand-père**] [**l'**] avait connu. [**Il**] avait failli se battre contre [**lui**]. Mais [**il**] savait reconnaître la grandeur de [**ses**] adversaires ; [**il**] [**l'**] avait dit vingt fois : [**il**] eût donné un de [**ses**] bras, pour qu'[**un tel homme**] fût né de ce côté du Rhin. Le sort [**l'**] avait voulu autrement : [**il**] [**l'**] admirait, et [**il**] [**l'**] avait combattu, – c'est-à-dire qu'[**il**] avait été sur le point de [**le**] combattre. Mais comme [**Napoléon**] n'était plus qu'à dix lieues, et qu'ils marchaient à [**sa**] rencontre, une subite panique avait dispersé la petite troupe dans une forêt, et chacun s'était enfui en criant : « Nous sommes trahis ! »

personnage-1 *personnage-2* *personnage-1* *personnage-2* *personnage-1* *personnage-1* *personnage-3* *personnage-1* *personnage-1* *personnage-2* *personnage-1* *personnage-1* *personnage-2* *personnage-1* *personnage-2* *personnage-2*

[18][祖父] 曾经 认识 [拿破仑]，差点儿 和 [他] 交战。但 [他] 是 赏识
zufu *cengjing* *renshi* *Napolun* *chadianer* *he* *ta* *jiaozhan* *dan* *ta* *shi* *shangshi*
grand-père jadis connaître Napoléon faillir avec 3P se-battre mais 3P être apprécier
personnage-1 *personnage-2* *personnage-2* *personnage-1*

[敌人] 的 伟大 的，[他] 说 过 几 十 遍：[他] 肯 牺牲 一 条
diren *de* *weida* *de* *ta* *shuo* *guo* *ji* *shi* *bian* *ta* *ken* *xisheng* *yi* *tiao*
adversaire DE grandeur DE 3P dire AUX quelque dizaine fois 3P préférer consacrer un CL
personnage-2 *personnage-1* *personnage-1*

手臂，要是 [**这样 一 个 人物**] 能够 生 在 莱茵河 的 这 一 边。可是
shoubi *yaoshi* *zheyang* *yi* *ge* *renwu* *nenggou* *sheng* *zai* *Laiyinhe* *de* *zhe* *yi* *bian* *keshi*
bras si ce un CL personnage pouvoir naître PREP Rhin DE ce un côté mais
personnage-2

天违人意：[**拿破仑**] 毕竟 是 法国人；于是 [祖父] 只 得 佩服 [他]，和
tianweirenyi *Napolun* *bijing* *shi* *faguoren* *yushi* *zufu* *zhi* *de* *peifu* *ta* *he*
sort-autre Napoléon malgré-tout être français donc grand-père seulement devoir admirer 3P avec
personnage-2 *personnage-2* *personnage-2*

[他] 鏖战，—就是说 差点儿 和 [**拿破仑**] 交锋。当时 [**拿破仑**] 离开
ta *aozhan* *jiushishuo* *chadianer* *he* *Napolun* *jiaofeng* *dangshi* *Napolun* *likai*
3P combattre c'est-à-dire sur-le-point avec Napoléon combattre à-ce-moment-là Napoléon distancier
personnage-2 *personnage-2* *personnage-2*

[祖父] 的 阵地 只 有 四十多 里， 祖父 他们 是 被 派去 迎击
zufu *de* *zhendi* *zhi* *you* *sishiduo* *li* *zufu* *tamen* *shi* *bei* *paiqu* *yingji*
grand-père DE champs seulement avoir quarantaine demi-kilomètre grand-père 3Ps être par envoyer attaquer
personnage-1

的，可是 那 一 小 队 人马 忽然 一 阵 慌乱， 往 树林 里 乱窜，大家
De *keshi* *na* *yi* *xiao* *dui* *renma* *huran* *yi* *zhen* *huangluan* *wang* *shulin* *li* *luancuan* *dajia*
DE mais ce-là un petit troupe armée subite un CL panique vers forêt dedans disperser tout-le-monde

一边 逃 一边 喊：“我们 上当 了！”
yibian *tao* *yibian* *han* *women* *shangdang* *le*
et s'enfuir et crier 1Ps être-trahi LE

□ « Lieu »

Les lieux sont un autre sous-type de référents concrets. Nous les annotons séparément des autres référents concrets étant donné leur rôle spécifique dans l'établissement et la fixation du cadre spatio-temporel des événements dans la narration (*cf.* Zhang & Xu, 张汉民 & 徐赳赳, 1992 ; Gao, 高彦梅, 2015). Il s'agit des expressions qui marquent les frontières des intrigues et des épisodes dans une narration longue. Or, les notions spatiales ne sont pas exprimées de la même manière en français et en chinois. Par exemple, comme le montrent les extraits parallèles [19] et [20], on dit « (passer) le seuil » (« lieu-3 ») en français mais '(traverser) la porte' *(kuajin)damen* (跨进) 大门 (« lieu-7 ») en chinois : les lieux ne sont pas les mêmes. Dans les extraits parallèles [21] et [22], le référent du lieu 'au-dessus-de son bâton

magique' *(zai)tade-moshu-gun-shang* (在) 他的魔术棍上 (« lieu-9 ») en chinois n'est exprimé en français que par un référent d'objet, « son bâton magique ».

[19][**La maison**] … [**le refuge contre tout ce qui est effrayant**] : l'ombre, la nuit, la peur, les choses
lieu-1 *lieu-2*

inconnues. Rien d'ennemi ne saurait passer [**le seuil**] … Le feu flambe. Une oie dorée tourne
lieu-3

mollement à la broche. Une délicieuse odeur de graisse et de chair croustillante embaume [**la chambre**].
lieu-4

[20][**家**] ····· 是 [抵御 一切 可怕的 东西 的 托庇所] 。 阴影， 黑夜， 恐怖， 不可知的 一切都 给
jia *shi* *diyu* *yiqie* *kepade* *dongxi* *de* *tuobisuo* *yinying* *heiye* *kongbu* *bukezhide* *yiqiedou* *gei*
maison être contrer tout effrayant chose DE refuge ombre nuit peur inconnu tout AUX
lieu-5 *lieu-6*

挡住 了 。 没有 一 个 敌人 能 跨进 [大门] ····· 炉火 融融， 金黄色的 鹅，
dangzhu *le* *meiyou* *yi* *ge* *diren* *neng* *kuajin* *damen* *luhuo* *rongrong* *jinhuangsede* *e*
arrêter LE sans un CL ennemi pouvoir traverser porte feu flambant doré oie
lieu-7

软绵绵的 在 铁串 上 转侧 。 满屋的 油 香 与 肉 香 。
ruanmianmiande *zai* *tiechuan* *shang* *zhuance* *manwude* *you* *xiang* *yu* *rou* *xiang*
mou PREP broche dessus tourner embaumant-chambre graisse odeur et chair odeur

[21]Il inventait d'attacher une vieille ficelle à son bâton magique, et il la jetait gravement dans [**le fleuve**],
lieu-8

attendant que le poisson vînt mordre.

[22]他 想出 把 一 根 旧 绳子 缚 在 [他的 **魔术 棍** 上] ， 一本正经的 丢 在
ta *xiangchu* *ba* *yi* *gen* *jiu* *shengzi* *fu* *zai* *tade* *moshu* *gun* *shang* *yibenzhengjingde* *diu* *zai*
3P inventer BA un CL vieux ficelle attacher à son magie bâton dessus gravement jeter à
lieu-9

[河 里] ， 等 鱼儿 来 咬。
he *li* *deng* *yuer* *lai* *yao*
fleuve dedans attendre poisson venir mordre
lieu-10

Nous supposons que le caractère variable de ces expressions spatiales peut provoquer des différences éventuelles quant aux chaînes de référence.

- « Autres concrets »

Dans la typologie des référents concrets de Flaux & Van De Velde (2000), deux types de référents et plusieurs de leurs sous-types sont regroupés sous cette étiquette : les référents « concrets dénombrables[7] », qui incluent les objets naturels animés, comme « un grillon », « des poissons », et inanimés, comme « une rivière », « les étoiles », les objets fabriqués physiques, comme « un bateau », « son piano » , et non physiques, comme « un discours », « une chanson », et les référents « concrets non dénombrables », comme « du gras », « de l'eau ». Ainsi, nous avons annoté toutes ces entités comme « autres concrets », comme on peut le voir dans les extraits [23] et [24].

[23]Gottfried lui parlait [**des étoiles**] et [**des nuages**] ; il lui apprenait à distinguer [**les souffles**
autres-concrets-1 *autres-concrets-2* *autres-concrets-3*

[**de** [**la terre**] **et de** [**l'air**] **et de** [**l'eau**]], [[**les chants**], [**les cris**], [**les bruits**] **du**
autres-concrets-4 *autres-concrets-5* *autres-concrets-6* *autres-concrets-7* *autres-concrets-8* *autres-concrets-9* *autres-concrets-10*

petit monde voletant, rampant, sautant ou nageant, qui grouille dans [**les ténèbres**]], et
autres-concrets-11

[7] Outre cette typologie des référents, voir aussi celle de Martin (1996) qui propose différents degrés entre les référents concrets et abstraits, que nous ne développerons pas ici.

[**les signes précurseurs de** [**la pluie**] **et** [**du beau temps**]], et [[**les instruments innombrables de**
autres-concrets-12 *autres-concrets-13* *autres-concrets-14* *autres-concrets-15*

[[**la symphonie de la nuit**]].
autres-concrets-16

[24]高脱弗烈特 和 他 谈着 [星辰], [云彩] , 教 他 辨别 [[泥土], [空气]
Gautuofuliete *he* *ta* *tanzhe* *xingchen* *yuncai* *jiao* *ta* *bianbie* *nitu* *kongqi*
Gottfried avec 3P parlant étoile nuage apprendre 3P distinguer terre air
autres-concrets-17 *autres-concrets-18* *autres-concrets-19* *autres-concrets-20*

和 [水]] 的 气息] , 辨别 在 [[黑暗] 中 飞舞 蠕动 , 跳跃
he *shui* *de* *qixi* *bianbie* *zai* *heian* *zhong* *feiwu* *rudong* *tiaoyue*
et eau DE souffle distinguer PREP ténèbres dedans voleter ramper sauter
autres-concrets-21 *autres-concrets-22* *autres-concrets-23* *autres-concrets-24* *autres-concrets-25*

浮游 的 万物] 的 [歌声] 、 [叫声] 、 [响声]] , 告诉 他 [[晴] [雨] 的
fuyou *de* *wanwu* *de* *gesheng* *jiaosheng* *xiangsheng* *gaosu* *ta* *qing* *yu* *de*
nager DE monde DE chant cri bruit dire 3P ensoleillé pluie DE
autres-concrets-26 *autres-concrets-27* *autres-concrets-28* *autres-concrets-29* *autres-concrets-30*

先兆] , [[夜 间 的 交响曲] 中 数不清的 乐器] 。
xianzhao *ye* *jian* *de* *jiaoxiangqu* *zhong* *shubuqingde* *yueqi*
signe-précurseur nuit pendant DE symphonie dedans innombrable instrument-de-musique
autres-concrets-31 *autres-concrets-32* *autres-concrets-33*

5.2.2.2.2. Référents abstraits

Au contraire des référents concrets, les référents abstraits ne sont pas directement perceptibles selon la couleur, la forme ou le goût. Nous les avons subdivisés en employant deux étiquettes, « temps » et « autres ».

- « Temps »

Les expressions temporelles renvoient à un moment ou à une durée qui forment un sous-type de référent abstrait, et elles sont annotées à part, tout comme les lieux, en raison de leur rôle spécifique dans l'établissement et la fixation du cadre spatio-temporel des événements dans la narration (*cf.* Zhang & Xu, 张汉民 & 徐赳赳, 1992 ; Gao, 高彦梅, 2015).

Ces expressions peuvent correspondre au *temps chronologique*, comme les mentions de date (« le 14 juillet »), de saison (« hiver ») et d'heure (« cinq heures du matin »), et au *temps non chronologique*, comme les durées (« quelques semaines »), les répétitions (« tous les lundis ») et les commencements (« le début de ») (*cf.* Gardes-Tamine, 2010 ; Gross, 2012). Comme le montrent les exemples d'annotation suivants, « le matin » (« temps-1 ») de [25] et 'matin' *qingchen* 清晨 (« temps-2 ») de [26] et « un soir » (« temps-3 ») de [27] et 'un soir' *yi-wan* 一晚 (« temps-5 ») de [28] expriment des temps chronologiques de moments ; « toute son enfance » (« temps-4 ») de [27] et 'son toute enfance' *ta-zhenggede- tongnianshidai* 他整个的童年时代 (« temps-6 ») de [28] expriment le temps non chronologique d'une durée :

[25][**Le matin**] … Ses parents dorment.
temps-1

[26][**清晨**] …… 父母 睡着。
qingchen *fumu* *shuizhe*
matin parent dormant
temps-2

[27]Le brillant météore, qui avait passé [**un soir**] dans le ciel de sa petite ville, eut une influence décisive
temps-3

sur l'esprit de Christophe. Pendant [**toute son enfance**], ce fut le modèle vivant, sur lequel il eut les
temps-4

yeux fixés.

[28]光明的 流星 在 小 城 的 天空 照耀 了 [一 **晚**] 之后， 克利斯朵夫 精神 上
guangmingde liuxing zai xiao cheng de tiankong zhaoyao le yi wan zhihou Kelisiduofu jingshen shang
brillant météore PREP petit ville DE ciel briller LE un soir après Christophe esprit dessus
temps-5

便 受到 确切不移的 影响。 在 [他 **整个的 童年时代**]， 哈斯莱 变成 他的 模范，
bian shoudao queqiebuyide yingxiang zai ta zhenggede tongnianshidai Hasilai biancheng tade mofan
donc avoir décisif influence PREP son entier enfance Hassler devenir son modèle
temps-6

他的 眼睛 始终 钉住 了 它。
tade yanjing shizhong dingzhu le ta
son yeux toujours fixer LE 3P

Au lieu d'exprimer le temps par la combinaison de verbes conjugués et d'expressions nominales temporelles, les locuteurs chinois ne disposent que de ces dernières, qui servent à marquer le temps au début ou à la charnière des événements. Les noms temporels en chinois sont susceptibles de jouer un rôle plus important que ceux du français étant donné l'absence de la conjugaison verbale dans cette langue.

- « Autres abstraits »

Mis à part le temps, les sous-types des référents abstraits sont nombreux, comme les *entités abstraites* (« la justice ») (*cf.* Landragin, 2016a), les *événements* (« le séisme », « la réponse », « la grippe ») (*cf.* Gross, 2012 : 243[8] ; Longo & Todirascu, 2014), les *qualités* (« la beauté », « la grandeur ») (*cf.* Dubois & Lagane, 1973 : 39) et les *actions* (« un acte », « une opération », « le nettoyage », « le départ ») (*cf.* Dubois & Lagane, 1973 : 39 ; Anscombre, 1986 : 11). Tous ces sous-types de référents abstraits sont annotés par la seule étiquette « autres abstraits », comme c'est le cas dans les extraits parallèles [29] et [30] :

[29]Jour immense, taciturne, que marque [**le rythme égal de l'ombre et de la lumière**], et [**le rythme**
abstrait-1 *abstrait-2*

de la vie de l'être engourdi qui rêve au fond de son berceau], – [**ses besoins impérieux,**
abstrat-3

douloureux ou joyeux, si réguliers que le jour et la nuit qui [**les**] **ramènent semblent ramenés**
abstrat-4

par [**eux**]].
abstrait-5

[30]在 摇篮 中 作梦 的 浑噩的 生物， 自有 [他 迫切的 **需要**]， [其中] 有
zai yaolan zhong zuomeng de hunede shengwu ziyou ta poquede xuyao qizhong you
PREP berceau dedans rêver DE engourdi être avoir son impérieux besoin dequel avoir
abstrait-6 *abstrait-7*

[**痛苦的**]， 也 有 [欢乐的]； 虽然 [**这些 需要**] 随着 昼夜 而 破灭， 但 [[它们]
tongkude ye you huanlede suiran zhexie xuyan suizhe zhouye er pomie dan tamen
douloureux aussi avoir joyeux mais ces besoin suivant jour-nuit et disparaître mais leur
abstrait-8 *abstrait-9* *abstrait-10* *abstrait-11*

整齐的 规律]， 反 象 是 昼夜 随着 [它们] 而 往复。
zhengqide guilv fan xiang shi zhouye suizhe tamen er wangfu
régulier rythme contrairement comme être jour-nuit suivant 3Ps et revnir
abstrait-12 *abstrait-13*

[8] Gross (2012 : 243) catégorise cinq types supplémentaires d'événements : (i) *fortuits* (« le séisme », « l'accident », « le gel », « la chute »), (ii) *organisés* (« le défilé », « la cérémonie », « la conférence »), (iii) *cycliques* (« l'anniversaire », « les Pâques »), (iv) les *actions événementielles* (« la réponse », « la signe », « l'attentat ») et (v) les *états événementiels* (« la grippe », « le retard », « le calme »).

Parfois les antécédents des référents abstraits sont longs et peuvent comporter une ou plusieurs phrases, voire un ou plusieurs paragraphes. Pour éviter de surcharger l'annotation, nous n'avons pas annoté les antécédents événementiels en phrases complètes. Nous considérons qu'une expression événementielle est devenue référentielle dès qu'elle est reprise par un pronom ; nous n'annotons que les reprises en forme pronominale. La même annotation est faite dans *Manuel DEMOCRAT* (2018) : les antécédents phrastiques ou paragraphiques ne sont pas annotés ; seules les reprises pronominales le sont, en tant que référentielles.

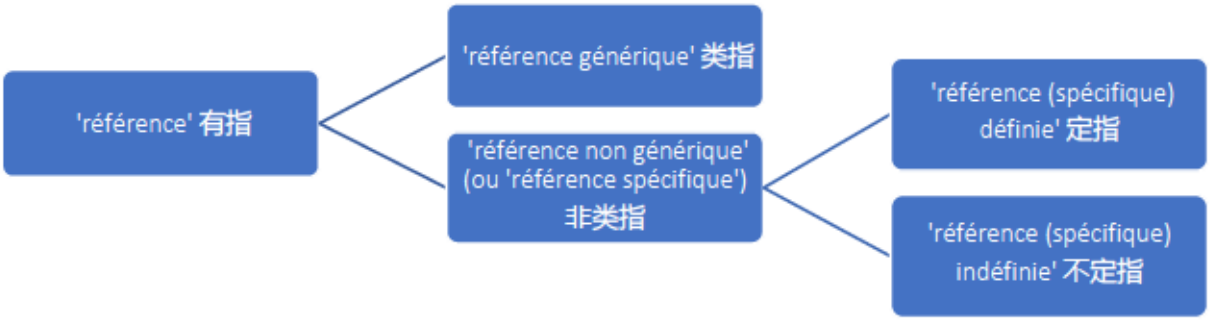
5.2.2.2.3. Référents coordonnés

Le terme de *référent coordonné* (Schnedecker, 2021 : 45) désigne une situation où plusieurs référents sont reliés par des coordinateurs comme « et », « avec » et « y compris » en français, et *he* 和 'et', *yiji* 以及 'ainsi que' et *baokuo* 包括 'y compris' en chinois. Dans notre annotation, cette étiquette est attribuée aux expressions qui coordonnent deux ou plusieurs référents de différentes catégories sémantiques. Par exemple, « grand-père et son piano » est annoté, d'abord, comme un « personnage » et un « référent concret », et, ensuite, comme un seul « référent coordonné ».

5.2.2.3. Mode de la référence

Les modes de référence sont présentés par diverses relations hiérarchiques selon les différents auteurs. Nous avons obtenu une typologie des modes de référence consensuelle pour le chinois et le français : celle de Dong (董秀芳, 2010) et de Xiong (熊岭) pour le chinois, et celle de la *Grammaire Méthodique du Français* (Riegel *et al.*, 2016) pour le français. Dong propose une typologie des modes de référence d'après l'observation de son corpus : la référence se subdivise en *référence générique* (类指) et *non générique* (非类指). Cette dernière est subdivisée en *référence définie* (定指) et *indéfinie* (不定指). La référence non générique est appelée *référence spécifique* dans d'autres ouvrages, comme *Dictionnaire d'Analyse du Discours* (Charaudeau & Maingueneau, 2002 : 489) et *Grammaire Méthodique du Français* (Riegel *et al.*, 2016 [1994] : 962). La hiérarchisation des modes de référence est présentée dans la figure (5.4) suivante avec notre traduction en français.

Figure (5.4) : Modes de référence en chinois de Dong (董秀芳, 2010, nous traduisons en français)



La *Grammaire Méthodique du Français* (Riegel *et al.*, 2016 [1994]) propose la définition suivante de la référence générique et de la référence spécifique :

> *La référence peut être :*
>
> *- **générique**, si l'on envisage la contrepartie référentielle de l'expression dans son extension maximale.*
>
> *- **spécifique**, si le référent est présenté comme existant et identifiable comme tel dans une situation donnée.*
>
> *(Riegel* et al.*, 2016 [1994] : 962)*

Au lieu de définir les trois modes par les aspects sémantiques, Xiong (熊岭, 2012), en citant Chen (陈平, 1987), les détermine selon les points de vue du locuteur et de l'interlocuteur :

> *"类指"的一类成员以一个抽象的整体出现，交谈者能够准确地把此类和彼类区分开来。*
>
> *La « référence générique » désigne une classe abstraite d'individus. L'interlocuteur arrive à distinguer les individus de cet ensemble de ceux d'un autre ensemble. (Xiong, 熊岭, 2012 : 34, nous traduisons)*
>
> *（"定指"标示）发话人使用某个名词性成分时，预料受话人能将所指对象与语境中某个特定事物等同起来，能把它与语境中某个可能存在的其他同类实体区分开。*
>
> *(La référence « définie » signifie que) Le locuteur utilise un SN en supposant que l'interlocuteur peut établir une équivalence avec un objet spécifique du contexte, et peut le distinguer des autres objets éventuels de la même catégorie sémantique dans le même contexte. (Chen, 陈平, 1987, cité par Xiong, 熊岭, 2012 : 34, nous traduisons)*
>
> *"不定指"是定指的反面，即发话人使用某个名词性成分时，预料受话人不能将所指对象与语境中某个特定事物等同起来，不能把它与语境中某个可能存在的其他同类实体区分开来。*
>
> *La référence « indéfinie » est le contraire de la référence définie, à savoir que le locuteur utilise un SN en supposant que l'interlocuteur ne peut pas établir d'équivalence avec un objet spécifique du contexte, et ne peut pas le distinguer des autres objets éventuels de la même catégorie sémantique dans le même contexte. (Xiong, 熊岭, 2012 : 35, nous traduisons)*

Par conséquent, nous choisissons d'annoter les expressions référentielles selon les trois modes de référence :

- « générique »
- « définie (spécifique) »
- « indéfinie (spécifique) ».

Les exemples de ces trois modes de référence sont mis en gras dans les extraits parallèles [31] et [32]. Par exemple, en français :

- « Jean-Michel » et « ses » réfèrent de manière spécifique et définie : ces expressions permettent d'identifier une seule personne ;
- « ses ambitions », « l'enfance » et « la musique » réfèrent de manière générique : ces expressions permettent seulement d'identifier un concept, une notion, une classe d'entités, sans pouvoir préciser un individu particulier de la classe ;
- « de grands dons pour la musique », « une facilité remarquable » et « de bonne heure » réfèrent de manière spécifique et indéfinie : ces expressions permettent d'identifier soit une entité au singulier sans préciser l'individu, soit un groupe d'entités en nombre indéterminé.

Les exemples chinois sont les suivants, entre autres :

- 'Jean-Michel' *Yuehan-Mixier* 约翰·米希尔, 'fils' *erzi* 儿子 et 'Melchior' *Manxiwo* 曼希沃 réfèrent de manière spécifique et définie ;
- 'ambition' *yexin* 野心, 'espoir' *xiwang* 希望 et 'musique' *yinyue* 音乐 réfèrent de manière générique ;
- 'tout le monde' *dajia* 大家 réfère de manière spécifique et indéfinie.

[31][**Jean-Michel**] avait reporté [[**ses**] **ambitions**] sur [[**son**] **fils**] ; et [**Melchior**] promit d'abord de
défini *défini* *générique* *défini* *défini* *défini*

[**les**] réaliser. [**Il**] avait, dès [**l'enfance**], [**de grands dons pour** [**la musique**]]. [**Il**] apprenait avec
défini *défini* *générique* *indéfini* *générique* *défini*

[**une facilité remarquable**], et [**de bonne heure**] [**il**] acquit, comme [**violoniste**], [**une virtuosité**
indéfini *indéfini* *défini* *générique* *indéfini*

qui fit de [**lui**] **pendant longtemps** [**le favori**], **presque** [**l'idole** [**des concerts de** [**la cour**]]].
défini *générique* *générique* *défini* *défini*

[32]于是 [**约翰·米希尔**] 把 [**野心**] 寄托 在 [[**儿子**] **身** 上] ; 而 [**曼希沃**] 最初 也 表现得
yushi *Yuehan-Mixier* *ba* *yexin* *jituo* *zai* *erzi* *shen* *shang* *er* *Manxiwo* *zuichu* *ye* *biaoxiande*
donc Jean-Michel BA ambition reporter PREP fils corps dessus et Melchior d'abord aussi promettant
défini *générique* *défini* *défini* *défini*

很 有 [**希望**] , [他] 从小 极 有 [[**音乐**] 天才] , [**学** 的 时候] 非常 容易 ,
hen *you* *xiwang* *ta* *congxiao* *ji* *you* *yinyue* *tiancai* *xue* *de* *shihou* *feichang* *rongyi*
très avoir espoir 3P dès-petit très avoir musique don apprendre DE temps très facile
générique *défini* *générique* *générique* *défini*

[[**提琴**] 的 **演技**] 很早 就 成熟 了 , [**大家**] 在 [[**音乐会**] 中] 捧 [他] ,
tiqin *de* *yanji* *henzao* *jiu* *chengshu* *le* *dajia* *zai* *yinyuehui* *zhong* *peng* *ta*
violon DE virtuosité tôt AUX maîtriser LE tout-le-monde PREP concert pendant apprécier 3P
générique *défini* *indéfini* *générique* *générique* *défini*

把 [他] 当做 [**偶像**] 。
ba *ta* *dangzuo* *ouxiang*
BA 3P comme idole
défini *générique*

5.2.2.4. Fonction syntaxique

La fonction syntaxique des expressions référentielles est beaucoup discutée dans les études, entre autres celles de Keenan & Comrie (1977) et de Rohder & Kehler (2014). Ces auteurs constatent des corrélations entre la saillance du référent et la fonction syntaxique de l'expression. Ils obtiennent des résultats plus ou moins différents selon leurs observations et leurs expérimentations. Les échelles de degré de référence de chacun d'eux sont les suivantes : les référents aux fonctions de sujet et de complément d'objet sont plus saillants que ceux qui sont aux autres fonctions syntaxiques :

> *Sujet > complément d'objet direct > complément d'objet indirect > complément oblique > complément de possession > construction comparative standard explicite*
>
> *(Keenan & Comrie, 1977 : 66)*

> *Sujet > complément d'objet > autre référents*
>
> *(Rohder & Kehler, 2014 : 919)*

Par ailleurs, étant donné les spécificités typologiques du français, le projet MC4 propose d'annoter les fonctions syntaxiques des expressions référentielles suivantes :

> *« Sujet », « complément d'objet direct », « complément d'objet indirect », « attribut », « complément du nom », « déterminant possessif », « circonstant », « apostrophe », « dislocation », « complément d'agent » et « autres » (notamment les termes isolés dans les titres)*

Le sujet et le complément d'objet sont mentionnés dans les échelles et la méthode d'annotation comme les places syntaxiques potentiellement prédisposées à avoir les expressions référentielles. Certaines autres fonctions sont également mentionnées d'après certains documents, comme le complément du nom, l'attribut, le circonstant et le déterminant possessif. Ces places contiennent aussi éventuellement des expressions référentielles.

Nous avons finalement choisi six fonctions pour annoter notre corpus franco-chinois : « sujet », « complément d'objet », « complément du nom », « circonstant » (de lieu et de temps), « attribut » (nominal) et « autres » (y compris toutes les autres fonctions). Ce choix a été effectué sur la base d'une combinaison des échelles susmentionnées, du *Manuel MC4* et de notre observation primitive du corpus *Jean-Christophe*.

- « Sujet »

Le sujet s'avère être la place préférable d'un nombre important d'expressions référentielles selon de nombreuses études comme celles de Keenan & Comrie (1977), Lyons (1978), Hobbs (1979), Kleiber (1981), Chen (陈平, 1987), Charolles (2002) et Rohder & Kehler (2014), entre autres.

Les échelles de Keenan & Comrie (1977) et de Rohder & Kehler (2014) mises à part, dans les tests cognitifs de Hobbs (1979), les résultats révèlent que le sujet a une plus forte tendance à être repris par rapport à celle du complément d'objet. Dans une phrase du type [33], 90 % des personnes considèrent que « he » coréfère à « John » à l'écrit et 75 % de personnes effectuent la même coréférence à l'oral, malgré les deux référents potentiels. Selon

le résultat de cette enquête, par rapport au complément d'objet, le sujet établit plus facilement la coréférence avec la reprise pronominale.

[33][John] can open [Bill]'s safe. [**He**] …
réf-1 *réf-2* *réf-?*

(Hobbs, 1979 : 80)

Pour expliquer cette prééminence du référent à la place du sujet, Rohde & Kehler (2014) considèrent que la position initiale du sujet devant les autres éléments syntaxiques permet de constituer un lien avec le contexte linguistique et extra-linguistique, surtout pour les pronoms en position de sujet, qui créent une forte tendance à la continuité référentielle par rapport à ceux qui figurent en position de complément d'objet ou occupent une autre fonction syntaxique. En raison de cette importance référentielle, les sujets exprimés ainsi que les sujets non exprimés sont annotés dans les deux langues.

Les sujets zéro coréfèrent souvent au sujet de la proposition précédente, comme le montrent « sujet-9 » et « sujet-10 » de l'extrait [35]. Ils se situent devant le verbe prédicatif dans les deux langues.

[34]– [**Père**], il doit être tard, disait affectueusement [**la jeune femme**]. Il faut rentrer chez vous, [**vous**]
sujet-1 *sujet-2* *sujet-3*
avez loin à aller.
– [**J'**]attends Melchior, répondit [**le vieillard**].
sujet-4 *sujet-5*

[35]"[父亲],	[时候]	不	早	了	吧, "	[妇]	恳切的	说。"[您]	得	回去	了,	[Ø]	还要
fuqin	*shihou*	*bu*	*zao*	*le*	*ba*	*fu*	*kenqiede*	*shuo*	*nin*	*dei*	*huiqu*	*le*	*haiyao*
père	temps	NEG	tôt	LE	INTERJ	femme	affectueusement	dire	vous	devoir	rentrer	LE	encore
sujet-6	*sujet-7*					*sujet-8*			*sujet-9*			*sujet-10*	

走	好	一	程	路	呢。"
zou	*hao*	*yi*	*cheng*	*lu*	*ne*
aller	bien	un	CL	chemin	INTERJ

"[我]	等着	曼希沃, "[老人]	回答。
wo	*dengzhe*	*Manxiwo*	*laoren*	*huida*
1P	attendant	Melchior	vieux-homme	répondre
sujet-11			*sujet-12*	

□ « Complément d'objet »

Le complément d'objet occupe la deuxième place dans les échelles de référence de Keenan & Comrie (1977) et de Rohder & Kehler (2014) ; sa possibilité de référence est moins forte que celle du sujet. Les compléments d'objet sont annotés comme une seule fonction syntaxique, sans distinction entre directs et indirects, en raison de leur forme morphologique identique en chinois (*cf*. Zhang, 张伯江, 1999). Comme le montre « CDO-1 » de l'extrait français [36], la forme du pronom « le » se distingue de celle du sujet « il », alors que ce type de distinction n'existe pas en chinois, comme 'le' *ta* 他 (« CDO-4 ») et 'il' *ta* 他 (« sujet ») de l'extrait chinois [37].

[36]Le chant des cloches [**le**] rendait très triste ; car il se rappelait [ses **espérances déçues**]. Il
CDO-1 *CDO-2*
pensait à [**ce qu'il faisait là**], à cette heure, dans la rue. Et, de honte, il pleurait.
CDO-3

[37]钟声	使	[他]	非常	悲伤,	因为	[他]	回想起	[幻灭的	希望]。	他	又	想到
Zhongsheng	*shi*	*ta*	*feichang*	*beishang*	*yinwei*	*ta*	*huixiangqi*	*huanmiede*	*xiwang*	*ta*	*you*	*xiangdao*
cloche-chant	rendre	3P	très	triste	car	3P	rappeler	déçu	espérance	3P	aussi	penser
		CDO-4				*sujet*		*CDO-5*				

此刻 冒 [雨] 街头 是 为的什么， 不禁 羞愧交迸的 哭 了。
cike *mao* *yu* *jietou* *shi* *weideshenme* *bujin* *xiukuijiaobengde* *ku* *le*
ce-moment sous pluie rue être pourquoi ne-pouvoir-s'empêcher-de honteux pleurer LE
CDO-6

□ « Complément du nom »

La *Grammaire Méthodique du Français* (Riegel *et al.*, 2016 [1994]) propose la formule suivante qui illustre la place du complément du nom en français dans une expression nominale, où « SN0 » signifie un SN modifié d'un complément du nom :

SN0 = (déterminant1) – tête lexicale – (déterminant2) – complément du nom

Les compléments du nom se situent toujours à droite du nom en français, contrairement à ceux du chinois qui sont toujours à gauche. En chinois, les expressions de cette fonction sont antéposées à la tête lexicale ; l'équivalent d'un SN modifié d'un complément du nom en chinois peut donc être formulé comme suit :

SN0 = (déterminant) – complément du nom – tête lexicale

Dans cette formule, le complément du nom est antéposé à la tête lexicale ; le déterminant peut seulement se situer devant le complément du nom au lieu de se situer après et avant la tête lexicale.

Les compléments du nom référentiels à gauche (en chinois) et à droite (en français) sont annotés dans notre corpus, comme le montrent les extraits [38] – [41].

[38]La flamme éclaire [la figure rouge [**du vieux Jean-Michel**]].
CDN

[39]灯光 照出 [[**老 约翰·米希尔**] 红红的 脸] 。
dengguang *zhaochu* *lao* *Yuehan-Mixier* *honghongde* *lian*
flamme éclairer vieux Jean-Michel rouge visage
CDN

[40]Les trois cloches tranquilles continuaient à sonner [la fête [**du lendemain**]].
CDN

[41]三 口 钟 庄严肃穆， 继续 在 那里 奏鸣， 报告 [[**明天**] 的 节日] 。
san *kou* *zhong* *zhuangyansumu* *jixu* *zai* *nali* *zouming* *baogao* *mingtian* *de* *jieri*
trois CL cloche sérieux-tranquille continuer PREP là-bas sonner raconter lendemain DE fête
CDN

□ « Circonstant »

Les circonstants peuvent concerner des candidats éventuels des référents, comme les lieux et les temps. Ils sont donc annotés dans le corpus.

[42]Il l'avait dit vingt fois : il eût donné un de ses bras, pour qu'un tel homme fût né de [**ce côté du Rhin**].
circonstant-1

'他 说 过 几十 遍： 他 肯 牺牲 一 条 手臂， 要是 这样 一 个 人物 能够
ta *shuo* *guo* *jishi* *bian* *ta* *ken* *xisheng* *yi* *tiao* *shoubi* *yaoshi* *zheyang* *yi* *ge* *renwu* *nenggou*
3P parler AUX des fois 3P vouloir donner un CL bras pour-que ce un CL homme pouvoir

生 在 [**莱茵河 的 这 一 边**] 。'
sheng *zai* *Laiyinhe* *de* *zhe* *yi* *bian*
naître PREP Rhin DE ce un côté
circonstant-2

Par rapport à ceux du français, les circonstants chinois comportent des spécificités relativement à leur place dans la phrase et à leur composition.

Quant à leur place, les circonstants cadratifs se situent au début des phrases existentielles en chinois ; en français, cette place est flexible (au début ou à la fin de la phrase) (*cf*. Lyons, 1974 [1967]). Comme le montre la phrase [43] de Lyons, la seule place acceptable du circonstant 'sur la table' *zhuo-shang* 桌上 est en son début ; autrement dit, la phrase [44] dont le circonstant est postposé à la fin, n'est pas acceptable en grammaire chinoise. Or, la traduction française peut avoir deux manières acceptables : le circonstant « sur la table » au début de la phrase ou à la fin, comme nous traduisons [44] et [45], malgré les focus différents des deux énoncés :

[43][**桌 上]　有 书。**
zhuo shang you shu
table dessus avoir livre
circonstant

[44]Il y a un livre sur la table.
[45]Sur la table, il y a un livre.

(Lyons, 1974 [1967] : 50, nous traduisons en français)

[46]*有 书 [**桌 上**]。
you shu zhuo shang
avoir livre table dessus
circonstant

Du point de vue de la composition, le mot de direction *shang* 上 signifiant « au-dessus de », « sur », « superposé », est souvent utilisé comme élément contracté ou suffixe postposé au nom du circonstant. Selon Guo (郭锐, 2002 : 210) et Xu (徐赳赳, 2018 : 40), le *shang* 上 'dessus' est classé comme *nom de direction* (方位名词) à la fois pour sa catégorie grammaticale et pour sa catégorie sémantique. Cette spécificité est également illustrée dans l'exemple [43] de Lyons avec l'explication suivante :

> *Le 上 « postpositionnel » (dessus, sommet, surface), appartient à l'ensemble des morphèmes qui sont utilisés dans les syntagmes locatifs. (Lyons, 1974 [1967] : 51)*

Nous résumons ce type de circonstants contractés chinois par la formule suivante : la préposition est suivie d'un circonstant qui contient un nom commun et un nom de direction / temps / lieu ; le nom général seul est annoté comme complément du nom :

Préposition – [[nom commun]$_{CDN}$ – nom de temps / lieu / direction]$_{circonstant}$

Les exemples de *shang* 上 dans le corpus *Jean-Christophe* sont annotés dans les extraits suivants :

[47]这 时　即使 [[**天**] 上]　打雷，孩子 也不会 听见。
zhe shi jishi tian shang dalei haizi yebuhui tingjian
ce temps même ciel dessus tonner enfant NEG entendre
CDN circonstant

'Le tonnerre pourrait tomber, l'enfant ne l'entendrait pas.'

[48][**窗**　上]　不时 还 有 一 声 两 声 的 雨点。
chuang shang bushi hai you yi sheng liang sheng de yudian
fenêtre dessus parfois encore avoir un CL deux CL DE pluie
CDN circonstant

'La vitre sonna une ou deux fois encore sous le doigt de la pluie.'

Le même type de contraction des expressions circonstancielles est observé avec les *noms de direction*, les *noms de temps* (时间名词) et les *noms de lieu* (处所名词). Ces trois catégories sémantiques sont également classées comme catégories grammaticales du chinois (*cf*. Guo, 郭锐, 2002 ; Xu, 徐赳赳, 2018).

Nous avons appliqué cette annotation à notre corpus, comme le montrent les exemples [49] et [50] : les trois circonstants de lieu 'berceau-intérieur' *yaolan-li* 摇篮里 (« circonstant-1 »), 'porte-extérieur' *men-wai* 门外 (« circonstant-2 ») et 'théâtre-intérieur' *xiyuan-li* 戏院里 (« circonstant-6 ») sont composés d'un nom général et d'un nom de lieu postposé. Les expressions sont annotées comme « circonstant » ; les noms généraux dans ces circonstants sont annotés comme « complément du nom ».

[49]初生的 婴儿 在 [[摇篮] 里] 扭动。老人 进来 虽然 把 木靴 脱 在
chushengde yinger zai yaolan li niudong laoren jinlai suiran ba muxue tuo zai
nouveau-né bébé PREP berceau intérieur s'agiter vieux-homme entrer bien-que BA sabot laisser PREP
CDN circonstant-1

[[门] 外]， [走路 的 时候] 地板 还是 格格的响， 孩子 哼啊嗐啊 的 哭 了。
men wai zoulu de shihou diban haishi gegedexiang haizi hengahaia de ku le
porte extérieur marcher DE temps plancher encore craquer enfant ONOMA DE pleurer LE
CDN circonstant-2 circonstant-3

'Le nouveau-né s'agite dans [**son berceau**]. Bien que le vieux ait laissé, pour entrer, ses sabots à
circonstant-4

[**la porte**], son pas a fait craquer le plancher ; l'enfant commence à geindre.'
circonstant-5

[50]我 想 他 在 [[戏院] 里] 罢。
wo xiang ta zai xiyuan li ba
1P croire 3P PREP théâtre intérieur INTERJ
CDN circonstant-6

'Je crois qu'il est [**au théâtre**].'
circonstant-7

- « Attribut »

L'annotation des attributs est discutable chez les sémanticiens et chez les informaticiens, qui préconisent les uns et les autres des attitudes distinctes.

Tout d'abord, les discussions linguistiques sur la fonction référentielle de diverses formes d'attributs sont nombreuses (*cf.* Riegel *et al.*, 2016 [1994] : 421), ce qui nous permet de choisir attentivement quels attributs sont à annoter et de quelle manière. Pour le français, Schnedecker (1997) propose la distinction entre les SN indéfinis attributifs et les SN sans déterminant attributifs comme suit :

> *« Être un N » décrit des propriétés constitutives ou durables du référent désigné par le SN sujet, aptes donc à constituer une classe.*
>
> *« Être ø N » prédique des états transitoires, précaires, correspondant à des « événements, à des faits spatio-temporellement dépendants, renvoyant à des points de référence dans le monde du discours donné ».*
>
> *(Schnedecker, 1997)*
>
> *Par contraste, celles dont le SN attribut a un déterminant indéfini sont considérées comme des constructions typantes ; elles « déterminent directement la classe des individus particuliers qui sont leurs occurrences » ou renvoient, le cas échéant, à des classes référentielles en relation d'inclusion. (Riegel, 1985, in Boone,* art. cit. *: 76, cité par Schnedecker, 2021 : 77)*

Pour le chinois, Zhang (张伯江, 1997) propose une distinction similaire des deux types d'attribut : les SN indéfinis attributifs tendent à référer une entité générique et exemplaire ; les SN sans déterminant attributifs tendent à désigner une qualité ou un état au lieu d'une entité telle quelle. En raison de ces deux typologies similaires pour les deux langues, les

référents des SN indéfinis attributifs sont annotés comme *génériques* ; les SN sans déterminant attributifs ne sont pas annotés comme expressions référentielles.

Ensuite, outre les discussions des linguistes, en ce qui concerne le traitement informatique des attributs, il existe deux types d'opinions relatives à leur annotation comme expressions référentielles ou non. D'un côté, sans parler de la fonction syntaxique d'*attribut* de phrase, les expressions *attributives* s'opposent déjà aux expressions référentielles dans leur définition (*cf.* Donnellan, 1966 : 285 ; Charolles, 2002 : 8) : les premières renvoient aux propriétés et les dernières aux entités. Certains auteurs, comme Hou (2021), choisissent de n'annoter aucun attribut nominal comme expression référentielle. D'un autre côté, de nombreuses études révèlent que les attributs nominaux ont une faible tendance à référer en français et en chinois par rapport aux autres fonctions syntaxiques (*cf.* Dubois & Lagane, 1973 ; *Guidelines*, 2008 ; Chen, 陈平, 2015 ; Li, 李先银, 2016). Cela montre justement que les attributs nominaux peuvent avoir une fonction référentielle. Selon le *Manuel DEMOCRAT* (2018), il est suggéré d'annoter tous les attributs nominaux comme expressions référentielles. Comme le montrent les exemples suivants, tous les attributs nominaux sont annotés comme expressions référentielles qui coréfèrent à leur sujet de phrase :

[51][Marie] est [**vendeuse dans une boulangerie**]. [Elle] est aussi [**la gérante d'une pizzeria**].
réf-1 *réf-1* *réf-1* *réf-1*

[52][Londres] est [**la capitale de l'Angleterre**].
réf-2 *réf-2*

[53]J'aime regarder [**l'étoile du matin**]. Mon frère préfère regarder [**l'étoile du soir**]. Nous savons pourtant que [l'étoile du matin] est [**l'étoile du soir**].
réf-3 *réf-4* *réf-3* *réf-4*

(*Manuel DEMOCRAT*, 2018 : 18)

Lors de l'annotation des textes, nous avons adopté la seconde position afin d'obtenir les attributs nominaux dans les deux langues et d'examiner de quelle manière et à quel degré ces expressions réfèrent en français et en chinois.

Enfin, nous distinguons les attributs coréférentiels et non coréférentiels au sujet. Dans les exemples [54] – [57], « c' » et « une des douleurs de sa vie » sont coréférentiels ; « Hassler » et « un génie » ne sont pas coréférentiels puisque le sujet est un individu spécifique défini, tandis que « un génie » réfère à une classe générique de personnes.

[54][C']était [**une des douleurs de sa vie**].
sujet *attribut-3*

[55][这] 是 [他 **终身** **隐** **痛** 之一] 。
zhe *shi* *ta* *zhongshen* *yin* *tong* *zhiyi*
ce être son toute-vie discret douleur un-des
sujet *attribut-4*

[56]Il s'écriait que [Hassler] était [**un génie**], comme on n'en voit qu'un par siècle.
sujet *attribut-1*

[57]他 大声的说 [哈斯莱] 是 [个 **天才**] , 一百 年 只会 出 一 个 的 那 种 天才 。
ta *dashengdeshuo* *Hasilai* *shi* *ge* *tiancai* *yibai* *nian* *zhihui* *chu* *yi* *ge* *de* *na* *zhong* *tiancai*
3P s'écrier Hassler être CL génie cent an seulement paraître un CL DE ce-là type génie
sujet *attribut-2*

- « Autres »

Nous avons choisi d'annoter les cinq fonctions syntaxiques qui sont le plus souvent mentionnées dans les théories de référence ; le reste des expressions référentielles seront annotées par « autres », comme les appellatifs et les éléments disloqués.

[58]Oh ! [**mon pauvre petit**], que tu es laid !
autres-1

[59]哦， [我的 小 乖乖]， 你 多 难看！
o *wode* *xiao* *guaiguai* *ni* *duo* *nankan*
ONOMA mon petit chéri 2P que laid
autres-2

[60][**Mon pauvre garçon**], tu n'es pas bien habile.
autres-3

[61][可怜的 孩子]， 你 真不 高明。
keliande *haizi* *ni* *zhenbu* *gaoming*
pauvre garçon 2P pas-du-tout habile
autres-4

5.2.2.5. Forme : déterminant, tête lexicale et modifieur

La forme de l'expression est beaucoup discutée dans les théories fonctionnelles et cognitives puisqu'elle est un indice qui peut permettre de prédire la continuité topicale et la perception du référent dans la représentation mentale du locuteur. Dans l'échelle de continuité topicale de Givón (1983), comme le montre la figure (5.5), la forme morphologique et la place syntaxique de l'expression impliquent la continuité de référent dans le discours : l'anaphore zéro représente la continuité maximale du référent, à savoir que ce sont les référents les plus accessibles, alors que les SN indéfinis servent à introduire ou à redémarrer un référent qui n'est plus continu dans le discours, ce qui correspond aux référents les moins accessibles pour les lecteurs / locuteurs.

Figure (5.5) : Échelle de continuité topicale de Givón (1983 : 17)

most continuous/accessible topic
↑ zero anaphora
unstressed/bound pronouns or grammatical agreement
stressed/independent pronouns
R-dislocated DEF-NP's
neutral-ordered DEF-NP's
L-dislocated DEF-NP's
Y-moved NP's ('contrastive topicalization')
cleft/focus constructions
↓ referential indefinite NP's
most discontinuous/inaccessible topic

Certaines théories cognitives cherchent à mettre en corrélation la forme de l'expression et l'importance cognitive du référent. Ces théories nous semblent intéressantes et importantes puisqu'elles sont au fondement d'observations sur des exemples en plusieurs langues, dont le japonais, l'amharique narratif écrit, l'ute, l'hébreux narratif biblique, l'espagnol latino-américain oral, l'anglais narratif, l'anglais oral, l'haoussa narratif et le chamorro. Il nous semble que cette échelle peut prédire les régularités en français et en chinois.

Les plus connues et les plus appropriées selon la perspective multilingue sont l'*Échelle d'accessibilité* d'Ariel (1990, 2013) et la *Hiérarchie du donné* de Gundel *et al*. (1993) qui s'adaptent à plusieurs langues, comme l'anglais, le français et l'hébreu. Il nous paraît pertinent d'appliquer ces corrélations formelles-cognitives aux expressions référentielles chinoises.

La notion d'*accessibilité* sert à décrire le degré de perception d'une entité du monde dans l'esprit du locuteur. La théorie d'Ariel (1990, 2013) cherche à répondre à deux questions concernant la relation entre l'état mental du locuteur et la forme des expressions linguistiques : i) comment définir l'*accessibilité mentale* ? ii) quelles expressions correspondent à quels degrés d'accessibilité mentale ? L'échelle d'accessibilité de diverses formes d'expressions référentielles est exposée dans la figure (5.6). Les prénoms et les noms avec et sans modifieur sont les moins accessibles aux locuteurs ; les zéros, les pronoms réfléchis et les marqueurs d'accords signifient que leur référent leur est plus accessible.

Figure (5.6) : Échelle d'accessibilité d'Ariel (1990 : 73) avec la traduction en français de Gardelle *et al*. (2019 : 7)

Low Accessibility	↑	**Faible degré d'Accessibilité**
Full name + modifier		Prénom et nom + modifieur
Full ('namy') name		Prénom et nom
Long definite description		Description définie longue
Short definite description		Description définie courte
Last name		Nom de famille
First name		Prénom
Distal demonstrative + modifier		Démonstratif distal + (SN) + modifieur
Proximal demonstrative + modifier		Démonstratif proximal + (SN) + modifieur
Distal demonstrative (+NP)		Démonstratif distal + (SN)
Proximal demonstrative (+NP)		Démonstratif proximal + (SN)
Stressed pronoun + gesture		Pronoms accentués + geste
Stressed pronoun		Pronoms accentués
Unstressed pronoun		Pronoums inaccentués
Cliticized pronoun		Pronoms clitiques
Extremely High Accessibility Markers (gaps, including pronoun, zero pronoun and wh-traces, reflexives, and Agreement)		Zéro, réfléchis, marqueurs d'accord
High Accessibility	↓	**Haut degré d'Accessibilité**

Selon l'approche multilingue, la théorie de l'accessibilité référentielle d'Ariel s'est révélée pertinente pour l'anglais et l'hébreu (Ariel, 1990, 2013), dont les échelles d'accessibilité sont similaires. Comme le montre le tableau (5.3) suivant, le groupe de *pronom accentué avec geste* n'est pas mentionné pour l'hébreu ; le groupe de l'accessibilité maximale en anglais contient plusieurs formes d'*éléments vides*, comme *pronom*, *pronom zéro*, *traces QU-*, *réfléchis* et *accord*, alors que seuls les *marqueurs des accords verbaux et personnels* et le *pronom zéro* sont mentionnés pour l'hébreu, et le pronom zéro présente un degré d'accessibilité plus haut que celui des marqueurs des accords verbaux et personnels.

Tableau (5.3) : Échelle d'accessibilité des marqueurs pour l'hébreu et pour l'anglais[9]

	En hébreu	En anglais
Accessibilité minimale	Nom propre complet + modifieur	Nom propre complet + modifieur
	Nom propre complet	Nom propre complet
	Description définie longue	Description définie longue
	Description définie brève	Description définie brève
	Nom patronymique	Nom patronymique
	Prénom	Prénom
	Démonstratif (d'éloignement →de proximité) + modifieur	Démonstratif d'éloignement + modifieur
		Démonstratif de proximité + modifieur
	Démonstratif (d'éloignement →de proximité) (+ SN)	Démonstratif d'éloignement (+ SN)
		Démonstratif de proximité (+ SN)
	-	**Pronom accentué avec geste**
	Pronom accentué	Pronom accentué
	Pronom atone	Pronom atone
	Pronom clitique	Pronom clitique
Accessibilité maximale	**Marqueurs des accords verbaux et personnels**	**Marqueurs d'accessibilité très élevée (éléments vides, comprenant pro, PRO et traces QU-, réfléchis et accord)**
	Pronom zéro	

La notion du *donné* (en anglais *giveness*) est proposée par Prince (1981). Elle est définie selon trois aspects, dont l'objectif est de transférer les nouvelles et anciennes informations entre les locuteurs et les interlocuteurs. Premièrement, le *donné* est défini en termes de *prédictibilité* (en anglais *predictability / recoverability*) :

> *The speaker assumes that the hearer CAN PREDICT OR COULD HAVE PREDICTED that a PARTICULAR LINGUISTIC ITEM will or would occur in a particular position WITHIN A SENTENCE. (Prince, 1981 : 226)*

Deuxièmement, le *donné* est défini en termes de *saillance* (en anglais *saliency*) :

> *The speaker assumes that the hearer has or could appropriately have somme particular thing / entity / … in his / her CONSCIOUSNESS at the time on hearing the utterance. (Prince, 1981 : 228)*

Troisièmement, le *donné* est défini en termes de *connaissances partagées* (en anglais *shared knowledge*) :

> *The seaker assumes that the hearer « knows », assumes, or can infer a particular thing (but is not necessarily thinking about it). (Prince, 1981 : 230)*

À partir de la notion de *donné*, Gundel *et al*. (1990, 1993) proposent la *Théorie de la Hiérarchie du Donné* (en anglais *Giveness Hierarchy*) qu'ils placent au fondement d'une étude empirique dans l'objectif de constituer une corrélation entre le degré de familiarité du référent pour les locuteurs et les formes d'expression de la langue. L'objectif de cette hiérarchie est de décrire le degré de familiarité du référent mentionné dans le discours. Elle consiste à mettre en corrélation le degré du donné et la forme d'expression d'un référent, comme le montre la figure (5.7).

> *Referential givenness–newness involves a relation between a linguistic expression and a corresponding non-linguistic entity in the speaker / hearer's mind, the discourse (model), or some real or possible world, depending on where the*

[9] Nous traduisons en français.

referents or corresponding meanings of these linguistic expressions are assumed to reside. (Gundel & Fretheim, 2004 : 176)

Dans le cadre de cette théorie, les expressions sont divisées en six groupes selon la catégorie morphologique, qui correspondent à six statuts cognitifs de locuteurs avec les définitions suivantes :

> *TYPE IDENTIFIABLE: the addressee is able to access a representation of the type of object described by the expression.*
>
> *REFERENTIAL: The speaker intends to refer to a particular object or objects.*
>
> *UNIQUELY IDENTIFIABLE: the addressee can identify the speaker's intended referent on the basis of the nominal alone.*
>
> *FAMILIAR: the addressee is able to uniquely identify the intended referent because he already has a representation of it in memory (in long-term memory if it has not been recently mentioned or perceived, or in short-term memory if it has).*
>
> *ACTIVATED: the referent is represented in current short-term memory.*
>
> *IN FOCUS: the referent is not only in short-term memory, but is also at the current center of attention.*
>
> *(Gundel* et al.*, 1993 : 276 – 279)*

Cette hiérarchie nous intéresse en particulier parce qu'elle est établie pour les formes d'expression en anglais et est ensuite élargie à quatre autres langues (le japonais, le chinois mandarin, le russe et l'espagnol). Les auteurs cherchent à prouver que ce modèle qui relie le statut cognitif à la forme de l'expression linguistique représente une certaine *universalité* qui concerne toutes les langues.

Figure (5.7) : Hiérarchie du donné référentiel de Gundel *et al.* (1993 : 275) avec la traduction en français de Gardelle *et al.* (2019 : 8)

au premier plan >	activé >	connu >	identifiable >	référentiel >	identifiable
('*in focus*')		('*familiar*')	comme unique		comme type
it	*this/that* pron *this* + N	*that* + N	*the* + N	*this* indéfini + N	*a* + N

D'après les deux échelles, les expressions référentielles sont annotées par la forme en trois temps :

i) selon le déterminant : seules les expressions nominales sont concernées ;

ii) selon les formes non nominales ;

iii) selon le modifieur et la tête lexicale : seules les expressions nominales sont concernées.

5.2.2.5.1. Mentions nominales : déterminants et noms propres

Cette annotation ne concerne que les mentions nominales. Les noms propres sont annotés dans un groupe à part comme « noms propres ». Les autres nominaux sont étiquetés selon leur type de déterminant en cinq groupes : « SN défini », « SN possessif », « SN démonstratif », « SN indéfini » et « SN sans déterminant ».

- « Nom propre »

Pour le français, d'après les descriptions morphologiques des noms propres dans la *Grammaire Méthodique du français* (Riegel *et al.*, 2016 [1994]) et les études de Leroy (2004) et Gary-Prieur & Noailly (2019), nous annotons tous les noms avec une majuscule à l'initiale comme « nom propre », peu importe s'ils ont un déterminant (*le Rhin*) ou non (*Paris*), et peu importe s'ils sont simples (*Napoléon*) ou composites (*les Plaisirs du Jeune Âge*).

Pour le chinois, les noms propres n'ont pas de spécificité morphologique qui les distingue des noms communs. C'est pour cette raison que certains auteurs, comme Guo (郭锐, 2002), considèrent que les noms constituent une seule catégorie, sans distinction entre propres et communs. Vu que les critères morpho-syntaxiques ne permettent pas de distinguer ces deux sous-types de noms, certains autres auteurs proposent d'appliquer des critères sémantiques. Cao (曹长远, 1988) relie la définition des noms propres chinois par l'unicité et la dénomination :

> *专有名词指的是在特定的时间与空间中存在的独一无二的事物的名称。*
> *(Cao, 曹长远, 1988 : 101)*
>
> *Le nom propre est une dénomination d'un objet unique dans un contexte spatio-temporel spécifique. (nous traduisons)*

Dans la typologie de Cao, les noms propres sont subdivisés tout d'abord en ceux qui incluent des noms communs et ceux qui n'en incluent pas ; ces deux groupes contiennent chacun des sous-groupes sémantiques. Selon ces critères morphologiques et sémantiques, tous les noms propres chinois ne correspondent pas aux noms propres français. Les noms propres chinois incluant des noms communs sont divisés ainsi :

- les noms de rivières, de fleuves, de mers, d'océans, de montagnes, d'îles, d'archipels, etc., comme 'Long-Fleuve' (« Yangtze ») *Chang-jiang* 长江 et 'Pacifique-Océan' (« Océan Pacifique ») *Taiping-yang* 太平洋 ;
- les noms d'années, d'époques, d'événements, etc., comme '1949-Année' (« l'année 1949 ») *yijiusijiu-nian* 1949 年, 'Tang-Dynastie' (« Dynastie Tang ») *Tang-chao* 唐朝 et 'Opium-Guerre' (« Guerre de l'Opium ») *Yapian-zhanzheng* 鸦片战争 ;
- les noms de pays, de régions, etc., comme 'Amérique-Pays' (« les États-Unis ») *Mei-guo* 美国 et 'Jiangsu-Province' (« Province du Jiangsu ») *Jiangsu-sheng* 江苏省 ;
- les noms de groupes, d'organisations, etc., comme 'Uni-Pays' (« Nations-Unies ») *Lianhe-guo* 联合国, 'Chinois-Communiste-Parti' (« Parti Communiste Chinois ») *Zhongguo-Gongchan-dang* 中国共产党 ;

- les noms de villes, de bâtiments, etc., comme 'Tian'anmen-Place' (« Place Tian'anmen ») *Tiananmen-guangchang* 天安门广场 et 'Eiffel-Tour' (« Tour Eiffel ») *Aifeier-tieta* 埃菲尔铁塔 ;
- les noms de journaux, de magazines, de publications, etc., comme 'Peuple-Journal' (« Journal du Peuple ») *Renmin-ribao* 人民日报 et 'Figaro-Journal' (« le Figaro ») *Feijialuo-bao* 费加罗报 ;
- les noms de planètes, de constellations, etc., comme 'Mars-Planète' (« Mars ») *Jin-xing* 金星, 'Andromède-Constellation' (« Andromède ») *Xiannü-zuo* 仙女座 ;
- les noms d'écoles, d'entreprises, etc., comme 'Pékin-Université' (« Université de Pékin ») *Beijing-daxue* 北京大学 et 'Carrefour-Supermarché' (« Carrefour ») *Jialefu-chaoshi* 家乐福超市 ;
- les autres noms propres composés.

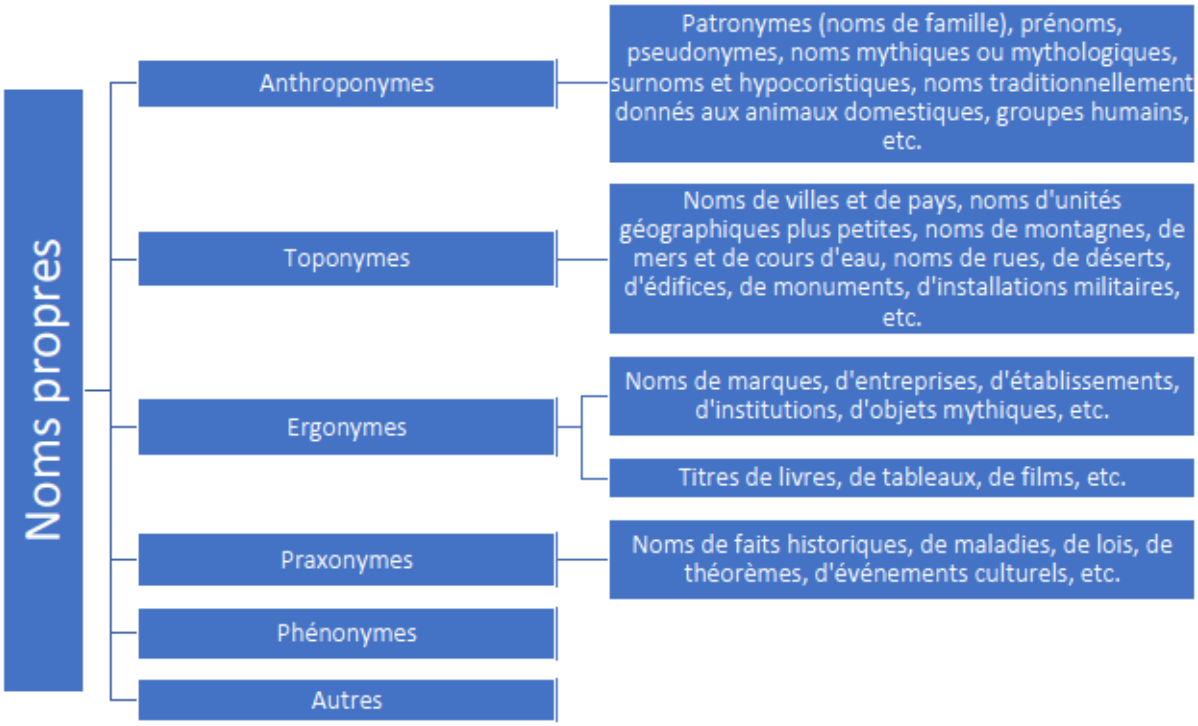
Les noms propres chinois sans nom commun sont :

- les noms et prénoms de personnes, comme 'Aristote' *Yalishiduode* 亚里士多德, 'Confucius' *Kongzi* 孔子 ;
- les noms d'autres objets particuliers, comme 'Soleil' *taiyang* 太阳, 'Lune' *yueliang* 月亮.

Nous utiliserons ces critères sémantiques pour l'annotation des noms propres chinois.

Il existe des critères sémantiques similaires relatifs aux noms propres pour le français, comme celui de Leroy (2004). Pour éviter d'utiliser les critères morphologiques, l'auteur propose un classement des noms propres français selon un principe sémantique : six catégories sémantiques sont signalées pour les noms propres : i) les anthroponymes (noms de personne) ; ii) les toponymes (noms de lieu) ; iii) les ergonymes (noms de réalisation ou de découverte humaine ayant une réalité matérielle) ; iv) les praxonymes (noms de réalisation ou de découverte humaine non matérielle) ; v) les phénonymes (noms de phénomène naturel) et vi) autres noms propres comme les noms de bateau, de bar, d'hôtel, de fusée, de programme, etc. Chaque catégorie comporte encore quelques sous-catégories sémantiques, comme le montre la figure (5.8).

Figure (5.8) : Classement sémantique des noms propres de Leroy (2004)



Selon ce critère sémantique, nous pouvons distinguer les noms propres des noms communs en chinois sans prendre en compte des critères morphologiques.

Nous appliquons en même temps les critères morphologiques et sémantiques pour le français. Comme nous le montrons dans les exemples parallèles [62] et [63], les noms propres des deux langues sont annotés.

[62]Les cloches de [**Saint-Martin**] chantèrent dans la nuit.
npr-1

[63]黑夜 里 传来 [圣·马丁 寺] 的 钟 声。
heiye li chuanlai Shengmading si de zhong sheng
nuit pendant chanter Saint-Martin église DE cloche voix
npr-2

- « SN défini »

Les SN définis ne concernent que les expressions référentielles en français initiées par les quatre articles définis « le », « la », « les » et « l' ». Il n'existe pas de parallélisme d'article défini en chinois. Ainsi, l'étiquette « SN défini » n'existe pas dans l'annotation des textes en chinois. Comme le montrent les extraits parallèles [64] et [65] où tous les SN définis en français sont traduits par des SN sans déterminant en chinois : « la digestion » est traduit par 'digérer nourriture DE travail' *xiaohua-shiwu-de-gongzuo* 消化食物的工作 ; « la figure » par 'figure' *lianpang* 脸庞, etc.

[64][**La digestion**] le plonge en une extase, où [**les figures**], [**les ombres**], [**l'abat-jour de** [**la lampe**]],
SN-déf *SN-déf* *SN-déf* *SN-déf* *SN-déf*
[**les langues de flammes qui dansent avec une pluie d'étoiles dans** [**la cheminée noire**]], tout
SN-déf *SN-déf*
prend une apparence réjouissante et magique.

[65][消化 食物 的 工作] 使 他 出了神：[脸庞]，[影子]，[灯罩]，在 [[黑魆魆的
xiaohua shiwu de gongzuo shi ta chuleshen lianpang yingzi dengzhao zai heixuxude
digérer nourriture DE travail rendre 3P extase-LE visage ombre lampe-abat-jour PREP noirâtre
SN-sans-dét *SN-sans-dét* *SN-sans-dét* *SN-sans-dét* *SN-sans-dét*

壁炉] 中 闪烁 飞舞 的 火舌]，一切都 有 一 副 可喜的 神奇的 面貌。
bilu zhong shanshuo feiwu de huoshe yiqiedou you yi fu kexide shenqide mianmao
cheminée dedans étinceler danser DE flamme tout avoir un CL réjouissant magique apparence
SN-sans-dét

Notre hypothèse porte sur le parallélisme entre les SN définis français et les SN sans déterminant chinois dans l'échelle d'accessibilité et dans l'échelle du donné : les deux catégories manifesteraient le même degré d'accessibilité et de donné. Cette hypothèse sera illustrée en même temps que les autres dans les figures à la fin de cette partie.

- « SN possessif »

Les SN possessifs sont marqués par les déterminants possessifs de chaque langue, comme « mon / ma / mes » en français pour la première personne du singulier, et 'mon / ma / mes' *wode* 我的 en chinois. À la différence des déterminants possessifs du chinois marqués par le genre et le nombre de personnes, ceux du français sont marqués, par ailleurs, par le nombre d'objets possédés. En raison de ces accords plus riches, les déterminants possessifs français peuvent refléter plus de caractéristiques référentielles que ceux du chinois. Toutes les variations des déterminants possessifs des deux langues figurent dans le tableau (5.4).

Tableau (5.4) : Déterminants possessifs en français et en chinois

Personne	Nbre et genre	Français	Chinois
1P	Singulier masculin	Mon	*wo(de)* 我 (的) 'de moi'
	Singulier féminin	Ma	
	Pluriel	Mes	
2P	Singulier masculin	Ton	*ni(de)* 你 (的) 'de toi'
	Singulier féminin	Ta	*nin(de)* 您 (的) 'de vous' (en politesse)
	Pluriel	Tes	
3P	Singulier masculin	Son	*ta(de)* 他 (的) 'de lui (masculin)'
	Singulier féminin	Sa	*ta(de)* 她 (的) 'd'elle'
	Pluriel	Ses	*ta(de)* 它 (的) 'de lui (inanimé)'
1Ps	Singulier	Notre	*women(de)* 我们[10] (的)
	Pluriel	Nos	
2Ps	Singulier	Votre	*nimen(de)* 你们 (的)
	Pluriel	Vos	
3Ps	Singulier	Leur	*tamen(de)* 他们 (的) 'd'eux (masculin)'
	Pluriel	Leurs	*tamen(de)* 她们 (的) 'd'elles'
			tamen(de) 它们 (的) 'd'eux (inanimé)'

[10] Une exception de modifieur postposé en chinois concerne le suffixe quantitatif *men* 们, équivalent au suffixe *-s* à la fin des noms en français, sauf que *men* 们 ne peut s'employer qu'avec quatre types de dénominations d'humain pour exprimer une quantité indéfinie : les pronoms personnels (*wo* 我 – *women* 我们 'je – nous'), les noms génériques d'humain (*ren* 人 – *renmen* 人们 'personne – gens'), les catégories d'humain (*nüshi* 女士 – *nüshimen* 女士们, 'madame – mesdames') et les noms propres d'humain (*Leifeng* 雷锋 – *Leifengmen* 雷锋们 'Lei Feng – les gens du type Lei Feng'). *Cf.* aussi Lv (吕叔湘, 1990) et Jiang (江蓝生, 2018).

Tout comme les déterminants définis, les possessifs sont un autre type de déterminants définis dans les deux langues. Ils sont annotés comme l'indiquent les extraits suivants :

[66][**Sa douleur**] s'évanouit, [**son cœur**] se mit à rire : et il glissa dans le rêve, avec un soupir d'abandon.
SN-poss-1 *SN-poss-2*

[67][他的 **痛苦**] 消散 了，心 也 笑开 了；他 轻松的 叹 了口 气，溜进 了
tade tongku xiaosan le xin ye xiaokai le ta qingsongde tan le kou qi liujin le
son douleur s'évanouir LE cœur aussi rire LE 3P légèrement soupir LE CL air glisser LE
SN-poss-3

梦乡。
mengxiang
rêverie

- « SN démonstratif »

Les SN démonstratifs sont initiés par les déterminants démonstratifs de proximité et d'éloignement dans chaque langue, c'est-à-dire « ce (-là) », « cet (-là) », « cette (-là) » et « ces (-là) » en français, et 'ce / cet / cette' *zhe* 这, 'ces' *zhexie* 这些, 'ce / cet / cette -là' *na* 那 et 'ces N-là' *naxie* 那些 en chinois. Les expressions démonstratives sont annotées comme mentions dans notre corpus, comme le montrent les exemples suivants :

[68]Louisa, passive, résignée, a accepté [**ce malheur**], comme elle accepte tout.
SN-dém-1

[69]懦弱隐忍的 **鲁意莎**，对 什么 都 是 逆来顺受的，就 一声不响的 接受 了 [**这**
nuoruoyinrende Luyisha dui shenme dou shi nilaishunshoude jiu yishengbuxiangde jieshou le zhe
résigné Louisa à quoi tout être passif donc silencieux accepter LE ce
SN-dém-2

桩 不幸]。
zhuang buxing
CL malheur

- « SN indéfini »

Les SN indéfinis sont introduits par les déterminants indéfinis de chaque langue. Prenons la définition des « indéfinis » en français de Corblin (1997) :

> *En français, on appelle alors indéfinis des groupes nominaux à tête nominale qui ne sont ni définis ni démonstratifs : un N, beaucoup de N, certains N, au moins deux N, deux N exactement, tout N, chaque N, deux N sur trois, trente pour cent des N, … (Corblin, 1997 : 8)*

Suivant cette définition large, tous les déterminants autres que les définis et les démonstratifs peuvent être classés dans le groupe des indéfinis. Ainsi, la liste des indéfinis est plus longue que celle des définis et des démonstratifs. D'après la typologie de la *Grammaire Méthodique du Français* (Riegel *et al.*, 2016 : 292), cette longue liste des déterminants indéfinis peut se diviser selon les groupes suivants :

- les « articles indéfinis », comme « un(e) », « des », « de » et « d' » ;
- les « articles partitifs », comme « du », « de la » et « de l' » ;
- les « déterminants indéfinis de la quantité déterminée », comme « un », « deux » et « trois » ;
- les « déterminants indéfinis de la quantité non déterminée », comme « quelques », « plusieurs », « divers » et « différents » ;

- les « déterminants indéfinis de la totalité distributive exhaustive », comme « chaque », « tout(e) » et « n'importe quel(le) ».

Les sous-types de déterminants indéfinis chinois équivalent à ceux du français, à l'exception des articles partitifs ; nous établissons ces équivalences dans la typologie des déterminants indéfinis du français :

- les articles indéfinis : 'un / une' *yige* 一个 et 'des' *yixie* 一些 ;
- les déterminants indéfinis de la quantité déterminée : 'un' *yige* 一个, 'deux' *liangge* 两个 et 'trois' *sange* 三个 ;
- les déterminants indéfinis de la quantité non déterminée : 'quelques' *jige* 几个, 'plusieurs' *henduoge* 很多个, 'divers' *duozhong* 多种 et 'différents' *butongzhong* 不同种 ;
- les déterminants indéfinis de la totalité distributive exhaustive : 'chaque' *meige* 每个, 'tout' *suoyoude* 所有的 et 'n'importe quel(le)' *wulunnayige* 无论哪一个.

Quelques exemples de SN indéfinis sont donnés dans les extraits [70] – [73].

[70][**Un valet**] […] lui demanda d'[**un ton protecteur**] ce qu'il venait faire.
SN-ind-1 *SN-ind-2*

[71][一 个 **当差**] […] 用 长辈的 口气 问 他 来意。
Yi ge dangchai yong zhangbeide kouqi wen ta laiyi
un CL valet utiliser protecteur ton demander 3P objectif
SN-ind-3

[72]Il cherchait [**un grillon**] pour en faire [**un cheval**] : il lui mettait doucement sa baguette sur le dos,
SN-ind-4 *SN-ind-5*

et disait [**une formule**]. L'insecte se sauvait : il lui barrait le chemin. Après [**quelques instants**],
SN-ind-6 *SN-ind-7*

il était couché à plat ventre, près de lui, et il le regardait. Il avait oublié son rôle de magicien, et s'amusait à retourner sur le dos la pauvre bête, en riant de ses contorsions.

[73]他 找 了 [一 只 **蟋蟀**] 想 叫 它 变成 [一 **骑** 马] : 他 把 棍子 轻轻的
Ta zhao le yi zhi xishuai xiang jiao ta biancheng yi ji ma ta ba gunzi qingqingde
3P chercher LE un CL grillon vouloir demander 3P devenir un CL cheval 3P BA baguette doucement
SN-ind-8 *SN-ind-9*

放 在 它的 背 上， 嘴 里 念着 咒语。蟋蟀 逃 了…… 他 挡住 它的
fang zai tade bei shang zui li nianzhe zhouyu xishuai tao le ta dangzhu tade
mettre PREP son dos dessus bouche dedans disant formule grillon s'enfuir LE 3P barrer son

去路。过了 [一会]， 他 躺 在 地 下， 靠近着 虫， 对 他 望着。他 忘
qulu guole yihui ta tang zai di xia kaojinzhe chong dui ta wangzhe ta wang
chemin après un-temps 3P se-coucher PREP terre dessus s'approchant insecte à 3P regardant 3P oublier
SN-ind-10

了 魔术师 的 角色， 只 把 可怜的 虫 仰天翻着， 看 它 扭来扭去 的 扯动
le moshushi de juese zhi ba keliande chong yangtianfanzhe kan ta niulainiuqu de chedong
LE magicien DE rôle seulement BA pauvre insecte retournant regarder 3P se-tourner DE bouger

身子， 笑了出来。
shenzi xiaolechulai
corps rire-LE

- « SN sans déterminant »

Les SN sans déterminant sont composés d'un nom, simple ou composé, avec ou sans modifieur, comme « pluie » dans l'extrait [74], 'nuit' *heiye* 黑夜 et 'pluie' *yutian* 雨天 dans l'extrait [75].

[74]Dans l'air mouillé de [**pluie**], elle cheminait comme un pas sur la mousse.
SN-sans-dét-1

[75][严肃 迟缓的 音调]，在 [[雨天] 潮润的 空气] 中] 进行，有如 踏 在
yansu chihuande yindiao zai yutian chaorunde kongqi zhong jinxing youru ta zai
grave lent voix PREP pluie mouillé air intérieur procéder comme cheminer PREP
SN-sans-dét-2 *SN-sans-dét-3* *SN-sans-dét-4* *SN-sans-dét-5*

[苔藓] 上] 的 脚步]。
taixian shang de jiaobu
mousse dessus DE pas
SN-sans-dét-6 *SN-sans-dét-7* *SN-sans-dét-8*

- « Autres »

L'ensemble des autres SN qui contiennent un déterminant autre sont annotés par l'étiquette « autres ».

5.2.2.5.2. Mentions non nominales

Les expressions référentielles non nominales sont classées de manière plus floue dans les échelles cognitives par rapport au classement des expressions nominales. Dans l'échelle d'accessibilité d'Ariel, elles sont subdivisées en *pronom accentué*, *pronom inaccentué*, *pronom clitique*, et les autres sont placées dans le même groupe comme *zéro, réfléchis et marqueurs d'accord*. Il nous semble que ce regroupement se constitue à partir d'observations basées sur le français et l'anglais. Il n'est pas pertinent pour le chinois puisque les pronoms de cette langue ne se distinguent pas de ces formes morpho-syntaxiques[11]. Dans la hiérarchie du donné de Gundel *et al.*, seul le pronom neutre de la troisième personne du singulier *it* est mentionné comme *au premier plan* ; les autres formes ne sont pas prises en compte. Ce groupe *it* ne nous convient pas non plus pour observer les fonctions des mentions non nominales dans les chaînes de référence puisque les autres formes non nominales sont aussi diverses.

À partir d'une observation primitive des formes non nominales des maillons dans le corpus *Jean-Christophe*, en prenant en compte les formes mentionnées dans les échelles cognitives, nous fixons d'abord trois étiquettes pour annoter les mentions non nominales selon le critère grammatical : *pronom personnel*, *déterminant possessif* et *zéro*. La quatrième étiquette est ensuite rajoutée selon le critère référentiel : *déictique*, que nous présenterons dans les sections ultérieures. La cinquième étiquette est dites *autres* et regroupe toutes les autres formes. Toutes les formes ne sont pas censées être utilisées aussi fréquemment les unes que les autres comme mentions et comme maillons des chaînes de référence. Par exemple, les pronoms personnels et les zéros peuvent être très récurrents selon les configurations des chaînes déjà établies par les linguistes que nous avons présentées dans le *Chapitre 1*, tandis que les pronoms démonstratifs et les pronoms possessifs peuvent l'être

[11] Les opinions sur la fonction référentielles des accords sont diverses. Du point de vue sémantique et cognitif, les accords grammaticaux sont des marqueurs anaphoriques et référentiels de haut degré d'accessibilité (Givón, 1983 ; Reichler-Béguelin, 1988 ; Ariel, 1990). Du point de vue morphologique traditionnel, les expressions anaphoriques, surtout les maillons dans les chaînes de référence, peuvent se présenter en catégories nominale et pronominale (*cf.* Chastain, 1975 ; Corblin, 1985, 1995 ; Charolles, 2002) ce qui exclut les accords. Cela nous paraît problématique pour annoter notre corpus franco-chinois. Sachant que le chinois est une langue analytique dépourvue des accords et de la conjugaison, nous choisissons de ne pas annoter les accords. Autrement dit, les accords inhérents des expressions nominales et pronominales en français sont annotés avec les expressions, mais les accords d'autres expressions, comme ceux des verbes et des participes, ne sont pas annotés.

moins et encore moins comme maillons des chaînes dans notre corpus constitué de roman-fleuve.

- « Pronom personnel »

Les pronoms personnels incluent les pronoms des trois personnes. Ces derniers sont annotés par deux étiquettes en raison de leur fonction référentielle distincte :

> *Les pronoms personnels des 1ère et 2ème personnes (je, tu, nous, vous) désignent les participants à la situation d'énonciation, et ont essentiellement une interprétation déictique. Les pronoms de 3ème personne (il, elle, ils, elles) désignent les non-participants et sont essentiellement anaphoriques. (Abeillé & Godard, 2021 : 983)*

Sous l'étiquette *pronom personnel*, nous annotons seulement les pronoms personnels de la troisième personne des trois genres au singulier et au pluriel dans les deux langues, à savoir « il(s) » *ta(men)* 他 (们), « elle(s) » *ta(men)* 她 (们), « il(s) » *ta(men)* 它 (们), et « on » *ren(men)* 人 (们)[12], avec leur variation en formes conjointes et disjointes, comme le montre le tableau (5.5) en bas. Les pronoms personnels de la première et de la deuxième personnes sont annotés à part comme *déictiques* selon le critère référentiel que nous présenterons dans la section suivante.

Tableau (5.5) : Étiquettes d'annotation des pronoms personnels en chinois et en français

Pronom personnel chinois	Pronom personnel français	Étiquette d'annotation
wo(ziji) 我 (自己)	Je, me, moi (-même)	« déictique »
ni(ziji) 你 (自己)	Tu, te, toi (-même)	« déictique »
ta(ziji) 他 (自己)	Il, le, se, lui (-même), y, en	« pronom personnel »
ta(ziji) 她 (自己)	Elle, la, se, elle-même, y, en	« pronom personnel »
ta(ziji) 它 (自己)	Il (inanimé), le, se, lui (-même), y, en	« pronom personnel »
nin(ziji) 您 (自己)	Vous (poli) (-même)	« déictique »
women(ziji) 我们 (自己)	Nous (-même)	« déictique »
nimen(ziji) 你们 (自己)	Vous (pluriel) (-mêmes)	« déictique »
tamen(ziji) 他们 (自己)	Ils, les, se, leur, eux (-mêmes), y, en	« pronom personnel »
tamen(ziji) 她们 (自己)	Elles, les, se, leur, elles-mêmes, y, en	« pronom personnel »
tamen(ziji) 它们 (自己)	Ils (inanimé)	« pronom personnel »
Ren(men)(ziji) 人 (们) (自己)	On, se, soi (-même)	« pronom personnel »

De manière générale, les pronoms personnels comportent en français différentes marques selon la personne, le nombre, le genre, la fonction syntaxique et le degré d'intégration dans la construction syntaxique (*i.e.* formes conjointes ou clitiques *versus* formes disjointes), comme « il », « lui », « elle », « ils » et « eux » (*cf.* Charolles, 2002 : 184). Les pronoms personnels en chinois ne varient pas en fonction du degré d'intégration, comme 'il /

[12] Selon le référent, « on » *ren(men)* 人 (们) peut être un pronom défini ou indéfini. Dans le second cas, il est annoté par l'étiquette « autres », que nous analyserons en détails dans la partie des résultats avec les exemples du corpus.

le / lui' *ta* 他 et 'ils / les / eux' *tamen* 他们. Illustrons, par l'annotation de quelques exemples, des pronoms personnels dans les extraits suivants :

[76][**Elle**] semblait écrasée par [**eux**].
pro-per-1 *pro-per-2*

[77][她] 似乎 被 [他们] 压倒 了。
ta *sihu* *bei* *tamen* *yadao* *le*
3P sembler par 3Ps écraser LE
pro-per-3 *pro-per-4*

[78][**Il**] touchait les fleurs, en [**leur**] enjoignant de se changer en carrosses dorés, comme [**on**]
pro-per-5 *pro-per-6* *pro-per-7*
[**lui**] avait dit qu'[**elles**] faisaient dans les contes.
pro-per-8 *pro-per-9*

[79][他] 指着 花 一 点， 吩咐 [它们] 变成 金色的 四轮车， 象
ta *zhizhe* *hua* *yi* *dian* *fenfu* *tamen* *biancheng* *jinsede* *siluncheche* *xiang*
3P touchant fleur un pointer enjoindre 3Ps devenir doré carrosse comme
pro-per-10 *pro-per-11*
童话 中 所说的 一样。
tonghua *zhong* *suoshuode* *yiyang*
conte dedans parlant pareil

□ « Déictique »

Cette étiquette référentielle n'est pas au même niveau grammatical que les autres étiquettes de catégorie. Ce groupe concerne les expressions référentielles déictiques non nominales et se divise en deux sous-groupes :

- les pronoms personnels de première et deuxième personnes, à savoir « je / nous », « tu / vous » en français et 'je / nous' *wo(men)* 我 (们), 'tu / vous' *ni(men)* 你 (们) en chinois, comme le montre le tableau (5.5) en haut[13] ;
- les adverbes spatio-temporels, comme « ici / là-bas » et « maintenant » en français, et 'ici / là-bas' *zheli* 这里 / *nali* 那里, 'maintenant' *xianzai* 现在 en chinois.

Les exemples des mentions de déictiques sont donnés dans les extraits parallèles [80] – [85].

[80][**Tu**] ne la manges pas ? Donne-la-[**moi**], dis, Christophe.
déictique-1 *déictique-2*

[81]"[你] 不 吃 吗？ 给 [我] 行不行， 克利斯朵夫？"
ni *bu* *chi* *ma* *gei* *wo* *xingbuxing* *Kelisiduofu*
2P NEG manger INTERR donner 1P INTERR Christophe
déictique-3 *déictique-4*

[82]Il commence par s'arrêter, […] pour savoir quelle histoire il se racontera [**aujourd'hui**].
déictique-5

[83]他 停下来， […] 盘算 [今天] 讲 哪 桩 故事。
ta *tingxialai* *pansuan* *jintian* *jiang* *na* *zhuang* *gushi*
3P s'arrêter réfléchir aujourd'hui raconter quel CL histoire
déictique-6

[84]C'est mon vieux veston bleu : il y a une tache [**là**].
déictique-7

[85]是 我的 旧 蓝 上装： [这儿] 还 有 块 污迹。
shi *wode* *jiu* *lan* *shangzhuang* *zheer* *hai* *you* *kuai* *wuji*
être mon vieux bleu veston là encore avoir CL tâche
déictique-8

[13] Grammaticalement, les pronoms déictiques font partie des pronoms personnels. Nous les annotons à part pour distinguer leur fonction référentielle de celle de pronom personnel de troisième personne.

 « Zéro »

Les pronoms sujets ne sont pas nécessairement exprimés dans certaines propositions ; ils sont alors appelés *zéro*.

En plus des sujet zéro, d'autres places syntaxiques peuvent également être vides, comme c'est le cas des compléments d'objet zéros. Pourtant, ils sont très peu récurrents en chinois selon l'étude quantitative de Hou & Sun (侯敏 & 孙建军, 2004), et sont fortement contraints par la transivité du verbe : dans leur corpus, les sujets zéro occupent 93 % des zéros. Ainsi, nous n'annotons que les sujet zéro dans notre corpus. Cette attitude est adoptée également par Ariel (1990), qui définit un sujet zéro dans son échelle d'accessibilité référentielle par un vide anaphorique à la place du sujet syntaxique.

En terme de traitement outillé, il peut être difficile d'annoter une expression sans forme explicite ou phonétique. Landragin (2011) propose deux manières d'annotation pour ce type de situation : soit ajouter un signe de support, comme « Ø », en représentant le zéro, soit annoter directement le mot suivant du zéro (ce mot est souvent un verbe). Cette seconde manière est choisie dans certains travaux comme ceux de Landragin (2011), de Schnedecker & Landragin (2014) et également dans le nôtre, où nous avons annoté les verbes en tant qu'ils représentent des zéros dans le corpus[14]. Les exemples parallèles des zéros dans les deux langues sont mis en gras dans les extraits [86] – [89].

[86]Un valet flânait sous le porche ; il arrêta l'enfant et [**lui**]*zéro-1* demanda d'un ton protecteur ce qu'il venait faire.

[87]一	个	当差	在	门洞	下面	闲荡，	[拦]住	了	孩子，	[用]	长辈的	口气	问	他	来意。
yi	*ge*	*dangchai*	*zai*	*mendong*	*xiamian*	*xiandang*	*lan zhu*	*le*	*haizi*	*yong*	*zhangbeide*	*kouqi*	*wen*	*ta*	*laiyi*
un	CL	valet	PREP	porche	dessous	flâner	arrêter *zéro-2*	LE	enfant	utiliser *zéro-3*	protecteur	ton	demander	3P	objectif

[88]但	因	他	天性	倔强，	[倒]反	更喜欢	那些	和声。	
dan	*yin*	*ta*	*tianxing*	*juejuang*	*dao fan*	*gengxihuan*	*naxie*	*hesheng*	
mais	car	son	nature	indiscipline	contrairement *zéro-4*	préférer	ces-là	accords	

'Comme il est naturellement indiscipliné, il ne les en aime que mieux.'

[89]我们	只	希望	他	一	件	事，	[就]	是	做	个	好人。	
women	*zhi*	*xiwang*	*ta*	*yi*	*jian*	*shi*	*jiu*	*shi*	*zuo*	*ge*	*haoren*	
1Ps	seulement	demander	3P	un	CL	chose	juste *zéro-5*	être	devenir	CL	brave-homme.	

'On ne lui demande qu'une chose, c'est de devenir un brave homme.'

 « Déterminant possessif »

Les déterminants possessifs sont le seul type de déterminants qui peuvent référer tout seuls, puisqu'ils peuvent renvoyer à un possesseur (« son » dans « son jardin », « mon » dans « mon piano », par exemple) et équivalent au complément du nom en structure « de quelqu'un / de quelque chose » (« ton » dans « ton père », « son » dans « son héros », par

[14] Contrairement à la façon dont nous avons choisi d'annoter le corpus, les phrases et les extraits que nous citons dans le corps du présent travail sont annotés par la première manière, à savoir que les places des zéros sont marquées par « Ø » pour qu'elles soient plus visibles par les lecteurs.

exemple). La forme morphologique varie en fonction de divers critères dans les deux langues. Ceux du français s'accordent avec la personne, le nombre et le genre du nom ; ceux du chinois avec le nombre et le genre du possesseur. Toutes les variations figurent dans le tableau (5.4) dans la partie « SN possessif » et sont annotées dans notre corpus comme le montrent les extraits [90] – [92][15].

[90]Louisa saisit l'enfant de [**ses**] mains fiévreuses et le cacha contre [**son**] sein.
dét-poss-1 *dét-poss-2*

[91]一 个 寒噤 沿着 [他的] 脊梁 直流 下去。
yi *ge* *hanjin* *yanzhe* *tade* *jiliang* *zhiliu* *xiaqu*
un CL frisson sur son échine courir dessous
dét-poss-3

'Un frisson lui court le long de l'échine.'

[92]鲁意莎 听着 钟声， 也 如梦如幻的 想着 [她] 过去的 苦难， 想着 睡 在
Luyisha *tingzhe* *zhongsheng* *ye* *rumengruhuande* *xiangzhe* *ta* *guoqude* *kunan* *xiangzhe* *shui* *zai*
Louisa écoutant cloche aussi rêvant pensant son passé misère pensant dormir PREP
dét-poss-4

身旁 的 亲爱的 婴儿 的 前程。
shenpang *de* *qinaide* *yinger* *de* *quancheng*
côté DE cher enfant DE futur

'Louisa rêvait aussi, en les écoutant, à [**ses**] misères passées et à ce que serait plus tard le cher petit enfant endormi auprès d'elle.'
dét-poss-5

- « Autres »

L'étiquette « autres » est annotée sur toutes les mentions des autres formes non nominales. Il s'agit de trois catégories principales : les pronoms indéfinis, les pronoms possessifs et les pronoms démonstratifs.

Les pronoms indéfinis comprennent un grand groupe de divers pronoms, comme « tout », « chacun », « l'autre », « un », « certains » et « tout » en français, et 'gens / quelqu'un' *(you)ren* (有) 人, 'certains' *youde* 有的 et 'qui / quelqu'un' *shui* 谁 en chinois.

En chinois, la frontière entre les pronoms indéfinis, interrogatifs et relatifs est parfois floue, sans aucune distinction formelle, ce qui leur confère une certaine polyvalence. Le Goffic & Wang (2002) mentionnent à ce sujet le pronom *shui* 谁, qui a trois usages : celui de pronom interrogatif (extrait [93]), équivalent à *qui* en français ; celui de pronom indéfini (extrait [94]), équivalent à *quelqu'un* ou *quelques-uns*, usage commun au français et au chinois ; celui enfin de pronom *intégratif* (connecteur subordonnant non relatif) (extrait [95]), qui existe spécifiquement en chinois sans équivalent en français. Par conséquent, la référence de ce groupe de pronoms indéfinis n'est pas aussi limpide que celle des définis. La catégorie grammaticale ne peut pas être considérée comme le seul critère de distinction entre les expressions référentielles et non référentielles.

[93]你 等 谁？
ni *deng* *shui*
2P attendre qui
'Qui attends-tu ?'

[15] La particule *de* 的 'DE' a diverses fonctions grammaticales (*cf.* Paris, 1977 ; Culioli, 1994). Dans nos gloses, les DE possessifs suivants les pronoms sont annotés comme parties des déterminants possessifs ; les autres sont annotés comme DE à part.

[94]你 等 **谁** 吗 ?
ni deng shui ma
2P attendre quelqu'un INTERR
'Tu attends quelqu'un ?'
[95]**谁** 笑 到 最后，**谁** 笑 得 最好。
shui xiao dao zuihou shui xiao de zuihao
qui rire PREP dernier qui rire AUX mieux
'Rira bien qui rira le dernier.'

(Le Goffic & Wang, 2002 : 453)

Les pronoms possessifs français (« le mien / la mienne / les miens / les miennes ») reflètent la personne, le genre et le nombre du référent ; ceux du chinois ('le mien / la mienne / les miens / les miennes' *wode* 我的) ne s'accordent qu'avec la personne et partagent les mêmes formes que celles des déterminants possessifs.

Les pronoms démonstratifs français (« ça », « ceci / cela », « celui-ci / celui-là », « celle-ci / celle-là », « ceux-ci / ceux-là» et « celles-ci / celles-là ») présentent également des formes plus variables qui reflètent le genre et le nombre du référent, alors que ceux du chinois n'ont que deux formes pour les référents, dont une désigne ceux de la proximité ('ceci / celui-ci' *zhe* 这) et l'autre ceux de l'éloignement ('cela / celui-là' *na* 那).

Les exemples parallèles de divers pronoms sont annotés dans les extraits [96] – [99].

[96][**Tout**] est à lui.
autres-1-pro-ind

[97][一切] 都 是 [他的]。
yiqie dou shi tade
tout tout être le-sien
autres-2-pro-ind *autres-3-pro-ppss*

[98]Il comptait, dans son absurde orgueil, succéder à son père, comme directeur de musique : [un autre]
autres-4-pro-ind
fut nommé.

[99]他 凭着 那 种 荒谬的 骄傲，满以为 能够 承继 父亲 作 乐队 指挥；
ta pingzhe na zhong huangmiude jiaoao manyiwei nenggou chengji fuqin zuo yuedui zhihui
3P comptant ce-là type absurde orgueil penser pouvoir succéder père comme orchestre chef

结果 是 任命 了 [别人]。
jieguo shi renming le bieren
enfin être nommer LE autrui
autres-5-pro-ind

Dans le traitement de corpus, certains linguistiques informaticiens signalent que les verbes ne sont pas annotés comme expressions référentielles de manière générale, sauf s'ils sont repris ou coréfèrent aux autres expressions (*cf.* Longo & Todirascu, 2014 ; Oberlé, 2016). Nous adoptons nous aussi cette manière de ne pas annoter les verbes comme expressions référentielles dans notre corpus.

5.2.2.5.3. Modifieur et tête lexicale des mentions nominales

Les expressions référentielles nominales peuvent se caractériser non seulement par leur déterminant, mais également par leur tête lexicale (nom commun et nom propre) ainsi que par leur *modifieur*[16]. Ariel (1990) distingue le « prénom et nom + modifieur » et le

[16] Les synonymes de *modifieur* sont divers et varient selon les auteurs. Par exemple, Dubois & Lagane (1973) les appellent des *noms simples* et *étendus*. Wilmet (1983) et Anscombre (1986) discutent des *extensions* des SN.

« prénom et nom » sans modifieur comme constituant deux groupes d'expressions référentielles avec deux degrés d'accessibilité distincts, ce qui implique que les modifieurs peuvent changer la fonction des mentions dans les chaînes de référence. En raison de cela, les mentions nominales sont annotées par les quatre étiquettes suivantes :

- « nom propre modifié »
- « nom commun sans modifieur »
- « nom propre modifié »
- « nom commun sans modifieur »

Un SN sans modifieur signifie que le SN n'est composé que d'un déterminant et d'une tête lexicale, sans aucun autre élément facultatif. Un SN modifié comporte non seulement un déterminant et une tête lexicale, mais également d'autres éléments optionnels et dépendants des noms[17] :

- les adjectifs, comme « l'enfant – le pauvre enfant », *haizi* 孩子 – *keliande-haizi* 可怜的孩子 ;
- les groupes prépositionnels, comme « l'enfant – l'enfant dans le jardin », *haizi* 孩子 – *gongyuan-li-de-haizi* 公园里的孩子 ;
- les propositions subordonnées relatives, marquées par les pronoms relatifs *qui*, *que*, *où*, *dont* en français, et par l'auxiliaire 的 en chinois, comme « l'enfant – l'enfant qui pleure », *haizi* 孩子 – *ku-de-haizi* 哭的孩子 ;
- les appositions, comme « l'enfant – l'enfant, violoniste de la cour », *haizi* 孩子 – *haizi-,-cour-violoniste* 孩子, 宫廷小提琴手.

Toutes ces formes sont annotées comme parties des expressions référentielles (*cf.* aussi Oberlé, 2016), quelle que soit leur position, antéposée ou postposée par rapport à la tête lexicale, leur longueur et leur nombre. Les exemples parallèles en français et en chinois des noms propres / communs modifiés et sans modifieur sont présentés dans les extraits [100] – [103].

[100] [**Un soir**] que [**Melchior**] dînait en [**ville**], [**Gottfried, resté seul dans** [**la salle** [**du bas**]]],
nco-sans-mod *npr-sans-mod* *nco-sans-mod* *npr-mod* *nco-mod* *nco-sans-mod*

tandis que [**Louisa**] couchait [**les deux petits**], sortit, et alla s'asseoir à [[**quelques pas de**
npr-sans-mod *nco-sans-mod* *nco-mod*

[**la maison**]], [**au bord** [**du fleuve**]].
nco-sans-mod *nco-mod* *nco-sans-mod*

[101] [[一 天] **晚上**] [**曼希沃**] 有 人 请 吃饭， [**高脱弗烈特** 一个人] 待 在
yi tian *wanshang* *Manxiwo* *you* *ren* *qing* *chifan* *gaotuofuliete* *yigeren* *dai* *zai*
un jour soir Melchior avoir ON inviter dîner Gottfried seul rester PREP
nco-sans-mod *nco-mod* *npr-sans-mod* *npr-mod*

Galmiche (1989) parle des *adjoints* des SN. Charolles (2002) parle des *SN complets* et *incomplets*. Dubois *et al.* (2012) et Schnedecker & Longo (2012) utilisent le terme *expansion* pour désigner le même type de composant dans les expressions référentielles nominales, en les distinguant par *SN expansé* et *non expansé*. Citons enfin la définition de Dubois *et al.* (2012 : 191) : « Est expansion dans une phrase tout terme ou tout groupe de termes que l'on peut supprimer de la phrase sans que celle-ci cesse d'être une phrase et sans que les rapports grammaticaux entre les termes soient modifiés. » Nous utilisons le terme *modifieur* d'Ariel dans notre annotation.

[17] *Cf.* Dubois & Lagane (1973), Teng (1981), Anscombre (1986), Corblin (1987), Leroy (2004), Liu (刘丹青, 2005), Yue (岳耀, 2010), Dubois (2012), Dubois *et al.* (2012), Li (2016), Riegel *et al.* (2016 [1994]).

[[楼] 下]， [鲁意莎] 安排 [两 个 小的] 去 睡觉 了， 他 便 出去 坐 在
lou xia Luyisha anpai liang ge xiaode qu shuijiao le ta bian chuqu zuo zai
étage dessous Louisa arranger deux CL petit aller dormir LE 3P donc sortir s'asseoir PREP
nco-mod nco-sans-mod npr-sans-mod nco-sans-mod

[[屋子] 附近的 河边]。
wuzi fujinde hebian
maison proche fleuve-bord
nco-sans-mod nco-mod

[102] Il y avait dans [**ces pages**] [**des chevaux emportés**], [**des épées**], [**les cris de** [**la guerre**]],
nco-sans-mod nco-mod nco-sans-mod nco-mod nco-sans-mod

[**l'orgueil** [**du triomphe**]].
nco-mod nco-sans-mod

[103] 在 [那些 乐章] 中， 有 [飞奔的 马]， [[刀剑] 的 击触]， [[战争] 的 呐喊]，
zai naxie yuezhang zhong you feibende ma daojian de chuji zhanzheng de nahan
PREP ces-là page dedans avoir emporté cheval épée DE heurt guerre DE cri
nco-sans-mod nco-mod nco-sans-mod nco-mod nco-sans-mod nco-mod

[[胜利] 的 欢呼]！
shengli de huanhu
triomphe DE cri
nco-sans-mod nco-mod

5.2.2.5.4. Échelles hypothétiques d'accessibilité des expressions référentielles en chinois

Selon l'observation de l'usage de chaque forme d'expressions référentielles, nous proposons deux modèles hypothétiques d'expressions référentielles chinoises : une échelle d'accessibilité, comme le montre la figure (5.9), et une hiérarchie du donné, comme le montre la figure (5.10). Les différences entre le français et le chinois sont mises en gras dans les deux figures ajustées. Les places de « description sans déterminant longue / courte » en chinois sont substituées à « description définie longue / courte » en français ; les groupes « pronoms accentués », « pronoms inaccentués », « pronoms clitiques » ainsi que celui qui comprend « zéro, réfléchis, marqueurs d'accord » en français sont remplacés par un autre regroupement des formes non nominales : « pronom personnel », « déictique », « déterminant possessif », « zéro » et « autres pronoms ».

Figure (5.9) : Échelle d'accessibilité hypothétique des expressions référentielles chinoises

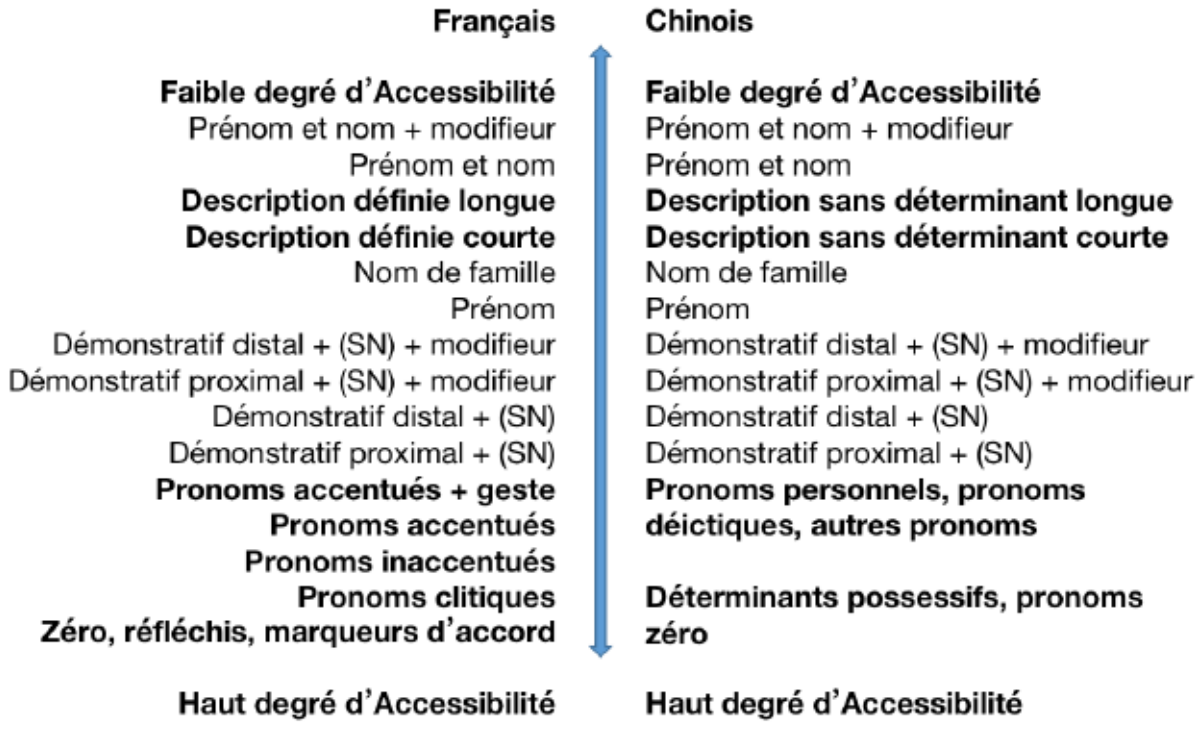
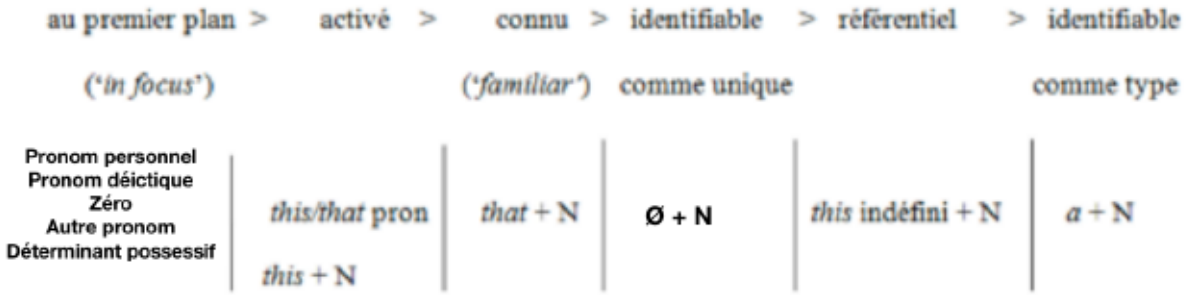


Figure (5.10) : Hiérarchie du donné hypothétique des expressions référentielles chinoises



5.2.3. Bilan

Après avoir effectué toutes les annotations, nous proposons, en guise de bilan, de montrer dans la figure (5.3) de la section *5.2.2. Annotation des caractéristiques des mentions*, les étiquettes utilisées pour annoter les propriétés linguistiques des mentions dans notre corpus. Reprenons les extraits parallèles [13] et [14] de la section *5.2.1. Annotation des mentions : singletons et maillons* comme exemples : une expression référentielle française comme « le nouveau-né » est annotée par six propriétés :

- place de mention : « maillon-1 » ;
- catégorie sémantique de référent : « personnage » ;
- mode de référence : « défini » ;
- fonction syntaxique : « sujet » ;

- type de déterminant : « SN défini » ;
- type de tête lexicale et de modifieur : « nom commun sans modifieur ».

Chushengde-yinger 初生的婴儿 'nouveau-né bébé' de l'extrait [14] en chinois est annoté par les mêmes étiquettes, excepté le type de déterminant comme « SN sans déterminant ».

5.3. Phase 3 – évaluations et calculs

Cette phase finale concerne les évaluations des annotations manuelles dans le corpus ainsi que la saisie des données quantitatives des annotations.

Il n'est pas difficile de comprendre que l'annotation manuelle d'une seule personne n'est pas suffisamment fiable et stable en raison de la fatigue, de l'interprétation personnelle et même parfois des erreurs, malgré le suivi attentif des conventions d'annotation, suivant les remarques de Mathet & Widlöcher (2016) : d'un côté, le résultat de l'annotation d'une seule personne ne peut suffire ; d'un autre côté, il n'est pas nécessaire de faire annoter le même texte par plus de deux personnes. Il en résulte que le meilleur nombre de personnes pour l'annotation d'un texte est deux. L'évaluation par d'autres annotateurs paraît donc importante (Xue *et al.*, 2002 ; Guo, 郭纯洁, 2015 ; Mathet & Widlöcher, 2016 ; Quignard *et al.*, 2021) afin d'éviter en particulier les biais interprétatifs subjectifs, et une double annotation est donc souhaitable dans le but de vérifier et de garantir la fiabilité de l'annotation manuelle. La personne qui effectue la seconde annotation est censée suivre les mêmes conventions d'annotation que celles du premier annotateur, en annotant séparément, sans discussion pendant l'annotation, ce qui permet d'aboutir à un *degré de consensualité* inter-annotateurs du même texte par deux personnes (Mathet & Widlöcher, 2016 ; Quignard *et al.*, 2021).

Le taux de consensualité reflète la proportion de désaccords ou d'erreurs éventuels entre les deux annotations. Plus le degré de consensualité est élevé, moins nombreuses sont les différences entre les deux annotations, et plus fiable est l'annotation de la première personne. Avant de procéder à l'évaluation inter-annotateurs, nous avons rectifié les conventions d'annotation ainsi que les annotations une vingtaine de fois afin de garantir la qualité de la première annotation. Après nos rectifications, une seconde annotation d'un échantillon[18] du corpus en français a été réalisée anonymement par un membre du projet DEMOCRAT. Deux échantillons de mille mots, au début des deux premiers épisodes, ont été annotés deux fois selon les mêmes conventions d'annotation ; l'annotation a été évaluée de la même manière que celle du corpus DEMOCRAT ; autrement dit, tout le corpus DEMOCRAT n'est pas doublement annoté. Certains échantillons sont tirés au hasard pour être soumis à la double annotation servant à calculer le taux de consensualité. Si les deux annotations sont

[18] Mathet & Widlöcher (2016) signalent qu'une double annotation sur un échantillon de corpus, au lieu de sa totalité, est suffisante pour garantir la consensualité de l'annotation : « Si plusieurs annotateurs sont globalement d'accord sur leurs annotations d'une partie d'un corpus, alors chacun d'eux devrait produire à peu près la même chose que ses collègues sur les autres parties du corpus. De la sorte, il n'est plus guère besoin de faire annoter l'intégralité du corpus par plusieurs annotateurs, mais on peut affecter chaque partie à un annotateur particulier. »

largement consensuelles, suite au calcul des désaccords, le corpus annoté constituera une référence en matière de fiabilité ; si, au contraire, le résultat des deux annotateurs affiche un taux élevé de désaccords, il faut absolument revoir la méthodologie de l'annotation : le problème est dû soit à des raisons linguistiques, soit à des raisons informatiques. Le taux d'accords visé dans le projet DEMOCRAT est de 0,75, ce qui signifie que 75 mentions sont annotées de la même manière toutes les 100 mentions. Une annotation parfaite, sans erreur n'existe pas, malgré les rectifications et les vérifications pendant et après les annotations par le même annotateur. Ce taux de consensualité est donc considéré comme acceptable pour les annotations du corpus DEMOCRAT ainsi que pour le nôtre. Selon le même mode de calcul que celui du corpus DEMOCRAT (Quignard *et al.*, 2021), le taux de consensualité de notre double annotation atteint 0,85 pour le premier échantillon et 0,79 pour le second. Ces deux taux dépassent donc le taux visé du corpus DEMOCRAT, premier grand corpus annoté en expressions référentielles et en chaînes de référence en français. Sachant qu'il n'existe aucune double annotation avec un consensus de cent pour cent, nous nous contentons de ce résultat de double annotation.

À la dernière étape du processus de constitution de notre corpus, les calculs et la visualisation servent à présenter les expressions référentielles et les chaînes de référence dans le corpus de manière quantitative. Comme nous l'avons dit dans la section *4.3. Outil de calcul et de visualisation : CRViewer* du *Chapitre 4*, les résultats des calculs et la visualisation des résultats sont réalisés par CRViewer en tableaux de chiffres, en histogrammes et en diagrammes en camembert.

5.4. Bilan

Nous avons cherché à constituer un corpus annoté en chaînes de référence en français et en chinois en fournissant les informations linguistiques les plus riches possibles sur toutes les expressions référentielles, y compris les catégories sémantiques des référents mentionnés dans le roman-fleuve, les fonctions syntaxiques des expressions référentielles, les formes spécifiques des expressions nominales (déterminants, têtes lexicales et modifieurs) et non nominales, les places des expressions référentielles par rapport aux chaînes et les modes de référence des expressions. Toutes les annotations ont été faites de manière exhaustive, mais certaines d'entre elles relatives aux caractéristiques linguistiques n'ont pas servi à expliquer les phénomènes des chaînes de référence. Par exemple, comme nous l'avions supposé, l'annotation de « mode de référence » n'est pas vraiment utile aux analyses des chaînes parce que peu de chaînes concernent les référents indéfinis.

Nous avons choisi d'annoter toutes les expressions référentielles car toutes sont censées constituer des chaînes, même si ce n'est finalement pas le cas. Après les avoir toutes annotées, nous avons obtenu une grande partie de mentions en singleton, et seulement une petite partie de mentions dans les chaînes.

Nous avons adapté les conventions monolingues du français et du chinois à un corpus bilingue. Par exemple, la distinction entre les compléments d'objet directs et indirects pour le français est éliminée dans notre guide franco-chinois ; l'étiquette « SN défini » est réservée aux textes en français ; celle de « nom propre » s'adapte à des critères formels pour le français et sémantiques pour le chinois.

L'annotation n'est pas réalisée une fois pour toutes : sa rectification, ainsi que celle des conventions d'annotation, sont effectuées périodiquement pour que ces dernières soient plus adaptées aux deux langues. En raison de son volume, chaque rectification de l'intégralité du corpus a duré quelques semaines. Nous avons dû rectifier une trentaine fois afin de réajuster systématiquement tous les critères d'annotation dans nos conventions et de diminuer le nombre d'annotations erronées aléatoires au minimum. Par ailleurs, pour prévenir d'éventuels désaccords ou des erreurs d'annotation individuelles, une double annotation nous a semblé la meilleure solution pour assurer la qualité d'un corpus annoté. Après de nombreuses rectifications d'annotations effectuées par nous-même, une deuxième annotation a été élaborée par un autre membre du projet DEMOCRAT sur deux échantillons de mille mots du corpus.

Notre corpus totalise 61 862 mots en français et 89 858 caractères chinois en bi-texte. Le premier livre de *Jean-Christophe* est découpé en 50 épisodes selon les trois principes de découpage que nous avons présentés dans la section sur le processus. Le nombre de mots français et celui du chinois révèlent un coefficient de 100 : 145, signifiant que cent mots français correspondent à 145 caractères parallèles chinois. Ce coefficient est similaire à la fois dans chaque épisode de texte et dans l'ensemble des 50 épisodes. La relative stabilité de ce coefficient signifie que le traducteur a respecté le même principe et le même style de traduction du début jusqu'à la fin des extraits que nous avons choisis dans ce roman-fleuve. Il est ainsi prévisible que certains indices de chaînes de référence relatifs à la longueur des mots ou des textes en chinois seront supérieurs à ceux du français en raison de cette longueur différente du texte. Par exemple, il se peut que la longueur des expressions référentielles et celle des maillons soit 1,45 fois plus importante en chinois qu'en français.

Mis à part le nombre de mots / caractères, les autres données sont issues de notre annotation manuelle du texte. Nous avons annoté toutes les expressions référentielles pour aboutir au nombre de 17 367 en français et 16 407 en chinois, au sein desquelles seulement 9 687 en français et 9 231 en chinois sont des maillons de chaînes. Autrement dit, environ la moitié des expressions référentielles annotées ne font pas l'objet de chaînes de référence. Nous ne prenons pas en compte les caractéristiques de ces expressions référentielles qui sont exclues des chaînes de référence. Les expressions référentielles renvoient respectivement à 7 875 référents mentionnés en français et 7 343 référents en chinois. Parmi ces expressions, 9 687 maillons en français et 9 231 maillons en chinois constituent respectivement 679 chaînes de référence en français et 656 chaînes de référence en chinois[19]. Les chiffres absolus du corpus et de chaque épisode sont présentés dans les tableaux (5.6) et (5.7) ci-après. Il s'agit du nombre de mots / caractères, de référents, d'expressions référentielles, de maillons et de chaînes de référence.

[19] Ces chiffres sont obtenus sur la base des 50 épisodes, c'est-à-dire que si un référent comme « Jean-Christophe » apparaît dans les 50 épisodes et fait chaîne dans 48 épisodes, il est compté comme « 50 référents » et « 48 chaînes » dans ce tableau.

Tableau (5.6) : Présentation quantitative du corpus

Nbre de	Chinois	Français	Total
Mots / caractères	89 858	61 862	151 720
Expressions référentielles	16 407	17 367	33 774
Référents	7 343	7 875	15 218
Maillons	9 231	9 687	18 918
Chaînes de référence	656	679	1 335

Tableau (5.7) : Présentation quantitative du corpus en épisodes

Nbre de	Mots / caractères		Paragraphes		Référents		ER		CR		Maillons	
	Fr	Ch	Fr	Ch	Fr	Ch	Fr	Ch	Fr	Ch	Fr	Ch
1	1588	2234	50	43	166	142	454	398	10	12	279	253
2	988	1494	5	5	116	119	257	263	12	9	149	149
3	526	982	16	16	48	72	153	184	7	10	108	119
4	850	1165	9	8	139	134	241	210	15	15	106	84
5	1188	1739	7	7	171	177	334	335	14	17	163	162
6	712	1056	7	7	91	95	210	218	13	14	123	130
7	913	1246	5	5	114	116	267	249	16	9	160	139
8	1345	1848	8	8	149	133	355	321	16	13	207	191
9	1822	2557	9	9	237	240	496	467	22	15	264	227
10	553	772	6	6	101	106	177	168	8	4	74	55
11	2117	3025	8	8	301	239	573	522	23	24	279	296
12	660	916	7	7	66	64	182	169	11	9	125	110
13	2089	3025	16	15	207	181	618	589	25	24	424	418
14	1306	1810	13	12	140	123	373	341	8	9	233	221
15	798	1118	19	18	77	70	237	193	12	8	167	129
16	1647	2471	11	8	188	170	517	490	12	19	320	325
17	411	558	3	3	56	48	125	111	6	4	73	61
18	2400	3567	14	14	380	393	688	674	29	31	303	278
19	771	1072	11	10	66	67	213	199	8	5	153	133
20	1490	2240	17	14	185	186	397	432	17	18	218	250
21	1433	2031	8	8	173	169	397	367	18	18	233	208
22	2038	2910	8	9	281	280	520	524	17	20	238	246
23	1050	1545	19	15	126	110	295	290	4	9	157	178
24	1196	1865	5	5	198	204	308	305	9	8	105	96
25	1533	2143	9	9	221	192	417	376	13	13	191	171
26	1011	1411	17	13	124	120	267	249	13	11	147	134
27	923	1381	4	4	150	138	267	261	9	10	111	117
28	1208	1809	18	19	146	120	387	329	14	11	245	208
29	619	873	4	4	102	113	170	169	6	4	71	56
30	1778	2525	35	33	194	191	478	484	19	17	295	299
31	1112	1667	4	4	155	131	319	302	12	13	166	178
32	1958	2853	61	49	168	205	506	524	25	20	351	328
33	1477	2187	30	25	124	112	403	356	17	18	285	259
34	1092	1290	11	9	110	99	214	206	9	13	106	114
35	2714	3873	19	17	348	271	795	717	36	35	465	454
36	1039	1519	12	10	127	118	297	279	20	18	183	167
37	1098	1597	5	5	137	143	316	306	17	14	188	168
38	907	1479	14	14	130	111	250	237	11	13	124	132
39	908	1313	3	3	127	114	254	236	8	8	131	125
40	1017	1482	4	4	151	140	275	262	13	11	127	123
41	1183	1682	3	3	192	153	366	313	10	11	179	156

42	2101	3112	42	38	232	212	610	562	29	24	400	352
43	688	1026	5	5	108	87	183	171	10	11	82	93
44	1172	1693	5	5	156	128	344	299	10	10	191	179
45	929	1450	30	27	102	81	283	263	9	8	189	184
46	1360	2096	35	28	147	142	409	394	11	10	264	258
47	1223	1793	7	7	174	153	360	323	11	9	189	175
48	947	1374	6	5	163	171	249	244	4	3	89	73
49	1218	1756	4	3	182	142	346	301	9	11	171	162
50	756	1228	4	5	129	118	215	225	2	6	86	108
Total	61862	89858	672	608	7875	7343	17367	16407	679	656	9687	9231
Moyenne	1237,2	1797,2	13,4	12,2	157,5	146,9	347,3	328,1	13,6	13,1	193,7	184,6
Total	151720		1280		15218		33774		1335		18918	

En moyenne, 100 référents dans les textes en français se réduisent à 93 référents en chinois ; 100 expressions référentielles en français sont traduites par 94 expressions référentielles en chinois ; 100 chaînes de référence en français correspondent à 97 chaînes en chinois et 100 maillons en français à 95 maillons en chinois.

Étant donné la démarche de constitution de notre corpus bilingue *Jean-Christophe*, les textes sont partis d'un état original, comme le montre le fragment [104], avec les titres, les notes et les images, et sont arrivés, en passant par un état primaire avec des annotations basiques, à l'état constitué [105], avec les annotations complètes des expressions référentielles, des chaînes et des propriétés linguistiques.

[104] AUX ÂMES LIBRES DE TOUTES LES NATIONS QUI SOUFFRENT, QUI LUTTENT ET QUI VAINCRONT.
<tome=R01>
L'aube
Dianzi, nell'alba che precede al giorno Quando l'anima tua dentro dormia...
PURG.IX
<partie=R011>
I.
Come, quando i vapori umidi e spessi A diradar cominciansi, la spera Del sol debilemente entra per essi…
PURG. XVII.
Le grondement du fleuve monte derrière la maison. La pluie bat les carreaux depuis le commencement du jour. Une buée d'eau ruisselle sur la vitre au coin fêlé. Le jour jaunâtre s'éteint. Il fait tiède et fade dans la chambre.

[105] {GrondementDuFleuve Le grondement {Rhin du fleuve}} monte derrière {Maison la maison}. {Pluie La pluie} bat {M1 les carreaux} depuis {M2 le commencement {JOUR du jour}}. {M3 Une buée d'eau} ruisselle sur {Vitre la vitre} {M4 au coin fêlé}. {M5 Le jour jaunâtre} s'éteint. Il fait tiède et fade dans {Chambre la chambre}.
{GrondementDuFleuve:cat="g generique",determinant="d N defini",expansion="x SN expanse",fonction="s sujet",place="n nul",type="o objet concret" Le grondement {Rhin:cat="d defini",determinant="d N defini",expansion="o SN nu",fonction="n complement du nom",place="n nul",type="o objet concret" du fleuve}} monte derrière {Maison:cat="d defini",determinant="d N defini",expansion="o SN nu",fonction="c circonstant",place="n nul",type="l lieu" la maison}. {Pluie:cat="g generique",determinant="d N defini",expansion="o SN nu",fonction="s sujet",place="n nul",type="o objet concret" La pluie} bat {M1:cat="d defini",determinant="d N defini",expansion="o SN nu",fonction="o complement d objet",place="n nul",type="o objet concret" les carreaux} depuis {M2:cat="d defini",determinant="d N defini",expansion="x SN expanse",fonction="c circonstant",place="n nul",type="t temps" le commencement {JOUR:cat="g generique",determinant="d N defini",expansion="o SN nu",fonction="n complement du nom",place="n nul",type="t temps" du jour}}. {M3:cat="i indefini",determinant="i N indefini",expansion="o SN nu",fonction="s sujet",place="n nul",type="o objet concret" Une buée d'eau} ruisselle sur {Vitre:cat="d defini",determinant="d N defini",expansion="o SN nu",fonction="c circonstant",place="n nul",type="o objet concret" la vitre} {M4:cat="d defini",determinant="d N defini",expansion="x SN expanse",fonction="c circonstant",place="n nul",type="l lieu" au coin fêlé}. {M5:cat="d defini",determinant="d N defini",expansion="x SN expanse",fonction="s sujet",place="n nul",type="t temps" Le jour jaunâtre} s'éteint. Il fait tiède et fade dans {Chambre:cat="d defini",determinant="d N defini",expansion="o SN nu",fonction="c circonstant",place="n nul",type="l lieu" la chambre}.

Une partie du corpus de notre thèse fait également partie du corpus DEMOCRAT : il s'agit des 20 000 premiers mots de *Jean-Christophe* en français annotés. Les textes ainsi que les annotations ont été publiés gratuitement en 2019 sur la plateforme Ortolang.

Bilan de la *Partie II*

Il nous paraît trop ambitieux de parler de la comparaison des chaînes de référence en français et en chinois de manière générale, étant donné les nombreuses limites imposées au corpus par les critères comme les époques, les genres et les styles d'auteurs dans les deux langues. Par conséquent, notre travail se fonde sur l'observation des différences des chaînes de référence à partir d'une étude de cas basée sur des textes spécifiques, à savoir les extraits du roman-fleuve français *Jean-Christophe* et sa traduction en chinois par Lei Fu, en langues contemporaines, qui répondent à divers critères de qualité.

Jean-Christophe a un statut représentatif selon deux aspects. D'un côté, en tant que son auteur a obtenu le Prix Nobel de Littérature, cette œuvre est importante dans le domaine de la littérature franco-chinoise ; d'un autre côté, en tant que genre de discours spécifique, le roman-fleuve comporte des chaînes de référence distinctes de celles d'autres genres. De ce point de vue, *Jean-Christophe* représente donc un corpus exemplaire.

Sur la base des études linguistiques, littéraires et culturelles déjà consacrées à ce livre, nous avons cherché à formuler quelques caractéristiques hypothétiques des chaînes de référence appartenant à ce genre de texte narratif long en français et en chinois. Ces textes sont en effet censés pouvoir nous permettre de repérer la variation et les points communs des expressions référentielles et des chaînes de référence équivalentes en français et en chinois dans des contextes similaires sous tous les aspects : lexical, sémantique, discursif.

L'objectif du découpage du roman-fleuve en épisodes était au début d'avoir des textes courts et homogènes ; mais à la fin, en raison des trois critères de découpage que nous avons utilisés, les épisodes ne présentent pas vraiment des textes courts de longueurs similaires : l'épisode le plus court compte 411 mots français et 558 caractères chinois, alors que le plus long comporte 2714 mots français et 3873 caractères chinois. Néanmoins, cette opération de découpage permet de décrire et de démontrer la répartition des maillons dans les extraits de ce roman-fleuve.

Dans ce chapitre, nous avons essayé de répondre aux questions méthodologiques relatives au traitement des expressions référentielles et des chaînes de référence de notre corpus. La résolution automatique actuelle ne nous a pas semblé pas suffisante pour traiter les expressions coréférentielles en français et en chinois en les unissant dans le même cadre logistique. Vu ces contraintes informatiques, nous avons choisi AiHanYu, SACR et CRViewer qui nous ont paru les plus pertinents pour traiter notre corpus franco-chinois. Ce sont des outils faciles à maîtriser et ils permettent d'économiser beaucoup d'efforts et de temps pendant l'annotation manuelle. Grâce à ces aides, nous avons réalisé le traitement en quatre phases et plusieurs étapes, en suivant les conventions d'annotation que nous pensons être pertinentes et reproductibles pour d'autres corpus franco-chinois. À partir de ce processus de traitement, nous avons obtenu un corpus franco-chinois annoté en expressions référentielles et en chaînes de référence. Toutes les expressions référentielles sont systématiquement enrichies de plusieurs propriétés linguistiques, et, d'autre part, les chaînes sont visualisées et schématisées par diverses formes de représentation. Ce travail rend possible l'obtention de données rentables pour dégager les variations des expressions référentielles et des chaînes de référence en français et en chinois.

Partie III

Observations sur le corpus *Jean-Christophe* : régularités et spécificités contrastives des chaînes de référence en français et en chinois

Chapitre 6 : Régularités des chaînes de référence dans un roman-fleuve français : aspects quantitatifs et qualitatifs

Jean-Christophe est un roman-fleuve de dix tomes. Notre corpus est composé du premier tome intitulé *L'Aube*, subdivisé en trois sections (sections I, II et III), et du premier tiers du deuxième tome intitulé *Le Matin* (section IV), ce qui représente les quatre premières sections situées au début de l'œuvre. Rappelons que nous avons découpé ces quatre sections en 50 épisodes, que nous avons délimités, d'une part, en fonction des marqueurs typographiques de l'œuvre, comme les numéros de sections et les paragraphes, et, d'autre part, selon les marqueurs spatio-temporels situés au début des paragraphes, comme « un soir » et « quelques semaines après ». Les observations des régularités et des spécificités des chaînes de référence en français se fondent donc sur ces textes.

Un roman consiste la plupart du temps en un récit au centre duquel se trouvent des personnages. La récurrence des expressions référant au personnage dominant couvre souvent l'ensemble du texte (*cf.* Schnedecker, 2014, 2019), constituant des chaînes *globales*, alors que les référents moins saillants sont censés former seulement des chaînes *locales* qui couvrent une ou quelques parties du texte. Aussi est-il normal que, parmi toutes les catégories de référents, ceux des humains se positionnent à la première place. Les expressions référentielles sont annotées selon la catégorie sémantique de leur référent par cinq étiquettes : les *humains*, les *entités concrètes*, les *entités abstraites*, les *lieux* et les *temps*. Le tableau (6.1) montre les chiffres des deux langues en parallèle : le pourcentage de chaque catégorie de référents dans la totalité des expressions référentielles, celui des catégories de maillons et, enfin, celui des catégories de référents dans les chaînes. Le nombre et la proportion de chaque type de référents sont assez proches dans les deux langues : la moitié ou légèrement plus des expressions référentielles et des chaînes renvoient à des humains (respectivement 50 % en français et 51 % en chinois pour le premier et 51 % en français et 55 % en chinois pour le second). Pourtant, les maillons occupent une plus grande proportion, respectivement 84 % en français et 83 % en chinois, ce qui indique que les chaînes d'humains sont beaucoup plus longues que celles des autres catégories de référents. Sur le plan syntaxique, les expressions référentielles se rapportant aux humains occupent régulièrement la fonction de sujet dans la phrase ; au niveau discursif, elles fonctionnent souvent comme topiques du discours. Ce sont donc les chaînes des personnages qui attirent tout d'abord notre attention, bien avant celles des autres types de référents.

Tableau (6.1) : Catégories sémantiques des référents dans *Jean-Christophe*

Pourcentage (nbre)	ER		Maillons		CR	
Langue	Fr	Ch	Fr	Ch	Fr	Ch
Humains	50 %	51 %	84 %	83 %	51 % (345)	55 % (360)
Concrets	17 %	19 %	9 %	9 %	32 % (215)	25 % (165)
Abstraits	24 %	17 %	5 %	5 %	8 % (57)	9 %(57)
Lieux	5 %	8 %	2 %	2 %	7 % (45)	10 % (64)
Temps	3 %	5 %	0,7 %	0,6 %	3 % (17)	2 % (10)
Autres	0,1 %	0,1 %	-	-	-	-

L'importance des personnages romanesques est signalée par Reuter (1988) et se manifeste, selon lui, dans les trois aspects suivants, qui pourraient servir d'explication aux résultats de l'annotation de notre corpus :

> *Proposition 1 : le personnage est un* marqueur typologique *: il différencie le récit des autres types de textes et indexe des sous-catégories génériques.*
>
> *Proposition 2 : le personnage est un* organisateur textuel *: il intègre des unités de niveau(x) inférieur(s), s'organise dans la structure du récit avec des éléments de même niveau, et constitue le soubassement nécessaire des configurations sémantiques.*
>
> *Proposition 3 : le personnage est un* lieu d'investissement *dans les textes, pour les producteurs et les récepteurs, de l'expérience du Sujet et du Social.*
>
> *(Reuter, 1988 : 6)*

La place primordiale du personnage est aussi évoquée par les approches cognitives. Dans l'échelle de saillance de Langacker (1991), l'importance des humains est plus grande que celle des autres référents animés et non animés : comme le montre l'échelle suivante, les trois sous-groupes d'humains sont plus saillants que les animés non humains ; les animaux sont plus saillants que les objets physiques ; ces derniers sont plus saillants que les entités abstraites :

> *[Speaker → hearer → human] → animal → physical object → abstract entity*
>
> *(Langacker, 1991 : 307)*

Les référents d'humains sont facilement et rapidement accessibles quelle que soit la catégorie grammaticale d'expression (*cf.* aussi Boucheix & Fayol, 1997 : 53). Le protagoniste et les personnages secondaires ont toujours une saillance plus remarquable que les personnages secondaires (Gernsbacher & Hargreaves, 1988 : 699 ; Boucheix & Fayol, 1997). Cette saillance des protagonistes humains est encore plus remarquable dans les discours du genre narratif ou à dominante narrative (Schnedecker & Longo, 2012 ; Schnedecker, 2017), comme le fait divers (Schnedecker, 2014), la nouvelle et le roman. Étant donné cette importance référentielle cruciale d'humains dans tous les aspects présentés, nous avons choisi d'examiner en premier lieu les chaînes des personnages.

Le tableau (6.2) présente la répartition des chaînes des personnages principaux qui émergent de notre corpus[1]. Les sections de l'œuvre sont découpées et numérotées en épisodes : la section I correspond aux épisodes 1 – 10, la section II aux épisodes 11 – 24, la section III aux épisodes 25 – 36 et la section IV aux épisodes 37 – 50. Chaque ligne du tableau (6.2) correspond à un épisode, et chaque colonne à un personnage, comme le montrent les noms abrégés aux en-têtes : « JC » (« Jean-Christophe »), « L » (« Louisa »), « M » (« Melchior »), « JM » (« Jean-Michel »), « Duc » (« grand-duc »), « ML » (« Melchior et Louisa »), « ER » (« Ernst et Rodolphe ») et « G » (« Gottfried »).

Pour un épisode qui comporte la chaîne d'un personnage, les cases sont marquées par le nombre de maillons de la chaîne dans cet épisode ; pour ceux où les chaînes des personnages sont absentes, c'est-à-dire ceux dénués d'expressions renvoyant aux personnages ou n'en comportant qu'une ou deux, les cases restent vides. Par exemple, la première ligne du

[1] En raison de la largeur limitée du tableau, toutes les chaînes des personnages du livre ne sont pas montrées dans le tableau (6.2).

tableau montre la répartition des chaînes des personnages dans le premier épisode : quatre chaînes renvoyant à quatre personnages émergent : celle de « Jean-Christophe », qui comprend 49 maillons, celle de « Louisa » avec 91 maillons, celle de « Melchior », qui compte 26 maillons et celle de « Jean-Michel », qui comporte 89 maillons. Les chaînes des personnages comme « grand-duc », « Melchior Louisa », « Ernest Rodolphe » et « Gottfried » sont absentes dans le premier épisode. À l'exception des singletons, beaucoup de personnages n'apparaissent que dans quelques épisodes (souvent dans 1 – 3 épisodes) avec un petit nombre de maillons (souvent moins de dix). Nous nommons ces chaînes de personnages *épisodiques* et *éphémères*. Dans notre échelle des chaînes, celles du protagoniste et des personnages secondaires correspondent à la définition des chaînes *globales*, alors que toutes les autres sont *locales*.

Tableau (6.2) : Répartition des chaînes des personnages dans le corpus en français

Section	Épisode	JC	L	M	JM	Duc	ML	ER	G
S I	1	48	91	26	88	-	-	-	-
	2	3	28	55	18	-	-	-	-
	3	38	17	4	28	-	-	-	-
	4	20	20	4	13	-	-	-	-
	5	92	11	4	-	-	-	-	-
	6	72	4	-	5	-	-	-	-
	7	86	-	-	-	-	-	-	-
	8	42	-	-	87	-	-	-	-
	9	95	-	-	33	-	-	-	-
	10	47	9	-	-	-	-	-	-
S II	11	-	-	61	118	4	-	-	-
	12	59	9	7	-	-	4	17	-
	13	240	42	-	-	-	-	-	-
	14	149	17	14	-	-	12	-	-
	15	77	37	22	-	-	-	-	-
	16	133	36	117	-	-	-	-	-
	17	54	-	4	-	-	3	-	-
	18	180	-	4	19	-	4	-	-
	19	78	44	-	-	-	-	-	-
	20	125	15	8	-	-	4	-	-
	21	119	4	21	6	-	-	-	-
	22	128	3	19	-	-	-	-	-
	23	131	-	35	-	-	3	-	-
	24	47	-	-	-	-	-	-	-
S III	25	113	-	-	37	-	-	-	-
	26	74	-	-	28	-	-	-	-
	27	43	-	3	-	4	-	-	-
	28	85	-	-	16	4	-	-	-
	29	45	4	-	-	-	-	-	-
	30	154	-	16	77	-	3	-	-
	31	38	10	4	9	-	-	-	70
	32	134	-	-	-	-	-	-	100
	33	110	-	18	21	3	-	-	57
	34	49	-	30	6	7	-	-	-
	35	246	-	38	30	21	-	-	-
	36	112	5	12	8	-	6	-	-
S IV	37	107	-	-	4	7	-	-	-
	38	49	7	5	20	-	-	-	-

	39	45	-	3	25	-	-	-	-
	40	69	-	-	3	-	-	-	20
	41	3	9	18	128	-	-	-	-
	42	166	26	-	74	-	-	12	34
	43	28	8	7	5	-	-	-	6
	44	60	9	75	11	-	-	-	-
	45	78	7	77	-	-	-	13	-
	46	139	13	26	-	-	-	-	-
	47	98	13	43	-	-	-	3	-
	48	82	-	-	-	-	-	-	-
	49	109	13	9	-	-	-	19	-
	50	81	-	-	-	-	-	-	-
Nbre de maillons (nbre d'épisodes)		4366 (49)	511 (28)	789 (32)	917 (27)	50 (7)	39 (8)	64 (5)	287 (6)

D'après le tableau (6.2) les différences de répartition entre les chaînes des personnages sont très importantes. Elles se manifestent par quatre aspects : la couverture de la chaîne (*i.e.* pourcentage de texte couvert par la chaîne), sa longueur ainsi que son démarrage et sa fin.

Premièrement, chaque chaîne couvre un seul ou plusieurs épisodes continus ou discontinus. Les cases remplies par le nombre de la chaîne en question désignent sa continuité dans des épisodes successifs, alors que les cases vides indiquent les moments de discontinuité dans les épisodes correspondants. Par exemple, la chaîne « Jean-Christophe » montre une continuité presque permanente dans le corpus qui couvre 49 épisodes, ce qui équivaut à 98 % du texte. Le seul moment de discontinuité se situe à l'épisode 11, auquel aucune expression ne renvoie. Par comparaison, les maillons de la chaîne « Ernest Rodolphe » se trouvent dans les épisodes 12, 42, 45, 47 et 49 ; cette chaîne, qui n'existe dans aucun des deux épisodes subséquents du corpus, illustre une grande discontinuité entre les épisodes. Ni les personnages, ni leurs maillons ne sont répartis dans les sections et dans les épisodes du corpus entier. Une seule chaîne (« Jean-Christophe ») s'étend à la majorité du corpus ; trois (« Louisa », « Melchior » et « Jean-Michel ») en couvrent environ une moitié, entre 27 – 32 épisodes, équivalents à 54 % – 64 % du corpus ; les autres n'occupent que quelques épisodes (5 – 8), c'est-à-dire 10 % – 16 % du corpus.

Deuxièmement, en ce qui concerne la longueur, la chaîne « Jean-Christophe » comporte le nombre de maillons le plus élevé (4366), et dépasse largement toutes les autres, qui comptent chacune moins de 1000 maillons. En effet, les personnages peuvent être facilement classés en trois groupes selon le nombre d'épisodes qu'ils couvrent et selon leur nombre de maillons : « Jean-Christophe » est le seul personnage dont la chaîne comporte plus de 4000 maillons, occupant la majorité du corpus ; « Louisa », « Melchior » et « Jean-Michel » sont les seuls dont la chaîne comporte entre 500 – 1000 maillons et couvre plus d'une moitié du corpus ; les chaînes de tous les autres personnages comportent moins de 300 maillons et, surtout, elles concernent moins de 20 % du corpus. Ce groupe de personnages inclut ceux qui se trouvent dans le tableau (6.2) et ceux que nous présenterons plus tard dans le tableau des « chaînes des personnages épisodiques et des personnages éphémères », qui seront détaillées dans la section *6.3. Chaînes des personnages épisodiques et éphémères* de ce chapitre.

Troisièmement, selon le tableau (6.2), toutes les chaînes n'apparaissent pas au même moment dans l'œuvre. Les quatre personnages qui ont les maillons les plus nombreux sont aussi ceux qui apparaissent le plus tôt, dès le premier épisode ; tous les autres entrent en

scène plus tard, à partir de la deuxième section (« grand-duc », « Melchior Louisa » et « Ernest Rodolphe ») et de la troisième (« Gottfried »). Les chaînes de ces personnages peuvent être absentes au début de l'œuvre, au milieu ou à la fin. Par exemple, la chaîne « Louisa » est absente dans les épisodes 23 – 28 ; celle de « Jean-Michel » l'est des épisodes 12 – 17 ; celle de « Jean-Michel » disparaît à partir de l'épisode 45 ; celle du « grand-duc » s'interrompt à partir de l'épisode 38 ; celle de « Melchior Louisa » est complètement absente dans la section IV, composée par les épisodes 37 – 50.

Par conséquent, la présentation des chaînes françaises dans le corpus *Jean-Christophe* suivra l'ordre de ces quatre groupes :

- le groupe 1 concerne une seule chaîne globale, celle du *protagoniste*, qui est répartie dans tout le corpus, avec plus de 4000 maillons ;
- le groupe 2 intéresse trois chaînes globales, à savoir celles des *personnages secondaires* : « Louisa », mère du protagoniste, « Melchior », son père, et « Jean-Michel », son grand-père. Elles sont distribuées dans une moitié environ du corpus, et leur nombre de maillons se situe entre 500 et 1000 ;
- le groupe 3 est relatif aux chaînes locales des personnages, c'est-à-dire celles des *personnages épisodiques* et *éphémères* qui sont réparties dans quelques épisodes avec des nombres de maillons moins élevés ;
- le groupe 4 concerne les chaînes locales des référents non humains.

6.1. Chaîne du personnage « Jean-Christophe »

Deux raisons nous poussent à observer en premier lieu la chaîne du personnage « Jean-Christophe ». D'abord, le titre même de ce roman-fleuve : il s'agit d'un nom propre qui dénomme un personnage, ce qui laisse présager que l'intrigue sera centrée sur celui-ci. Ensuite, *Jean-Christophe* est appelé *roman passif* par Leblond (2010 : 64), qui situe ce type d'œuvre à l'échelle de l'individu. Un roman passif « s'attache aux pas d'un personnage » et « s'organise autour d'une vie », celle du protagoniste ; l'ordre du récit est « donné du dehors par l'unité d'une époque ou d'une existence humaine ». À la notion de *roman passif* s'oppose celle de *roman actif*, qui se situe à l'échelle du monde, à savoir que l'ordre du roman « est créé par une libre disposition du romancier » (Leblond, 2010 : 63). La structure d'un roman actif se rapproche de celle du drame, « tendue vers un dénouement » (Lebond, 2010 : 66). Dans notre cas, le titre implique la place primordiale de l'existence humaine de Jean-Christophe. Ainsi, la chaîne « Jean-Christophe » est supposée être celle du seul personnage dominant du début à la fin du roman. C'est pour ces deux raisons que nous présenterons la configuration de la chaîne « Jean-Christophe » en premier lieu, avant d'aborder les particularités des autres chaînes de personnages et d'autres types de référents. D'après le tableau (6.2), la chaîne « Jean-Christophe » confirme l'hypothèse de sa place centrale dans un roman passif et s'avère être la seule chaîne continue ainsi que la plus longue du corpus en français et en chinois. Dans cette section, la chaîne « Jean-Christophe » en français sera examinée selon les questions suivantes :

- de quelles expressions se compose-t-elle ?
- comment le personnage est-il introduit et comment sa chaîne redémarre-t-elle dans l'œuvre ?

- pour quelles raisons et dans quels contextes phrastiques et discursifs les maillons sont-ils utilisés ?

6.1.1. Présentation générale

La composition de la chaîne sera tout d'abord présentée sous l'angle de la proportion des maillons de chaque catégorie morpho-syntaxique. Ensuite, deux indices reflètent quelques-unes de ses caractéristiques dans sa totalité : la longueur moyenne des maillons représente la composition de la chaîne et la densité référentielle indique sa répartition dans le roman-fleuve.

6.1.1.1. Catégories grammaticales des maillons

D'après les résultats de notre annotation, la chaîne « Jean-Christophe » comprend 16 % de maillons nominaux, dont 11 % sont des noms propres (« Christophe », « monsieur Krafft »), 3 % des SN définis (« le nouveau-né », « l'enfant »), 1 % des SN possessifs (« son petit-fils », « son garçon »), 0,2 % des SN démonstratifs (« cet enfant », « ce gamin »), 0,2 % des SN indéfinis (« un public complaisant », « un petit garçon »), 0,2 % des SN sans déterminant (« gamin »), et d'un seul maillon (égal à moins de 0,1 %) qui est SN exclamatif. Les 84 % de maillons restants sont incarnés par des catégories non nominales, à savoir les pronoms personnels de la troisième personne (60 %, comme « il » et « lui »), les déterminants possessifs (16 %, comme « son » et « sa »), les déictiques (4 %, comme « je » et « tu »), les zéros (4 %) et les autres pronoms (0,1 %), y compris principalement les pronoms démonstratifs (comme « ce » et « cela ») et indéfinis (comme « certains » et « tout le monde »). La proportion de toutes ces catégories est synthétisée dans le tableau (6.3) suivant.

Tableau (6.3) : Proportion des maillons de la chaîne « Jean-Christophe » selon la catégorie morpho-syntaxique

Maillon	Catégorie	Pourcentage (nbre de maillons)	
Nominales	Nom propre	11 % (503)	16 %
	SN défini	3 % (139)	
	SN possessif	1 % (37)	
	SN démonstratif	0,2 % (10)	
	SN indéfini	0,2 % (8)	
	SN sans déterminant	0,2 % (8)	
	Autres	0,1 % (1)	
Non nominales	Pronom personnel de troisième personne	60 % (2643)	84 %
	Déterminant possessif	16 % (720)	
	Déictique	4 % (171)	
	Zéro	4 % (154)	
	Autres pronoms	0,1 % (3)	

Avant de préciser les endroits d'occurrences et les conditions spécifiques d'usage de chaque catégorie, quelques hypothèses peuvent déjà être formulées.

Pour évoquer le protagoniste, le nom propre est la catégorie privilégiée par rapport à celle des noms communs de divers déterminants et sans déterminant. Est-ce que d'autres personnages du roman-fleuve sont évoqués aussi souvent par des noms propres ? Comment le protagoniste réalise-t-il sa saillance en se rapportant aux autres référents dans un long texte narratif ? Est-ce que les noms propres et les noms communs se répartissent de manière équilibrée dans les différents types de séquences textuelles ?

Les pronoms personnels de troisième personne, à la première place, ainsi que les déterminants possessifs, à la deuxième, occupent environ trois quarts de la chaîne et semblent jouer le rôle fondamental de maintien de la continuité de la chaîne d'un bout à l'autre de l'œuvre. Les reprises du protagoniste par ces deux catégories ne sont pas censées susciter d'ambiguïtés référentielles ; ce sont les contextes qui ne contiennent aucun autre personnage masculin au singulier, ce qui favorisent leur usage et le fait que le protagoniste exprimé par une forme nominale est le référent le plus proche de ces reprises pronominales, selon le principe de proximité. Pourtant, sachant que le roman-fleuve est susceptible de contenir plusieurs référents, comment les pronoms personnels de troisième personne et les déterminants possessifs permettent-ils la continuité de la chaîne du protagoniste sans susciter d'éventuelles ambiguïtés dans les situations multi-référentielles ? Jusqu'à quelle distance inter-maillonnaire le principe de proximité peut-il fonctionner et jusqu'à combien de référents concurrents ?

Les 4 % de déictiques sont répartis *a priori* dans les séquences dialogales. Leur proportion peut refléter celle des séquences dialogales dans ce genre narratif. Mais existe-t-il des séquences narratives qui contiennent également des déictiques ayant des usages exceptionnels dans le roman-fleuve ?

Les zéros sont utilisés à partir de la deuxième proposition coordonnée au même sujet que la première. 4 % de zéros suggèrent ainsi une proportion similaire des propositions coordonnées. A l'exception de cet emploi de zéro sous cette condition phrastique spécifique, existe-t-il d'autres contraintes au niveau inter-phrastique qui favorisent ou défavorisent les occurrences de zéro en français dans le roman-fleuve ?

Nous vérifierons, dans les sections suivantes, si les remarques attendues de chaque catégorie peuvent être corroborées, ou s'il y a des phénomènes et des emplois inattendus en fonction de conditions textuelles spécifiques.

6.1.1.2. Densité référentielle et corrélation avec la catégorie des maillons

La densité référentielle est définie par Schnedecker (2021 : 262) comme le « ratio du nombre d'expressions référentielles et du nombre de mots d'un texte » (*cf.* aussi Schnedecker, 2014). Cet indice reflète directement la proportion d'expressions d'un référent dans des textes d'une même longueur. Selon cette définition, la densité de la chaîne « Jean-Christophe » est égale à 0,071, à savoir qu'il y a 7,1 maillons de cette chaîne tous les cent mots français.

$$0{,}071\ (densité\ référentielle) = \frac{4366\ (\text{nombre de maillons})}{61862\ (\text{nombre de mots du texte})}$$

Sachant que la longueur moyenne des maillons ainsi que la densité référentielle sont deux indices assez variables selon les genres discursifs (*cf*. Schnedecker, 2021) en français, nous nous demandons s'ils varient en fonction des langues dans les textes d'un même genre discursif, autrement dit, si cette variation a lieu entre ceux du français et ceux du chinois. Nous essaierons de répondre à cette question dans la section *7.1. Chaîne du protagoniste 'Christophe'* Kelisiduofu *克利斯朵夫* du *Chapitre 7* par le truchement de la comparaison des chiffres issus du corpus en chinois.

Pour aller plus loin, nous nous intéressons aux corrélations entre la catégorie des maillons et la densité référentielle : le contexte d'une intense densité référentielle favoriserait-il une utilisation plus importante de maillons nominaux ou non nominaux ? La densité référentielle moyenne de 0,071 est prise comme seuil : les épisodes qui ont une densité référentielle supérieure à celui-ci sont dits épisodes *denses* (première colonne du tableau (6.4)). La proportion des maillons non nominaux de tout le corpus est de 86 % ; donc les épisodes qui présentent une proportion de maillons non nominaux supérieure à ce pourcentage sont dits épisodes de proportion *dense* (deuxième colonne du tableau (6.4)).

Tableau (6.4) : Corrélation entre la densité référentielle et la proportion des catégories non nominales

		Densité référentielle	Proportion des catégories non nominales
S I	1	0,030	79 %
	2	-	67 %
	3	0,072 (dense)	87 % (dense)
	4	0,025	71 %
	5	0,078 (dense)	89 % (dense)
	6	0,101 (dense)	99 % (dense)
	7	0,094 (dense)	95 % (dense)
	8	0,031	69 %
	9	0,052	84 %
	10	0,084 (dense)	89 % (dense)
S II	11	-	-
	12	0,089 (dense)	81 %
	13	0,115 (dense)	87 % (dense)
	14	0,114 (dense)	95 % (dense)
	15	0,096 (dense)	86 % (dense)
	16	0,081 (dense)	81 %
	17	0,131 (dense)	94 % (dense)
	18	0,075 (dense)	93 % (dense)
	19	0,101 (dense)	94 % (dense)
	20	0,084 (dense)	94 % (dense)
	21	0,083 (dense)	87 % (dense)
	22	0,064	90 % (dense)
	23	0,109 (dense)	92 % (dense)
	24	0,039	81 %
S III	25	0,074 (dense)	89 % (dense)
	26	0,073 (dense)	95 % (dense)
	27	0,047	79 %
	28	0,070	84 %
	29	0,073 (dense)	89 % (dense)
	30	0,087 (dense)	86 % (dense)
	31	0,034	89 % (dense)

	32	0,068	82 %
	33	0,074 (dense)	84 %
	34	0,045	82 %
	35	0,091 (dense)	88 % (dense)
	36	0,108 (dense)	91 % (dense)
S IV	37	0,097 (dense)	93 % (dense)
	38	0,054	65 %
	39	0,050	87 % (dense)
	40	0,068	90 % (dense)
	41	0	67 %
	42	0,079 (dense)	81 %
	43	0,041	86 % (dense)
	44	0,051	72 %
	45	0,084 (dense)	68 %
	46	0,102 (dense)	79 %
	47	0,080 (dense)	82 %
	48	0,087 (dense)	95 % (dense)
	49	0,089 (dense)	90 % (dense)
	50	0,107 (dense)	98 % (dense)
Total		31 épisodes *denses*	29 épisodes *denses*

Comparons les épisodes denses des deux colonnes. Il n'est pas difficile de constater qu'une grande partie de ces épisodes se correspondent. Par exemple, parmi les 31 épisodes dont la densité référentielle est supérieure à 0,071, 24 présentent des proportions denses de maillons non nominaux, coïncidence égale à 77 %. Réciproquement, parmi les 29 épisodes qui manifestent des proportions denses, 23 ont des densités référentielles supérieures à la moyenne de 0,071, coïncidence égale à 79 %. Par conséquent, il est fort probable que les longues successions de maillons non nominaux se répartissent plus souvent dans les épisodes à forte densité référentielle.

6.1.1.3. Titre, premier maillon et deuxième maillon

Sachant que Jean-Christophe est le personnage principal de l'œuvre, dont la chaîne comporte le plus grand nombre de maillons, nous nous posons la question de savoir comment il est introduit dans le récit : à quel moment et par quelle catégorie.

À première vue, le titre du livre comporte le nom propre d'un personnage nommé « Jean-Christophe ». Le résultat de l'annotation révèle que la chaîne de ce personnage est la plus longue puisqu'elle comprend 4366 maillons. À ce propos, il convient de se demander si ce nom propre constituant le titre du livre peut être considéré comme le premier maillon de cette chaîne ou s'il vaudrait mieux le traiter à part. En effet, ces deux points de vue sur le rôle du titre dépendent en quelque sorte du type de texte. Pour les textes journalistiques et scientifiques, Schnedecker (2021) propose des arguments pour deux traitements concurrents. D'un côté, les mentions présentes dans les titres peuvent être considérées comme premier maillon de chaîne en raison de la relation très forte entre le titre, le chapeau et le corps du texte dans ces deux types de textes. D'après l'auteure, cette relation forte peut être interprétée selon les quatre aspects suivants :

- Sémiotique : le titre joue un rôle crucial pour l'interprétation du texte ;

- Psycho-cognitif : le titre a un impact décisif sur la compréhension et mémorisation des textes ;

- Typo-dispositionnel : titre et texte sont dans un rapport de contiguïté spatio-textuelle étroit ;

- Sémantico-discursif : les relations sémantiques entre le titre et le texte sont indéniables ; on évoque souvent une relation d'élaboration entre le titre et le texte, le second développant ce qu'énonce le premier.

(Schnedecker, 2021 : 181)

D'autre part et à l'inverse, le titre peut être traité autrement pour des raisons typo-dispositionnelles en ce qui concerne les œuvres littéraires. À la différence des titres des textes de presse qui sont physiquement proches des corps de textes, le titre d'un roman est situé sur la couverture, distant du corps du texte de quelques pages dans les éditions imprimées. Il ne faudrait pas ignorer les effets de cet espace sur la cohésion textuelle, tout comme les autres formes typographiques dans le corps du texte, y compris les sous-titres, les numéros des sections et les paragraphes. Par conséquent, notre analyse du premier maillon ne concerne que celui qui figure dans le corps du texte : le nom propre dans le titre du livre n'est pas considéré par nous comme le premier maillon de la chaîne du protagoniste.

Selon notre point de vue, le titre du roman-fleuve ne constituant pas le premier maillon de la chaîne « Jean-Christophe », celui-ci, « le nouveau-né », apparaît, d'après notre annotation, à la place de sujet syntaxique au début de la première phrase dans le deuxième paragraphe du premier épisode du corpus, comme le montre l'extrait [1] ci-dessous[2]. Les maillons suivants de la même chaîne sont aussi annotés par des crochets.

[1] [**Jean-Christophe**]
titre

Le grondement du fleuve monte derrière la maison. La pluie bat les carreaux depuis le commencement du jour. Une buée d'eau ruisselle sur la vitre au coin fêlé. Le jour jaunâtre s'éteint. Il fait tiède et fade dans la chambre.
[**Le nouveau-né**] s'agite dans [son] berceau. Bien que le vieux ait laissé, pour entrer, ses sabots à
premier-maillon

la porte, son pas a fait craquer le plancher : [l'enfant] commence à geindre. La mère se penche hors de son lit, afin de [le] rassurer ; et le grand-père allume la lampe en tâtonnant, pour que [le petit] n'ait pas peur de la nuit. La flamme éclaire la figure rouge du vieux Jean-Michel, sa barbe blanche et rude, son air bourru et ses yeux vifs. Il vient près du berceau. Son manteau sent le mouillé ; il traîne en marchant ses gros chaussons bleus. Louisa lui fait signe de ne pas s'approcher. Elle est d'un blond presque blanc ; ses traits sont tirés ; sa douce figure mouton est marquée de taches de rousseur ; elle a des lèvres pâles et grosses, qui ne parviennent pas à se rejoindre et qui sourient avec timidité ; elle couve [l'enfant] des yeux – des yeux très bleus, très vagues, où la prunelle est un point tout petit, mais infiniment tendre.

« Jean-Christophe » est le premier personnage mentionné dans le livre. Cet extrait comporte sa mise en scène avant celle des autres protagonistes, comme « le vieux » et « la mère », dont l'introduction a lieu dans le même paragraphe après la sienne. Cette place primordiale du premier maillon de la chaîne « Jean-Christophe » confirme les observations des études qui, comme celles de Schnedecker & Longo (2012), de Schnedecker (1997, 2014, 2017, 2021) et de Manuélian & Schnedecker (2021), attestent également que les personnages qui

[2] La plupart des exemples et des extraits cités dans la *Partie III* de notre thèse sont issus de notre corpus *Jean-Christophe*. Pour éviter les notes redondantes, leur source « Rolland, 2007 [1912], traduit par Fu, 2017 [1941] » n'est pas mentionnée à chaque citation ; les exemples issus d'autres sources sont, eux, notés.

apparaissent en premier lieu dans un texte sont susceptibles d'être les personnages les plus importants dans la narration.

En ce qui concerne la catégorie du premier maillon d'une chaîne de référence, Charolles (1987) propose quatre types d'expressions référentielles susceptibles d'être utilisées pour démarrer une chaîne en français : *SN indéfini*, *nom propre*, *SN défini complet*[3] et *SN démonstratif*. Le premier maillon « le nouveau-né » de la chaîne « Jean-Christophe » est un *SN défini* incomplet, qui ne correspond à aucune catégorie de Charolles. Les quatre catégories sont ainsi testées dans les exemples [2]* – [5]* : elles remplacent « le nouveau-né » utilisé dans l'extrait [1] issu du roman-fleuve originaire :

- *SN indéfini*, comme « un nouveau-né », « un bébé », « un garçon », dans le test [2]* ;
- *nom propre*, comme « Jean-Christophe », « Christophe », dans le test [3]* ;
- *SN défini complet*, comme « le nouveau-né qui s'appelle Jean-Christophe », « le bébé chez les Krafft », dans le test [4]* ;
- *SN démonstratif*, comme « ce nouveau-né », « ce bébé », « ce garçon », dans le test [5]*.

Manifestement, toutes les catégories ne sont pas aussi acceptables les unes que les autres et chacune ne produit pas le même effet.

[2] ***Un nouveau-né** / **Un bébé** / **Un garçon** s'agite dans son berceau.
premier-maillon

[3] ***Jean-Christophe** / **Christophe** s'agite dans son berceau.
premier-maillon

[4] ***Le nouveau-né qui s'appelle Jean-Christophe** / **Le bébé chez les Krafft** s'agite dans son berceau.
premier-maillon

[5] ***Ce nouveau-né** / **Ce bébé** / **Ce garçon** s'agite dans son berceau.
premier-maillon

Selon Charolles, le SN indéfini est la catégorie le plus souvent utilisée au début d'un texte pour introduire un personnage qui n'est pas encore au centre de l'attention des lecteurs et qui n'est pas concerné dans les autres chaînes déjà constituées. Sans doute, le remplacement par un SN indéfini est acceptable : le protagoniste est introduit comme un personnage complètement nouveau. Par rapport à un SN indéfini, le narrateur a choisi un SN défini « le nouveau-né » : selon son point de vue, ce personnage n'est pas introduit comme complètement nouveau, puisque le seul personnage déjà mentionné dans le cadre fictif du roman-fleuve est repérable dans le titre « Jean-Christophe », malgré la grande distance typographique. Les lecteurs se rappellent cette mention présente dans le titre : le nom propre du titre du livre est censé référer au même personnage que celui de cette expression définie. Selon qu'une chaîne est initiée par un SN indéfini ou défini, les effets du premier maillon sont radicalement différents. D'après l'interprétation de Corblin (1995), l'indéfini et le défini permettent chacun de créer un contexte spécifique. Lorsque le premier maillon est un SN indéfini, le début du roman peut créer un monde fictionnel par le biais d'un personnage complètement nouveau (et aussi d'autres référents nouveaux), comme c'est le cas dans notre exemple avec « un nouveau-né ». Si tous les SN définis qui introduisent un nouveau référent

[3] Les SN définis complets peuvent produire des effets cognitifs et référentiels similaires à ceux des noms propres avec le recours aux connaissances encyclopédiques et aux mémoires de court / long terme de lecteurs. Au niveau syntaxique, ils peuvent souvent se substituer dans le même contexte discursif (*cf.* Roussarie, 2007 ; Charolles, 2002).

dans le texte sont remplacés par des SN indéfinis, le début du roman-fleuve sera identique à l'extrait [6]*, dans lequel est créé un monde inconnu avec des personnages et des objets complètement inédits et distanciés de ceux du monde réel. Certaines expressions manifestent des emplois agrammaticaux, comme « une pluie », « un commencement de jour » et « une nuit » ; d'autres, comme « un fleuve » et « un jour », ne permettent pas d'établir une relation référentielle avec ce qui est connu par les lecteurs dans le monde réel. Quant aux personnages, il est difficile de suggérer aux lecteurs d'établir un lien familial entre « un nouveau-né », « un vieux » et « une mère ». Le remplacement mécanique de tous les déterminants par les indéfinis provoque donc des agrammaticalités.

[6] ***Un** (*indéfini*) grondement de fleuve monte derrière **une** (*indéfini*) maison. **Une** (*indéfini-agrammatical*) pluie bat **des** carreaux depuis **un** (*indéfini-agrammatical*) commencement de **jour** (*indéfini*). Une buée d'eau ruisselle sur **une** (*indéfini*) vitre à **un** (*indéfini*) coin fêlé. Le jour jaunâtre s'éteint. Il fait tiède et fade dans **une** (*indéfini*) chambre.

[**Un** (*premier-maillon*) nouveau-né] s'agite dans [son] berceau. Bien qu'**un** (*indéfini*) vieux ait laissé, pour entrer, ses sabots à **une** porte, son pas a fait craquer **un** plancher : [l'enfant] commence à geindre. **Une** (*indéfini*) mère se penche hors d'**un** (*indéfini*) lit, afin de [le] rassurer ; et le grand-père allume **une** (*indéfini*) lampe en tâtonnant, pour que [le petit] n'ait pas peur d'**une** (*indéfini-agrammatical*) nuit. […]

Au contraire, lorsque les premiers maillons sont des SN définis, le début du roman-fleuve peut plonger les lecteurs *in medias res*, dans le monde fictif, ce à quoi contribue le SN défini, qui indique que le référent est présenté comme déjà connu par les lecteurs, d'où l'usage de « le nouveau-né » dans *Jean-Christophe*.

> Le N *s'interprète comme signal qu'on se situe dans un univers de discours déjà connu, où il y a un seul individu passible du signalement. Pour cette raison, le défini est une forme qui associe un individu à l'ensemble d'un univers discursif déjà posé, une forme qui utilise un acquis discursif étendu, contenant plusieurs individus déjà déterminés ; elle signale qu'on « reste » dans l'univers constitué. On peut comprendre par là que le défini soit un facteur de cohésion global d'une séquence discursive, faisant appel à un ensemble de mentions antérieures. (Corblin, 1995 : 68)*

Le narrateur a ainsi cherché à construire un monde fictif, peuplé de personnages et d'entités censés être « déjà connus », ce par l'emploi des SN définis, comme « le vieux », « la mère » pour les humains, et « le grondement du fleuve », « la maison » et « les carreaux », pour les entités non humaines, entre certaines desquelles les SN définis impliquent en même temps des relations d'anaphore associative.

La deuxième catégorie possible du premier maillon est le *nom propre* selon Charolles. À la différence du SN défini, un nom propre d'humain à la place de premier maillon signifie, comme Schnedecker (2021 : 189) le signale, que « le référent visé est censé être peu accessible » aux lecteurs. Selon le point de vue du narrateur, cette explication ne s'adapte pas au cas de *Jean-Christophe*, puisque ce personnage est déjà mentionné dans le titre du livre. La répétition du même nom propre au début du texte du test [3]* risquerait de laisser entendre que l'identité du personnage est redondante et inutilement réitérée.

La troisième catégorie possible du premier maillon est les *SN définis complets*, tandis que le premier maillon « le nouveau-né » apparaît sous forme dite incomplète. Tout ce que les

lecteurs peuvent présager de ce personnage se limite à l'information fournie dans le titre selon laquelle ce nouveau-né est susceptible de s'appeler « Jean-Christophe ». Ainsi, comme le montre le test [4]*, il est grammaticalement acceptable de compléter le SN défini « le nouveau-né » par une proposition subordonnée relative « qui s'appelle Jean-Christophe ».

Pour terminer les tests effectués dans le cadre de la typologie de Charolles, examinons la dernière catégorie potentielle du premier maillon, qui est *SN démonstratif* (*cf.* aussi Corblin, 1995). Les expressions démonstratives, comme « ce nouveau-né », « ce bébé » et « ce garçon » du test [5]*, paraissent encore moins pertinentes par rapport aux trois autres catégories. En effet, afin de référer à un personnage, les démonstratifs peuvent accompagner aussi bien une relation déictique qu'une relation anaphorique (*cf.* Corblin, 1987 ; Kleiber, 1996 ; De Mulder, 1998 ; Gary-Prieur & Léonard, 1998 ; Cornish, 2017 ; voir aussi leur définition dans le *Chapitre 2*), alors que le contexte extra-linguistique de la lecture d'une œuvre littéraire ne fournit pas d'indices permettant aux lecteurs de repérer un personnage présent. Nous chercherons à savoir s'il existe et comment ces deux fonctions de démonstratifs sont exploitées dans les extraits du corpus.

Le deuxième maillon est également considéré comme une place importante qui permet à la fois d'établir une relation coréférentielle avec le premier maillon et de maintenir la saillance du référent pour les maillons suivants. La forme du deuxième maillon peut servir à confirmer la saillance du référent (Salles, 2015). Dans la proposition de Charolles (1987) sur la configuration d'une chaîne de référence en français, les formes les plus susceptibles de se situer à la deuxième place sont au nombre de sept. Selon les configurations prototypiques, c'est un SN indéfini qui occupe la place du premier maillon. Nous illustrons ainsi les configurations des deux premiers maillons par cette adaptation avec le premier maillon « un nouveau-né ». Les sept formes du deuxième maillon permettent théoriquement de constituer une relation anaphorique ou coréférentielle avec leur antécédent : *nom propre*, *SN défini complet*, *SN défini incomplet*, *SN démonstratif*, *nom de qualité*, *SN indéfini* et *pronom*. Comme le montre l'extrait [1], le deuxième maillon de la chaîne « Jean-Christophe » est un déterminant possessif, « son (berceau) ». En effet, les déterminants possessifs sont considérés comme faisant partie des marqueurs de très haut degré d'accessibilité dans l'échelle d'accessibilité d'Ariel (1990). Cette huitième possibilité d'un déterminant possessif à la deuxième place de la chaîne permet de constituer une relation anaphorique et coréférentielle encore plus forte que les autres formes.

6.1.1.4. Redémarrage de la chaîne de référence après une longue distance

La chaîne « Jean-Christophe » comporte au total 4366 maillons en français, couvrant les quatre sections et 49 épisodes sur 50 des extraits du livre. Or, nous avons repéré deux endroits où ne se trouve aucun maillon de cette chaîne ; par conséquent, les deux ruptures de la chaîne donnent ensuite lieu à deux redémarrages de celle-ci.

La première rupture se trouve à l'épisode 2, qui ne contient que 3 maillons à sa dernière phrase. Autrement dit, entre le dernier maillon de l'épisode 1 et le premier maillon de l'épisode 2 se situe une distance qui couvre presque l'intégralité de l'épisode 2. Comme le montre l'extrait [7], la chaîne est redémarrée par un nom propre modifié « le petit Jean-Christophe »

après une narration relative aux autres personnages, qu'une longue distance de 883 mots sépare de l'antécédent.

[7] À peine marié, il se montra atterré de ce qu'il avait fait. Il ne le cacha point à la pauvre Louisa, qui, tout humble, lui en demandait pardon. Il n'était pas méchant, et le lui accordait volontiers ; mais, l'instant d'après, ses remords le reprenaient, au milieu de ses amis, ou chez ses riches élèves, maintenant dédaigneuses, qui ne tressaillaient plus au frôlement de sa main, quand il voulait rectifier la pose de leurs doigts sur le clavier. Il revenait alors avec une mine sombre, où Louisa, le cœur serré, lisait du premier coup d'œil les habituels reproches ; ou bien il s'attardait dans des stations au cabaret ; il y puisait le contentement de soi et l'indulgence pour autrui. Ces soirs-là il rentrait avec des éclats de rire, qui semblaient plus tristes à Louisa que les sous-entendus et la sourde rancune des autres jours. Elle se sentait un peu responsable des accès de déraison, où disparaissaient à chaque fois, avec l'argent de la maison, les faibles restes du bon sens de son mari. Melchior s'enlisait. À un âge où il aurait dû travailler sans répit à développer son médiocre talent, il se laissait glisser le long de la pente ; et d'autres prenaient sa place.
Mais qu'importait sans doute à la force inconnue qui l'avait rapproché de la servante aux cheveux de lin ? Il avait rempli son rôle ; et [**le petit Jean-Christophe**] venait de prendre pied sur cette terre,
redémarrage

où le poussait son destin.

La seconde rupture se trouve à l'épisode 11, le seul épisode qui ne comporte aucun maillon de la chaîne « Jean-Christophe ». L'absence de maillon dans cet épisode constitue en effet la distance la plus longue entre deux maillons de la chaîne. L'épisode 11 comporte 2117 mots français (voir le tableau (5.7) du *Chapitre 5*), donc la distance entre le dernier maillon de l'épisode 10, « le », et le premier maillon de l'épisode 12, « le petit Christophe », est plus que 2117 mots, sans compter le nombre de mots après « le » de l'épisode 10 et avant « le petit Christophe » de l'épisode 12. Le début de l'épisode 12 contenant le maillon de redémarrage est présenté dans l'extrait [8].

[8] Ce fut au moment où la situation de la famille devenait le plus difficile, que [**le petit Christophe**]
redémarrage-1

commença à comprendre ce qui se passait autour de [lui].
mention-2

[Il] n'était plus seul enfant. Melchior faisait un enfant à sa femme chaque année, sans s'inquiéter de
mention-3

ce qui en arriverait plus tard. Deux étaient morts en bas âge. Deux autres avaient trois et quatre ans. Melchior ne s'en occupait jamais. Louisa, forcée de sortir, les confiait à [Christophe, qui avait
mention-4

maintenant six ans].

Ce maillon de redémarrage se situe à la fois au début d'un épisode, selon notre découpage, et au début d'une section, selon la partition du romancier. Ces sections n'ont pas de sous-titre ; elles ne sont marquées que par des chiffres romains. Ces numéros de sections permettent donc aux lecteurs d'être avertis des nouvelles intrigues du roman-fleuve mais sans connaître leur sujet. Un épisode sans maillon de « Jean-Christophe » peut affaiblir encore la mémoire établie par la section I sur le personnage. Par rapport au premier maillon de la chaîne, « le nouveau-né », situé au début du roman-fleuve, le maillon de redémarrage « le petit Christophe » de l'extrait [8] semble mieux correspondre aux configurations de Charolles pour le début de la chaîne de référence : *nom propre* ou *SN défini complet*. Gary-Prieur & Noailly (2019) définissent les expressions du type « le petit Christophe » comme *nom propre modifié*, en configuration de *déterminant – adjectif – nom propre* (2019 : 26).

> *L'adjectif épithète antéposé apporte soit une caractéristique permanente du référent, soit le point de vue du locuteur sur l'individu concerné. […] Dans tous les cas, la propriété dénotée par l'adjectif antéposé est supposée préalablement admise par les deux participants. (Gary-Prieur & Noailly, 2019 : 26 – 27)*

L'usage de ce nom propre modifié correspond bien à ces interprétations de Gary-Prieur & Noailly : la caractéristique physique « petit » n'est pas permanente au personnage puisqu'il doit être considéré comme un référent évolutif qui grandit au fils du temps et qui portera d'autres propriétés dans la suite du roman-fleuve. Cet adjectif exprime plutôt le point de vue du romancier puisque le personnage apparaît « petit » à ce moment-là de la narration.

6.1.1.5. Répartition des maillons

Malgré la place dominante de la chaîne « Jean-Christophe » dans l'intégralité du roman-fleuve, la répartition des maillons présente une grande variation selon les différentes sections et les différents épisodes. Par exemple, l'épisode 35 comporte 246 maillons (il s'agit de l'épisode ayant le nombre le plus important de maillons), alors que les épisodes 2 et 41 n'en contiennent que 3, ce qui permet à peine de constituer une chaîne. La chaîne du protagoniste peut même être absente à des moments de la narration, comme nous l'avons vu au sujet de l'épisode 11. La question se pose donc de savoir à quel moment est utilisée telle ou telle catégorie de maillon pour maintenir la chaîne du personnage, pour marquer des évolutions ontologiques et pour redémarrer les intrigues le concernant. Cette répartition non équilibrée de maillons du protagoniste indiquerait le référent central du roman-fleuve aux niveaux global et local : le protagoniste peut être le personnage central pour la globalité du roman-fleuve, mais d'autres personnages peuvent dépasser son importance dans des sections et des épisodes de ce genre de long texte narratif : l'alternance du centrage aurait lieu entre les référents selon la répartition de maillons.

Afin d'apporter des éléments de réponse, nous avons listé tous les maillons de la chaîne « Jean-Christophe » dans l'*Annexe 4 : Maillons français dans la chaîne « Jean-Christophe »*, selon leur ordre d'apparition dans chaque section et dans chaque épisode. Un extrait des premières lignes de l'*Annexe 4* est montré ci-dessous dans le tableau (6.5).

Tableau (6.5) : Maillons de la chaîne « Jean-Christophe » et leur répartition dans les sections et les épisodes en français (début de l'*Annexe 4*)

		JC	Maillons
S I	1	48	Le nouveau-né – son – l'enfant – le – [le petit – l'enfant – L'enfant] – Ø – Son – l – lui – le – lui – il – il – le – Sa – sa – ses – il – il – cela – L'enfant – le – Il – il – sa – l – le – le petit – d'aussi laid – l'enfant – le – le – mon pauvre petit – tu – tu – t – lui – l – l'enfant – il – L'enfant – Il – son – il – [du petit – l'enfant]
	2	3	le petit Jean-Christophe – le – son
	3	38	l'enfant – son – Il – Il – son – il – il – Il – lui – il – Sa – le – il – il – lui – lui – l – son – lui – L'enfant – Sa – son – lui – il – son – son – sa – le – [mon jésus – mon petit poisson d'or] – Il – sa – cette misérable masse inconsciente et informe – lui – l

Dans les colonnes à gauche du tableau se trouvent les numéros des sections et des épisodes du texte, suivis du nombre de maillons de la chaîne « Jean-Christophe » dans chaque épisode ; dans la colonne de droite sont listés tous les maillons du corpus. Pour mieux montrer la répartition des maillons selon leur catégorie morpho-syntaxique, les maillons nominaux sont marqués par un fond de couleur :

Catégories nominales	SN sans déterminant	Nom propre	SN défini	SN possessif	SN démonstratif	SN indéfini
Catégories non nominales	Déterminants possessifs		Zéro	Déictiques		Autres
Autres	Suites des maillons nominaux		Maillons nominaux modifiés			

Les pronoms personnels de la troisième personne ne sont marqués par aucune couleur. Ils constituent le plus grand pourcentage de la chaîne, ce qui correspond également aux résultats de beaucoup d'autres études, comme le résument Masseron & Schnedecker (1995) :

> *En effet, à partir de nombreuses analyses opérées sur des corpus très éclectiques (Ariel, 1990 ; Combettes, 1986 ; De Weck, 1991 ; Schneuwly, 1988), il est désormais bien admis que le pronom est la forme la plus couramment usitée dans les chaînes – dans des rapports qui vont du tiers à la moitié des expressions référentielles totale. Dans cette optique, le pronom serait la forme « non marquée » de la reprise. (Masseron & Schnedecker, 1995 : 424)*

Sur la base de l'observation des maillons exposés dans l'*Annexe 4*, nous cherchons à synthétiser dans les parties suivantes les moments de l'apparition de chaque maillon nominal et non nominal de la chaîne « Jean-Christophe », afin de faire ressortir quelques interprétations et régularités de leur emploi.

Comme nous l'avons montré dans le tableau (6.3), les maillons nominaux occupent au total 16 % des maillons dans la chaîne « Jean-Christophe ». Selon Capin *et al.* (2021), ces maillons nominaux, y compris les SN et les noms propres, au milieu du texte, peuvent signifier trois situations en français : i) éviter des risques d'ambiguïté référentielle lorsqu'il y a la co-présence de personnages de même genre ; ii) coïncider avec d'autres marques de structuration textuelle et iii) faire office de marque de structuration textuelle. Outre ces trois situations, nous constatons également que ces maillons nominaux peuvent être utilisés de manière successive dans certains endroits du roman-fleuve. Il nous paraît donc intéressant d'observer s'il existe d'autres contraintes spécifiques qui conditionnent l'apparition de ces suites qui sont beaucoup plus rares et beaucoup plus courtes que celles des maillons non nominaux. Nous parlons donc de « suites des maillons nominaux » pour définir deux ou plus de deux maillons nominaux utilisés successivement dans les chaînes de référence. Au total, 96 suites de maillons nominaux sont obtenues dans la chaîne du protagoniste en français, contenant entre 2 et 9 maillons nominaux successifs. Les suites de 2 maillons concernent plus de deux tiers des cas, soit 69 % (67 / 96). Le nombre de suites diminue en fonction de l'augmentation de la longueur. Un bilan est ainsi établi :

- 67 suites de 2 maillons nominaux ;
- 18 suites de 3 maillons nominaux ;
- 7 suites de 4 maillons nominaux ; elles se situent dans les épisodes 4, 8, 13, 34, 35 (avec deux suites) et 46 ;
- 3 suites de 5 ou 6 maillons nominaux ; elles se situent dans les épisodes 8, 14 et 38 ;
- une suite de 9 maillons nominaux ; elle se situe dans l'épisode 47[4].

[4] Les 13 épisodes restants (2, 5, 6, 7, 11, 17, 24, 25, 26, 31, 39, 41 et 48) ne contiennent aucune suite de maillons nominaux.

Notre hypothèse porte sur deux indices qui pourraient déterminer l'usage des suites d'expressions nominales : la distance inter-maillonnaire et le nombre de référents dans le contexte. En accord avec ce qu'observe Ariel, nous avons remarqué dans notre corpus que la distance inter-maillonnaire peut être plus longue que celle de la moyenne ou celle entre deux maillons non nominaux. Par exemple, l'extrait [9] suivant contient une suite de 4 maillons nominaux (de « M-nominal-1 » à « M-nominal-4 »), en compétition avec deux autres référents (« baron » et « un employé »), et une suite de 4 maillons non nominaux (de « M-non-nominal-1 » à « M-non-nominal-4 »). Tous les maillons nominaux sont situés dans les différentes phrases, avec une distance moyenne de 26 mots (104 / 4). Les expressions coréférentielles de la même phrase sont en forme non nominale, avec une distance de 3,5 mots (14 / 4).

[9] Tout le monde le connaissait aux bureaux. Il demanda à parler à Son Excellence l'intendant des théâtres, baron de Hammer Langbach. Un employé, jeune, gras, chauve, le teint fleuri, avec un gilet blanc et une cravate rose, lui serra familièrement la main, et se mit à parler de l'opéra de la veille. [**Christophe**] répéta sa question. L'employé répondit que Son Excellence était occupée en ce
M-nominal-1

moment, mais que, si [**Christophe**] avait une requête à lui présenter, on la lui ferait passer avec
M-nominal-2

d'autres pièces, qu'on allait lui porter à signer. [**Christophe**] tendit la lettre. L'employé y jeta les
M-nominal-3

yeux, et poussa une exclamation de surprise :
– Ah ! par exemple ! fit-il gaiement. Voilà une bonne idée ! Il y a longtemps qu'il aurait dû s'aviser de cela ! De toute sa vie, il n'a rien fait de mieux. Ah ! le vieux pochard ! Comment diable a-t-il pu s'y résoudre ?
Il s'arrêta net. [**Christophe**] lui avait arraché le papier des mains, et [**Ø**] criait, blême de colère :
M-nominal-4 *M-non-nominal-1*

– [**Je**] vous défends !… [**Je**] vous défends de [**m'**] insulter !
M-non-nominal-2 *M-non-nominal-3* *M-non-nominal-4*

Nous avons mesuré la distance inter-maillonnaire moyenne en mesure du nombre de maillons des suites. Le bilan des distances moyennes entre deux maillons est affiché dans le tableau (6.6). Au contraire de notre hypothèse, la distance (première colonne du tableau) ne démontre pas de corrélation rapportée à la longueur des suites. La plus grande distance concerne les suites de 4 maillons, à savoir 37,1 mots entre deux maillons, alors que la plus petite distance est constatée dans l'intégralité des maillons nominaux et non nominaux du corpus, à savoir 14,2 entre deux maillons. Les détails de chaque suite de maillons nominaux sont donnés dans le tableau situé à la fin de cette partie.

En ce qui concerne le nombre de référents co-présents, les résultats (deuxième colonne du tableau) ne révèlent aucune corrélation significative. Le nombre de référents n'augmente ni ne diminue en fonction de la longueur des suites des maillons nominaux. Les référents co-présents se situent entre 0,7 (pour les suites de 2 maillons) et 3,1 (pour la suite de 9 maillons). Notre hypothèse relative aux deux corrélations potentielles semble donc échouer à expliquer les raisons de l'apparition des suites de maillons nominaux.

Tableau (6.6) : Caractéristiques des suites de maillons nominaux dans la chaîne « Jean-Christophe »

Nbre de	Mots (distance moyenne)	Réf
Dans les 67 suites de 2 maillons	7,9	0,7
Dans les 18 suites de 3 maillons	24,4	1,3
Dans les 7 suites de 4 maillons	37,1	1,5
Dans les 3 suites de 5 ou 6 maillons	28,0	0,9
Dans la suite de 9 maillons	19,2	3,1
De tous les maillons dans le corpus	11,7	1,5

Etablissons un bilan provisoire. Cette section consacrée à la présentation générale de la chaîne de référence du protagoniste dans le roman-fleuve français *Jean-Christophe* permet d'en avoir une vision globale. La plupart des observations sont orientées par les résultats des études antérieures, alors que les chiffres que nous avons obtenus ne confirment pas toutes nos hypothèses. D'un côté, celles sur la corrélation entre des paramètres se sont révélées fausses, comme celle entre la longueur des suites de maillons nominaux, la distance inter-maillonnaire et le nombre de référents ; pourtant, d'un autre côté, les chiffres peuvent parfois être aléatoires. Par exemple, nous n'avons obtenu qu'une seule suite de 9 maillons nominaux, dont les caractéristiques risquent d'être insuffisamment typiques et représentatives pour toutes les suites de maillons nominaux. Dans les sections suivantes, nous préciserons davantage la performance de chaque groupe de maillons dans la chaîne du protagoniste.

6.1.2. Divers maillons nominaux

Après la présentation générale des maillons nominaux, nous abordons dans cette partie l'usage des différentes catégories nominales des maillons dans la chaîne « Jean-Christophe » en français selon le déterminant. Les maillons seront analysés par ordre d'accessibilité décroissante dans l'échelle d'Ariel (1990) par les six catégories suivantes : nom propre, SN défini, SN possessif, SN démonstratif, SN indéfini et SN sans déterminant. De nombreux auteurs, notamment Reichler-Béguelin (1988), Galmiche (1989), Corblin (1995) et Schnedecker (1992, 1997, 2021), cherchent à synthétiser les contextes discursifs spécifiques et les caractéristiques des différentes expressions référentielles nominales. À partir de leurs remarques et conclusions, nos hypothèses portent sur cinq facteurs qui peuvent déterminer la catégorie grammaticale du maillon, et, réciproquement, quelle forme peut être utilisée sous quelles conditions discursives. Pour obtenir d'éventuelles conclusions quantitatives, les six groupes de maillons nominaux dans la chaîne « Jean-Christophe » sont respectivement interrogés selon les cinq critères suivants :

- Critère I : *fonction*

Il s'agit de la fonction syntaxique des maillons dans le cadre phrastique. Selon notre annotation, les fonctions des maillons peuvent être *sujet*, *complément d'objet*, *complément du nom*, *circonstant* et *autres fonctions*. Par exemple, dans l'extrait [10], « Christophe » (« M-1 ») est *sujet* ; « Christophe » (« M-3 ») est *complément d'objet*.

[10][**Christophe**]$_{M-1}$, machinalement, donna le chiffon de lettre ; mais [**il**]$_{M-2}$ se lança dans un flot de paroles embrouillées, tendant toujours la main pour ravoir la lettre. L'Excellence déplia soigneusement le papier, le lut, regarda [**Christophe**]$_{M-3}$, [**le**]$_{M-4}$ laissa patauger dans [**ses**]$_{M-5}$ explications, puis [**l'**]$_{M-6}$ interrompit, et dit, avec un éclair malicieux dans les yeux :

– C'est bien, [**monsieur Krafft**]$_{M-7}$. La demande est accordée.

De la main, il [**lui**]$_{M-8}$ donna congé et se replongea dans ses écritures.

[**Christophe**]$_{M-9}$ sortit, consterné.

– Sans rancune, [**Christophe**]$_{M-10}$! [**lui**]$_{M-11}$ dit cordialement l'employé, quand [**l'enfant**]$_{M-12}$ repassa par le bureau. [**Christophe**]$_{M-13}$ se laissa prendre et secouer la main, sans oser lever les yeux.

[**Il**]$_{M-14}$ se retrouva hors du château. [**Il**]$_{M-15}$ était glacé de honte.

Nous nous intéressons à cet indice parce que, du point de vue des approches cognitives, la position phrastique reflète la saillance du référent : les référents exprimés en position de sujet sont souvent plus saillants que ceux occupant d'autres positions (*cf*. Schmid, 2007 : 128). Les maillons du protagoniste sont donc supposés se présenter davantage au sujet qu'aux autres positions.

- Critère II : *distance*

Ce critère renvoie à la distance entre un maillon et celui qui le précède calculé nombre de phrases dans le contexte à gauche (*cf*. Givón, 1983 : 13). Par exemple, dans l'extrait [10], la distance entre « Christophe » (« M-3 ») et son antécédent « il » (« M-2 ») est marquée comme « 1 (phrase) » puisque l'antécédent se situe dans la phrase précédente ; celle entre « le » (« M-4 ») et son antécédent « Christophe » (« M-3 ») est marquée comme « 0 (phrase) » puisque les deux maillons se situent dans la même phrase.

Selon plusieurs auteurs, comme Givón (1983) et Ariel (1990), le nombre de phrases[5] est choisi comme mesure pour décrire la distance ; d'autres mesures existent, comme le nombre de syllabes, de mots, de lignes et de paragraphes (*cf*. Schnedecker, 1992, 1997). Nous choisissons le nombre de phrases pour deux raisons. D'une part, les mesures de petites unités comme les syllabes et les mots peuvent apporter des différences internes suscitées par les deux systèmes linguistiques, à savoir qu'une phrase n'a pas la même longueur lorsqu'elle est exprimée en français et en chinois. Par exemple, la phrase française [11], issue de notre corpus, compte 12 syllabes, 9 mots et 46 caractères alphabétiques ; la phrase parallèle chinoise [12] compte 12 syllabes, 6 mots et 26 caractères alphabétiques[6]. Les phrases parallèles dans les deux langues n'ont pas la même longueur lorsque la mesure choisie est différente ; mais le nombre de phrases est identique.

[5] La notion de *phrase* que nous utilisons concerne non seulement celles qui sont marquées par le point final « . », le point d'interrogation « ? », le point d'exclamation « ! », mais également par le point-virgule « ; ».

[6] Ici, les caractères sont les unités informatiques, qui n'ont rien à voir avec les caractères chinois. Un caractère chinois, ainsi que la ponctuation chinoise, équivalent à deux caractères informatiques. Le nombre de caractères alphabétiques peut refléter le nombre de lignes : une ligne française et une ligne chinoise ne contiennent pas le même nombre de syllabes ou de mots.

[11]La suite des jours est comme un même jour.
[12]循环不已的 日月 仍 好似 一 日。
xunhuanbuyide riyue reng haosi yi ri
circulant jour-nuit même comme un jour

D'autre part, les paragraphes nous semblent trop longs comme unités de mesure. Le risque existe qu'une grande partie des distances inter-maillonnaires soient 0 ou 1 ; le résultat ne permettrait donc pas de montrer correctement les différences.

Selon le critère de continuité de Givón (1983 : 14), la plus petite distance est mesurée comme « 1 (phrase) », représentant la continuité maximale du topique, et la plus grande distance est définie arbitrairement par l'auteur comme « 20 (phrases) », impliquant la discontinuité du topique. Chaque distance est marquée par une forme grammaticale, dite *dispositif d'encodage* (*coding device*) :

> *- Au-delà de 20 phrases, on atteint le seuil de la discontinuité : un topique non réinstancié dans l'intervalle sera considéré comme désactivé. […]*
>
> *- 15 phrases : description définie complète ;*
>
> *- 10 phrases : description définie ;*
>
> *- 2 – 3 phrases : pronom tonique ;*
>
> *- 1 – 2 phrases : pronom clitique ;*
>
> *- 0 phrase : anaphore zéro.*
>
> *(Givón, 1983, cité par Schnedecker, 1992 : 204)*

L'observation de quelques annotations au début du corpus nous permet de prévoir que le nombre de la plupart des distances inter-maillonnaires peut se situer éventuellement entre 1 et 9 phrases pour qu'un topique garde sa continuité. Nous reprenons donc le même critère de Givón pour notre étude : la distance est considérée comme *longue* si le nombre de phrases dépasse 10 et comme *très longue* s'il dépasse 20.

- Critère III : *référents* (« réf »)

Il s'agit du nombre de référents intervenus entre un maillon et son antécédent. Par exemple, entre les deux premiers maillons de l'extrait [10] se trouve un seul référent « le chiffon de lettre » ; donc le nombre de référent est 1[7].

Dans le cadre de l'analyse textuelle de Schnedecker (1992, 1997), deux types de contexte existent relativement au nombre de référents : situation *multi-référentielle* et *mono-référentielle* (voir leur définition dans la section *3.2.3.2. Séquences mono-référentielles et multi-référentielles* du *Chapitre 3*). Dans le contexte multi-référentiel, les référents sont subdivisés en *leaders*, *outsiders* et *challengers*, avec les définitions suivantes[8] :

> *Leader : potentiellement, un référent doté d'un nom propre, également sujet et agent ayant déjà fait l'objet de mentions préalables dans le contexte ;*

[7] Excepté le recensement des nombres exacts de référents, les nombres ne sont pas recensés dans les distances dites « longue » et « très longue » : dans ces cas-là, les nombres de référents sont marqués directement comme « beaucoup ».

[8] *Cf.* aussi Schnedecker (1992 : 201).

> *Outsider : un référent non animé en position de complément adnominal ou secondaire, apparaissant en première mention dans la situation et présenté, en la circonstance, dans cette première mention comme étant flambant neuf.*
>
> *Partant de là, le bon* challenger, *à même de soutenir la compétition avec un référent* leader *sera celui des référents d'une situation multi-référentielle qui réunit l'ensemble des caractéristiques suivantes : être dénommé par un nom propre, être sujet syntaxique et agent sémantique, avoir été préalablement activé dans le cotexte.*
>
> *(Schnedecker, 1997 : 68)*

Selon cette typologie des référents, le personnage Jean-Christophe est un *leader*, puisque les maillons correspondent à tous les critères susmentionnés : sémantiquement, ils renvoient à un référent animé ; formellement, ce personnage est doté d'une désignation par un nom propre ; syntaxiquement, beaucoup de maillons sont sujets de phrase ; discursivement, ce personnage fait déjà l'objet de mentions préalables. Les autres expressions référentielles renvoient donc à des référents *outsiders* et *challengers*[9]. Leurs nombres décident si la situation contextuelle est *multi-référentielle* (lorsque ils sont plusieurs) ou *mono-référentielle* (lorsque leurs nombres sont zéro) et peuvent ainsi influencer la forme des maillons du *leader*.

- Critère IV : *début*

Ce critère signifie que l'expression se situe au début d'une séquence textuelle : paragraphe, épisode et section du roman-fleuve. Ce sont des endroits importants dans le déroulement narratif pour quatre raisons : ils marquent, au premier temps, les changements typo-graphiques ; dans un deuxième temps, des changements spatio-temporels ; dans untroisième temps, des changements de point de vue, et dans un quatrième temps, la saillance du référent, puisque les expressions en première position correspondent à des référents plus saillants que ceux en positions ultérieures (*cf*. Reichler-Béguelin, 1997 : 102). Les référents installés à la première place servent de fondement et d'autres référents s'ajoutent au fur et à mesure de l'avancement de la structure (Gernsbacher, 1989 : 143, citée par Schnedecker, 1997 : 76).

Pour ces raisons-là, nous choisissons d'observer si les maillons se situent aux endroits dits importants ou non. Quatre positions sont donc examinées : i) début de paragraphe : dans la première phrase d'un paragraphe ; ii) début d'épisode : dans la première phrase d'un épisode ; iii) début de section : dans la première phrase d'une section et iv) non début : l'expression n'est pas dans la première phrase dans aucun type de séquence textuelle. Comme le montre l'extrait [10], les maillons numérotés 1, 8, 9 et 14 sont considérés comme en *début* de paragraphe ; les autres maillons ne sont pas au *début* dans aucune séquence textuelle.

- Critère V : *discours*

Ce dernier critère a pour objectif de décrire la position d'une expression par rapport au discours rapporté direct et au discours épistolaire. Selon ce critère, nous définissons deux positions. La première concerne les maillons qui se situent dans les séquences dialogales et épistolaires relatives au changement des tours de parole. Ces séquences contiennent trois

[9] Une notion similaire est proposée par Ariel (1990 : 28) : chaînes de *compétiteurs* locaux ou épisodiques, dont les référents sont en compétition de saillance avec celle du protagoniste, référent topical du roman-fleuve.

parties : i) la dernière phrase qui précède les tours de parole, ii) les tours de parole et iii) la première phrase après les tours de parole. Par exemple, une séquence dialogale est présentée dans l'extrait [10]. Les maillons numérotés 3 – 12 sont considérés comme en position de *discours* (appellation abrégée de « discours rapporté direct »). C'est une position en relation directe avec les tours de parole des personnages. Les changements de prise de parole peuvent impliquer en même temps des changements de point de vue (*cf*. Achard-Bayle, 2000). La seconde position concerne les maillons dans le reste du texte, comme les maillons numérotés 1 – 2 et 13 – 15. L'ensemble des séquences dialogales totalisent 4939 mots, équivalant à 8 % du corpus français[10]. Ce pourcentage nous permet de savoir si les séquences dialogales contiennent les formes d'expressions de proportions similaires par rapport à la moyenne de tous les textes du corpus.

En bref, les caractéristiques discursives des maillons dans les chaînes de référence peuvent être reflétées en quelque sorte par ces cinq critères ; nous exposerons les résultats quantitatifs en les illustrant par des exemples pour les six groupes de maillons nominaux.

6.1.2.1. Noms propres : répartition quantitative et tactique

Le nom propre est la catégorie nominale le plus souvent utilisée dans cette chaîne. Le nombre de noms propres atteint 409[11]. En raison de son rôle, cette catégorie a reçu des interprétations variables selon différentes approches. Nous présenterons d'abord les cadres théoriques où s'inscrivent les noms propres, et ensuite nous utiliserons ces théories pour expliquer les occurrences régulières et exceptionnelles des noms propres dans la chaîne « Jean-Christophe ».

6.1.2.1.1. Coïncidences avec les changements textuels

Par le biais de l'approche sémiologique, Hamon (1972) résume l'usage des noms propres désignant le protagoniste dans les œuvres littéraires comme *une distribution différentielle*, concernant « un mode d'accentuation purement quantitatif et tactique ». Elle comporte deux facettes :

> *Apparition fréquente ;*
>
> *Apparition aux moments marqués du récit (début / fin des séquences et du récit).*
>
> *(Hamon, 1972 : 91)*

L'apparition des noms propres est dite *quantitative* selon Hamon puisque leur occurrence n'est pas censée être unique ou épisodique. Cette remarque est confirmée par notre recensement, selon lequel les noms propres s'élèvent au total à 409, concernant 11 %

[10] Voir toutes les séquences dialogales dans l'*Annexe 6 : Séquences dialogales et épistolaire en français.*

[11] Les détails des noms propres et les résultats quantitatifs sont présentés dans l'*Annexe 7 : Caractéristiques discursives des noms propres français dans la chaîne « Jean-Christophe.*

de maillons dans la chaîne « Jean-Christophe », ce qui en constitue la catégorie nominale la plus importante.

L'apparition des noms propres est dite *tactique* puisque leurs occurrences au début ou à la fin des séquences textuelles peuvent marquer le déroulement narratif. Schnedecker (1997) signale que ce dernier peut être reflété par l'usage des noms propres au moment :

- des changements de points de vue sur le référent ;
- des changements de paragraphes, d'épisodes, des changements typographiques ;
- des changements de cadres spatio-temporels ;
- du redémarrage de la chaîne.

Deux moments susmentionnés peuvent coïncider avec la localisation des noms propres au début des séquences textuelles : les changements typographiques et spatio-temporels. Sachant que les épisodes du corpus sont découpés selon un principe sémantique qui correspond au changement spatio-temporel[12], si un nom propre paraît au début d'un épisode, il peut être considéré comme coïncidence éventuelle avec un changement typographique ou avec un changement spatio-temporel.

Selon notre recensement, 21 % des noms propres de la chaîne « Jean-Christophe » se situent en début de paragraphes, équivalant à 85 maillons, dont 8 sont en début d'épisodes, équivalant à 2 % des noms propres[13]. Nous comparerons ce chiffre avec celui des autres catégories pour savoir si ce pourcentage est élevé ou non. Comme le montre l'extrait [13], le nom propre « Christophe » se localise au début du premier paragraphe de l'épisode 27, constituant un changement typographique, et « à quelque temps de là » marque un changement temporel.

[13]À quelque temps de là, un événement musical surexcita les pensées de [**Christophe**]. François-Marie Hassler, l'auteur du premier opéra qui l'avait bouleversé, allait venir. Il devait diriger un concert de ses œuvres. La ville fut en émoi. Le jeune maître était violemment discuté en Allemagne ; et, pendant quinze jours, on ne parla que de lui. Ce fut bien autre chose, quand il fut arrivé. Les amis de Melchior et ceux du vieux Jean-Michel venaient constamment aux nouvelles ; et ils en apportaient d'extravagantes sur les habitudes du musicien et ses excentricités. L'enfant suivait ces récits avec une attention passionnée. L'idée que le grand homme était là, dans sa ville, qu'il respirait le même air, qu'il foulait les mêmes pavés, le jetait dans un état d'exaltation muette. Il ne vivait plus que dans l'espérance de le voir.

Les noms propres peuvent coïncider avec le redémarrage d'une chaîne après une longue distance avec leur antécédent. Cela peut expliquer le fait que la distance entre un nom propre et son antécédent peut être plus longue par rapport aux autres catégories de maillons, ce que nous démontrerons par une comparaison de toutes les catégories à la fin de cette partie.

Dans notre corpus, la distance inter-maillonnaire moyenne entre un nom propre et son antécédent est de 2,9 phrases, comme le montre l'extrait [14].

[14]La princesse prétendait se connaître en musique ; elle [**l'**]*antécédent* aimait fort, sans faire de différence entre la bonne et la mauvaise. Elle imposait à [**Christophe**]*npr* des programmes baroques, où de plates rapsodies coudoyaient les chefs-d'œuvre.

[12] Voir la section *5.1.2. Étape 2 : partition des textes en épisodes* du *Chapitre 5* pour la présentation des principes du découpage en épisodes.

[13] Aucun nom propre ne se situe au début d'une section du roman-fleuve.

Par ailleurs, la moyenne de la distance ne reflète pas de grande variation entre tous les noms propres : une petite partie des noms propres se situent dans la même phrase que leur antécédent (24 fois, voir l'*Annexe 7*), comme le montre l'extrait [15], où nous avons marqué « 0 » comme distance entre eux, contrairement à celles qui sont marquées comme « longue » (12 fois) et « très longue » (5 fois) pour une distance respectivement de 10 – 19 phrases et de plus de 20 phrases.

[15]Il tâchait aussi de soutirer de l'argent à [Christophe], par toutes sortes de câlineries et de ruses
antécédent

ingénieuses, qui donnaient envie de rire à [**Christophe**], bien qu'il n'en eût guère sujet.
npr

6.1.2.1.2. Coïncidence avec les changements de point de vue

Baumer (2017) rejoint l'opinion de Schnedecker sur les occurrences des noms propres à des endroits particuliers dans le déroulement textuel, en signalant que ce groupe d'expressions peuvent servir à « certains phénomènes d'installation, de construction ou d'alternance du point de vue narratif » (2017 : 77). Dans la narration, un des facteurs qui influence le choix des expressions référentielles est le *point de vue*, comme le signalent les deux auteurs. Cette notion est définie par Baumer (2017 : 74), au niveau linguistique, comme un des vecteurs de « la représentation d'un objet de discours, quel qu'il soit, par une origine subjective ». Les points de vue étant implicites, ils ne sont pas facilement mesurables. Or, les expressions sont explicites et peuvent être quantifiées et démontrées par des exemples spécifiques.

Le changement de point de vue n'est pas un facteur quantifiable, puisque ce sont les exemples et les extraits du corpus qui peuvent le démontrer au cas par cas de manière qualitative. Comme remarque Achard-Bayle (2000), le point de vue narratif est fortement lié au dialogue, représenté par les discours rapportés directs dans les récits :

> *Dans les passages dialogués des fictions, les personnages parlent « réellement ». Le point de vue est donc directement représenté. La difficulté de caractériser le point de vue (l'origo énonciative par rapport à l'origo perceptive) vient des séquences proprement narratives, y compris et particulièrement celles où la narration est au style indirect libre – style indirect libre qui d'ailleurs est à l'origine même d'importants travaux contemporains sur le point de vue et le discours narratif (S. Erlich, A. Banfield). (Achard-Bayle, 2000 : 6)*

Il nous semble que les tours de parole dans les séquences dialogales peuvent refléter partiellement les changements de point de vue puisque ce sont les personnages qui prennent directement la parole, non sans adopter les attitudes subjectives, sur un événement ou sur un autre personnage. C'est pour cette raison que nous avons recensé le pourcentage des noms propres de la chaîne « Jean-Christophe » dans les séquences dialogales. Par exemple, ce type de séquence est souligné dans l'extrait [16] : trois personnages, « Christophe », « l'employé » et « le baron », prennent tour à tour la parole. La plupart des noms propres « Christophe » apparaissent au début de chaque tour de parole des trois interlocuteurs. Ils sont parfois appellatifs et appositifs en co-occurrence des déictiques.

[16][Il] passa dans le cabinet du directeur. [Christophe] attendit, sous les regards des autres employés.
M-1 *M-2*

[Il] ne savait pas ce qu'[il] devait faire. [Il] songea à se sauver, avant qu'on [lui] rapportât la réponse ;
M-3 M-4 M-5 M-6

et [il] s'y disposait, quand la porte se rouvrit :
M-7

– Son Excellence veut bien [te] recevoir, [lui] dit le trop serviable employé.
M-8 M-9

[Christophe] dut entrer.
M-10

Son Excellence le baron Hammer Langbach, un petit vieux, propret, avec des favoris, des moustaches, et le menton rasé, regarda [**Christophe**] par-dessus ses lunettes d'or, sans
M-11

s'interrompre d'écrire, ni répondre d'un signe de tête à [ses] saluts embarrassés.
M-12

– Ainsi, dit-il après un moment, [vous] demandez, [**monsieur Krafft**] ?…
M-13 M-14

– Votre Excellence, dit précipitamment [**Christophe**], [je] vous prie de [me] pardonner. [J'] ai réfléchi.
M-15 M-16 M-17 M-18

[Je] ne demande plus rien.
M-19

Le vieillard ne chercha pas à avoir une explication de ce revirement subit. Il regarda plus attentivement [**Christophe**], toussota, et dit :
M-20

– Voudriez-[vous] me donner, [**monsieur Krafft**], la lettre que [vous] tenez à la main ?
M-21 M-22 M-23

[**Christophe**] s'aperçut que le regard de l'intendant était fixé sur le papier qu'[il] continuait, sans
M-24 M-25

y penser, à froisser dans [son] poing.
M-26

– C'est inutile, Votre Excellence, balbutia-t-[il]. Ce n'est plus la peine maintenant.
M-27

– Donnez, je [vous] prie, reprit tranquillement le vieillard, comme s'il n'avait pas entendu.
M-28

[**Christophe**], machinalement, donna le chiffon de lettre ; mais [il] se lança dans un flot de paroles
M-29 M-30

embrouillées, tendant toujours la main pour ravoir la lettre. L'Excellence déplia soigneusement le papier, le lut, regarda [**Christophe**], [le] laissa patauger dans [ses] explications, puis [l'] interrompit,
M-31 M-32 M-33 M-34

et dit, avec un éclair malicieux dans les yeux :
– C'est bien, [**monsieur Krafft**]. La demande est accordée.
M-35

De la main, il [lui] donna congé et se replongea dans ses écritures.
M-36

[**Christophe**] sortit, consterné.
M-37

– Sans rancune, [**Christophe**] ! [lui] dit cordialement l'employé, quand [l'enfant] repassa par le
M-38 M-39 M-40

bureau. [Christophe] se laissa prendre et secouer la main, sans oser lever les yeux.
M-41

21 % (86 / 409) des noms propres apparaissent ici dans les séquences dialogales[14]. Par rapport aux 8 % de séquences dialogales présentes dans l'intégralité du corpus, il semble que les noms propres coïncident beaucoup plus souvent avec les dialogues qu'avec les narrations.

[14] Dont un seul se situe dans une séquence épistolaire qui est également la seule séquence épistolaire du corpus. Pour simplifier notre présentation, nous parlons de « séquences dialogales » pour désigner les séquences dialogales et la seule séquence épistolaire.

En guise de bilan, nous présentons les détails des indices de chaque nom propre et les résultats quantitatifs dans l'*Annexe 7*. Cette catégorie concerne le nombre le plus élevé de maillons dans cette chaîne, totalisant 409 maillons. Du fait de ce grand nombre, cela nous a pris, d'abord, beaucoup de temps pour obtenir tous les indices susmentionnés maillon par maillon, et, ensuite, cela nous a aussi fourni des arguments pour juger, selon l'approche quantitative, si un nom propre peut être présenté comme « beaucoup », « souvent » ou « peu », utilisé sous telle et telle condition par rapport aux autres catégories. C'est pourquoi nous avons présenté les résultats quantitatifs des noms propres de manière plutôt brève, sans trop d'explications ou de commentaires. Nous comparerons ces chiffres avec ceux des autres catégories pour mettre en évidence des tendances éventuelles et des conditions favorables pour l'usage des noms propres ou pour celui d'une autre catégorie.

6.1.2.2. SN définis

Les SN définis français font partie des formes médianes de degré d'accessibilité référentielle selon Ariel. Cette forme concerne 3 % de maillons dans la chaîne « Jean-Christophe », totalisant 140 occurrences. Les informations détaillées de chaque SN défini sont exposées dans l'*Annexe 8 : Caractéristiques discursives des SN définis français dans la chaîne « Jean-Christophe »*. Les têtes lexicales les plus fréquentes des SN définis s'avèrent être « enfant » et « petit[15] », avec l'article défini « le » sans modifieur. « L'enfant » et « le petit » ont respectivement 83 et 32 occurrences dans la chaîne. Toutes les autres têtes lexicales occupent au total 17 % (23 occurrences) de SN définis, comme « nouveau-né », « être », « maître », « cavalier », « auditoire », « garçon », « venu », « fils » (2 occurrences), « pauvre » et « homme » ; ces noms renvoient également au protagoniste selon leur situation contextuelle d'emploi. Les exemples « l'enfant » et « le petit » sont montrés dans les extraits [17] et [18] :

[17]Louisa lui fait signe de ne pas s'approcher. Elle est d'un blond presque blanc ; ses traits sont tirés ; sa douce figure mouton est marquée de taches de rousseur ; elle a des lèvres pâles et grosses, qui ne parviennent pas à se rejoindre et qui sourient avec timidité ; elle couve [**l'enfant**] des yeux – des
SN-déf

yeux très bleus, très vagues, où la prunelle est un point tout petit, mais infiniment tendre.

[18]La mère se penche hors de son lit, afin de le rassurer ; et le grand-père allume la lampe en tâtonnant, pour que [**le petit**] n'ait pas peur de la nuit.
SN-déf

Ces deux appellations du protagoniste sont susceptibles de ne se concentrer qu'au début du roman-fleuve puisqu'il s'agit du récit de la vie de celui-ci depuis sa naissance. Il est prévisible qu'elles disparaissent au fur et à mesure de la croissance du personnage. Les changements de tête lexicale indiquent bien son évolution ontologique depuis la première apparition dans l'œuvre comme un « nouveau-né », en tant que premier maillon de la chaîne, jusqu'à « (petit) homme », en passant par diverses têtes lexicales qui reflètent l'évolution de l'âge, comme « petit », « enfant » et « garçon ». Cela suppose que l'ordre des maillons ne peut pas être modifié, malgré la relation coréférentielle symétrique et transitive entre les expressions coréférentielles inclues dans la chaîne.

[15] « Petit » est un adjectif converti en nom dans le corpus.

6.1.2.2.1. Distance inter-maillonnaire

Selon la corrélation entre le nombre de phrases et la forme d'expression de Givón, les différentes distances inter-maillonnaires sont associées aux différentes catégories de maillons : les descriptions définies, complètes et incomplètes, sont censées pouvoir être utilisées après une longue distance de plus de 10 phrases pour garantir la continuité de la chaîne. Selon le modèle de Givón, il se peut que la distance inter-maillonnaire des SN définis soit moins longue que celle des noms propres et plus longue que les autres catégories d'expressions ; mais le critère quantitatif de Givón nous semble arbitraire puisque son étude ne se fonde pas sur l'observation d'un corpus.

Notre observation sur le corpus *Jean-Christophe* nous fournit des résultats relativement différents. La distance moyenne entre un SN défini et son antécédent est de 3,1 phrases, un peu plus longue que celle des noms propres (2,9 phrases) ; mais les deux sont toutefois très proches. La distinction des moyennes semble ici moins significative que celle indiquée par Givón, parce que, comme le montre l'*Annexe 8*, un grand nombre de distances sont relativement petites. Il semble que ce sont les 8 distances dites *longues* et *très longues* qui augmentent la moyenne de la distance.

6.1.2.2.2. SN défini comme premier maillon de la chaîne et des séquences textuelles

Contrairement aux noms propres, qui n'apparaissent pas au tout début du roman-fleuve, les SN définis sont fréquemment utilisés dans les premiers paragraphes et épisodes. C'est aussi un SN défini qui démarre la chaîne « Jean-Christophe », comme le montre l'extrait [1] dans la section *6.1.1.3. Titre, premier maillon et deuxième maillon*. L'usage du SN défini « le nouveau-né » comme premier maillon implique que ce personnage, qualifié en tant que *nouveau-né*, ne sera peut-être pas encore nommé, jusqu'à l'apparition du premier nom propre modifié « le petit Jean-Christophe » au dernier paragraphe de l'épisode 2, comme le montre l'extrait [19] :

[19]Mais qu'importait sans doute à la force inconnue qui l'avait rapproché de la servante aux cheveux de lin ? Il avait rempli son rôle ; et [**le petit Jean-Christophe**] venait de prendre pied sur cette terre, où le poussait son destin.

Le nom du personnage est distribué quelque temps après sa première mention dans le roman-fleuve. Pendant les deux premiers épisodes sans dénomination du protagoniste, les lecteurs peuvent se demander s'il existe une coréférence entre « le nouveau-né » et le prénom « Jean-Christophe » mentionné dans le titre du livre. Ce retard de dénomination du personnage ouvre aux lecteurs un espace de curiosité, attire leur attention et les incite à continuer la lecture.

Excepté cette amorce comme premier maillon de la chaîne, les SN définis apparaissent aussi comme premier maillon au début des paragraphes et des épisodes. 19 % des SN définis se situent au début des séquences textuelles, équivalant à 27 occurrences, dont un seul SN défini est au début d'un épisode et le reste au début de paragraphes [16]. Concluons

[16] Aucun SN défini ne se situe au début d'une section du livre.

provisoirement que les SN définis coïncident plus souvent avec les changements typographiques qu'avec des changements spatio-temporels.

6.1.2.2.3. Noms propres et SN définis : contextes multi-référentiels et fonctions syntaxiques

Dans la chaîne du protagoniste en français, entre un nom propre et son antécédent se trouvent en moyenne 6,4 référents concurrents dits *outsiders* et *challengers*. Cette moyenne pour un SN défini est de 7,1 référents. Pour ces deux groupes de maillons, il peut y avoir d'assez nombreux autres référents *outsiders* et *challengers* lors de leurs occurrences. Tout comme les noms propres, les SN définis peuvent aussi servir à éviter les ambiguïtés référentielles lors de co-présence de plusieurs référents, surtout lorsque ces derniers ont le même genre et nombre. Cette moyenne nous semble assez élevée. La raison principale en est qu'il y a au total 7 longues et très longues distances inter-maillonnaires, équivalant à 20 % des cas ; le nombre des autres référents est donc assez élevé dans ces distances.

Les fonctions syntaxiques engagées par les noms propres et les SN définis sont synthétisées dans le tableau (6.7). Comme on peut le voir, la partie majoritaire des deux groupes de maillons sont sujets de phrase, mais cette proportion est beaucoup plus élevée en ce qui concerne les noms propres (78 %) que les SN définis (57 %). Lorsque le protagoniste est mentionné par un nom propre, catégorie la plus saillante, il est aussi probablement sujet de phrase, fonction syntaxique la plus saillante. Cette spécificité est moins évidente pour les SN définis.

Tableau (6.7) : Fonctions syntaxiques des noms propres et des SN définis dans la chaîne « Jean-Christophe »

Fonction syntaxique	Noms propres	SN définis
Sujet	78 % (319)	57 % (79)
Complément d'objet	5 % (22)	23 % (32)
Complément du nom	8 % (32)	12 % (17)
Circonstant	6 % (23)	7 % (9)
Autres	3 % (12)	2 % (3)

Résumons provisoirement. Les SN définis sont à la deuxième place des maillons nominaux, bien moins nombreux que les noms propres, mais beaucoup plus nombreux que les autres maillons nominaux. La plupart se présentent par le biais des expressions « l'enfant » et « le petit », mais d'autres formes variables sont aussi nombreuses. Une grande moitié de SN définis, et ceux de l'autre moitié se répartissent à toutes les autres positions syntaxiques. La distance moyenne inter-maillonnaire ainsi que le nombre de référents dans le contexte sont très proches de ceux des noms propres. Un cinquième de ce groupe de maillons coïncide avec le début des séquences textuelles.

6.1.2.3. SN possessifs

Les SN possessifs concernent 38 maillons, soit environ 1 % de composants de la chaîne. Toutes les occurrences sont présentées avec leur contexte et leurs caractéristiques discursives dans l'*Annexe 9 : SN possessifs français dans la chaîne « Jean-Christophe »*.

La distance inter-maillonnaire entre un SN possessif et son antécédent est de 1,7 phrases en moyenne ; elle est plus courte que celle des SN définis (2,9 phrases) et celle des noms propres (3,1 phrases). Aucun SN possessif n'est sollicité après une longue distance de plus de 10 phrases. Cela signifie que les SN possessifs ne sont pas, *a priori,* utilisés pour redémarrer la chaîne après une longue distance.

Le nombre de référents entre un SN possessif et son antécédent est de 5,7 en moyenne. Nous avons constaté que ce nombre de référents peut être de zéro pour les noms propres et les SN définis, alors qu'il n'est pas difficile de constater que le même nombre de SN possessifs est au moins un, puisque leur usage implique forcément un autre personnage. C'est le référent du déterminant possessif qui intervient entre le SN possessif et son antécédent. Cela implique également que les SN possessifs ne sont *a priori* pas utilisés comme maillons du protagoniste en contexte mono-référentiel. Les SN possessifs figurent dans les contextes où d'autres personnages surviennent. Les relations entre Jean-Christophe et d'autres personnages sont instaurées par les déterminants possessifs ; dans ce cas-là, Jean-Christophe devient secondaire du point de vue de ces personnages.

Les déterminants possessifs sont les seuls déterminants qui réfèrent. Nous analyserons leurs caractéristiques avec les autres maillons non nominaux plus tard. La particularité de ce groupe de déterminants est qu'ils s'accordent avec les personnes et toutes les personnes ne s'emploient pas dans les mêmes conditions discursives. Les SN possessifs peuvent ainsi être subdivisés selon les trois personnes : ceux de la première (« je / nous »), ceux de la deuxième (« tu / vous ») et ceux de la troisième (« il (s) », « elle (s) »). Les déterminants possessifs de première et deuxième personnes sont réservés aux séquences dialogales du roman-fleuve. Le statut du narrateur n'intervient pas dans les histoires entre les personnages fictifs créés dans le cadre du roman-fleuve. Selon notre recensement, 11 SN possessifs sont à la première personne ; ils se situent tous dans les discours rapportés directs ; aucun n'est à la deuxième personne. Les 27 SN possessifs restants sont à la troisième personne ; ils se trouvent tous dans la narration à proprement parler, sans aucune exception. Les contextes discursifs des deux groupes de SN possessifs sont bien distingués l'un de l'autre. 66 % des SN possessifs sont présents dans la narration *stricto sensu*. Dans ces passages, ils coïncident souvent avec des expressions subjectives, comme « affection », de l'extrait [20], ou « critique », de l'extrait [21].

[20]Le vieux avait **une immense affection** pour [**son petit-fils**] ; et ce lui était une joie de trouver en lui
SN-poss

un public complaisant.

[21]Grand-père ne redoutait pas **la critique de** [**son jeune public**], et il s'abandonnait à son emphase
SN-poss

ordinaire ; il ne craignait pas de répéter les mêmes termes, de ne pas finir les phrases, ou même, quand il était perdu au milieu de son discours, de dire tout ce qui lui passait par la tête, pour boucher les trous de sa pensée.

Les SN possessifs se situent à toutes les positions syntaxiques, mais toutes les fonctions ne présentent pas des proportions similaires à celles des noms propres et des SN définis. Les pourcentages des fonctions de SN possessifs sont donnés dans le tableau (6.8). Les sujets

occupent la plus petite proportion (8 %), avec seulement 3 maillons ; celle des autres fonctions est répartie entre les compléments d'objet (21 %, avec 8 maillons), les compléments du nom (16 %, avec 6 maillons) et les circonstants (24 %, avec 9 maillons). La plus grande proportion de maillons se trouve sous l'étiquette « autres » (32 %, avec 12 maillons), qui concerne ici particulièrement les appellatifs.

Tableau (6.8) : Fonctions syntaxiques des SN possessifs dans la chaîne « Jean-Christophe »

	Nbre	Pourcentage
Sujets	3	8 %
Compléments d'objet	8	21 %
Compléments du nom	6	16 %
Circonstants	9	24 %
Appellatifs	12	32 %

Les exemples de chaque fonction sont visibles dans les extraits [22] – [28].

[22]Louisa, le cœur serré, comprenait vaguement que [**son petit garçon**] se privait de manger, pour
SN-poss-sujet
que les autres eussent davantage ; elle repoussait cette pensée ; mais elle y revenait toujours.

[23]Elle ne lui disait rien ; mais, une ou deux fois, quand les autres étaient sortis, les enfants dans la rue, Melchior à ses affaires, elle priait [**son aîné**] de rester, pour lui rendre quelque petit service.
SN-poss-CDO

[24]À ce moment, il aperçut dans le corridor, à l'entrée de la loge, grand-père, rayonnant et honteux, qui aurait bien voulu se montrer et dire aussi son mot, mais qui n'osait, parce qu'on ne lui avait pas adressé la parole : il jouissait de loin de la gloire de [**son petit-fils**].
SN-poss-CDN

[25]La vitre sonna une ou deux fois encore sous le doigt de la pluie. Les cloches, plus lentement, chantèrent et s'éteignirent ; et Louisa s'endormit auprès de [**son enfant**].
SN-poss-circonstant

[26]Oh ! [**mon pauvre petit**], dit-elle toute honteuse.
SN-poss-appellatif

[27][**Mon pauvre garçon**], tu n'es pas bien habile.
SN-poss-appellatif

[28]Il ne faut plus jamais parler de cela, [**mon chéri**], dors tranquillement.
SN-poss-appellatif

Les SN possessifs appellatifs sont tous utilisés dans les discours rapportés directs. Ils se composent souvent de diverses têtes lexicales avec des modifieurs. Ces derniers n'impliquent pas forcément les caractéristiques physiques du personnage mais plutôt des relations familiales ou intimes entre Jean-Christophe et les autres interlocuteurs dans les séquences dialogales, pour produire des effets affectifs, comme le montrent les extraits [26] – [28]. Cette propriété est encore plus évidente en comparaison des SN possessifs appellatifs avec les SN sans déterminant. En effet, ces premiers peuvent être remplacés par ces derniers sans modifier la structure syntaxique principale ni influencer la référence des expressions, alors qu'au niveau sémantique, les relations familiales et intimes entre les interlocuteurs disparaissent après ce type de suppression, comme le montrent les tests [29]* – [31]* dans les trois derniers extraits précédents :

[29]*Oh ! [**Pauvre petit**], dit-elle toute honteuse.
test - SN-sans-dét

[30]*[**Pauvre garçon**], tu n'es pas bien habile.
test - SN-sans-dét

[31]*Il ne faut plus jamais parler de cela, [**chéri**], dors tranquillement.
test - SN-sans-dét

Établissons un résumé. Les SN possessifs sont en général peu utilisés dans la chaîne du protagoniste, avec seulement 1 % d'occurrences. Le sujet n'est plus la position principale de cette catégorie. Aucun SN possessif n'est utilisé pour redémarrer une chaîne de référence. Ils ne coïncident pas souvent avec les changements typographiques et spatio-temporels : seuls deux SN possessifs (5 %) se situent en début de paragraphe ; aucun ne se situe en début d'épisode et de section. Ils sont utilisés dans les contextes multi-référentiels, tout comme les noms propres et les SN définis. Les contextes discursifs des SN possessifs sont radicalement différents en fonction de la personne du déterminant : ceux de la première personne ne sont utilisés que dans les séquences dialogales, contrairement à ceux de la troisième, qui se situent tous dans les séquences narratives à proprement parler.

6.1.2.4. SN démonstratifs : marqueurs de trois sous-types de rupture

Tous les indices discursifs des SN démonstratifs sont exposés dans le tableau (6.9), et avec leur contexte dans l'*Annexe 10 : SN démonstratifs français dans la chaîne « Jean-Christophe »*. Leurs caractéristiques principales peuvent être résumées en deux points : ils sont peu utilisés dans cette chaîne et ils peuvent coïncider avec certains types de rupture discursive dans la narration.

Tableau (6.9) : Caractéristiques discursives des SN démonstratifs français dans la chaîne « Jean-Christophe »

S	E	Maillons	Fonction	Distance	Réf	Discours	Début
I	3	Cette misérable masse inconsciente et informe	S	2	2	N	N
	5	Cet âne-là	S	0	2	DD	N
	10	Ce petit être	C	2	4	N	PA
II	13	Ce petit garçon	A	2	3	N	N
	20	Cet enfant	S	0	2	DD	N
	22	Ce petit garçon	A	0	2	N	N
		Ce gamin	S	3	3	DD	N
III	38	Cet individu	S	0	0	DD	N
IV	47	Cet enfant de quatorze ans, qui le jugeait	C	1	2	N	N
	50	Cet enfant	C	2	3	N	N
Total / moyenne			50 % S	0,8	2,3	4 D	1 PA

Corblin (1995) compare beaucoup les SN démonstratifs avec les SN définis puisque ce sont là deux formes d'expressions référentielles qui peuvent désigner des entités définies. L'auteur remarque que les SN démonstratifs sont censés être employés moins souvent que les SN définis dans les chaînes longues.

> *Dans les textes réels, le défini est répétable et constitue souvent le corps des chaînes de référence longues, le démonstratif est plus rare, et ne semble pas de nature à constituer le corps d'une chaîne longue dans un discours. (Corblin, 1995 : 67)*

L'auteur remarque également qu'ils ne peuvent pas être utilisés de manière répétitive, en proposant les explications suivantes :

> *La raison principale semble résider dans le fait que le démonstratif, à chaque occurrence ressaisit l'objet comme un N particulier, opposé implicitement aux autres, négligeant le fait que l'objet soit isolable dans l'univers discursif au moyen d'un signalement (ou par anaphore pronominale).*
>
> *Le démonstratif, en raison de cette indépendance vis-à-vis du contexte discursif, et de sa valeur de saisie propre d'un objet récemment mentionné par opposition à la classe virtuelle des* N, *tend à ne pouvoir apparaître qu'en certains points des chaînes de référence longues, notamment de celles qui constituent le facteur principal de cohésion d'un segment discursif.*
>
> *Schématiquement, son usage tend à être restreint aux emplois dans lesquels sa valeur propre de reclassification est motivée par le développement discursif. (Corblin, 1995 : 69)*

Dans notre corpus, les SN démonstratifs sont bien moins nombreux que les noms propres et les SN définis. Il n'y a que 10 SN démonstratifs dans la totalité de la chaîne, équivalant à 0,2 % de maillons. Le nombre de référents concurrents est de 2,3 en moyenne ; la distance moyenne entre un SN démonstratif et son antécédent n'est que de 0,8 phrase. Les SN démonstratifs sont peu nombreux en raison de leur fonction particulière dans le déroulement textuel, comme le montre l'analyse de ce groupe de maillons dans la partie suivante.

Charolles (2002) et Schnedecker (2021) signalent que l'usage des démonstratifs peut « marquer une *rupture* dans l'univers de discours ambiant » (Schnedecker, 2021 : 192). Une rupture prototypique peut être interprétée *a priori* par un changement typographique ou un changement spatio-temporel. Mais un seul SN démonstratif est obtenu en début de paragraphe, c'est-à-dire que ce sont les autres types de rupture qui sont concernés. D'après notre observation, la rupture se présente ici dans trois types de contextes discursifs relatifs aux gestes de prise de parole.

Premièrement, quatre SN démonstratifs sur 10 se situent dans les séquences dialogales, comme le montre « cet âne-là » de l'extrait [32]. Avec la prise de parole, le SN démonstratif termine la narration et marque le début de l'opinion d'un autre personnage, ce qui constitue une rupture entre le narrateur et un personnage du livre.

[32]L'enfant se laisse bercer par ces bruits. Il chantonne tout bas, puis moins bas, puis tout haut, puis très haut, jusqu'à ce que de nouveau la voix exaspérée du père crie : « [**Cet âne-là**] ne se taira donc
SN-dém
jamais ! Attends un peu, je vais te tirer les oreilles ! »

Deuxièmement, dans la narration à proprement parler, sans séquences dialogales, les SN démonstratifs sont en co-occurrences des verbes de parole qui peuvent créer un contexte semblable à celui des discours rapportés indirects. Ce type de contexte concerne 3 SN démonstratifs sur 10, comme « demander » dans l'extrait [33], co-occurrent avec « ce petit garçon ». Tout comme le premier type de contexte, celui-ci, en discours rapporté indirect, est également en lien avec la rupture entre le point de vue du narrateur et celui des personnages concernés par la prise de parole indirecte.

[33]Christophe en fut saisi. Il se dissimula dans son coin, pour ne pas être aperçu ; mais cela ne servit à rien. La dame **demanda** qui était [**ce petit garçon**] ; Louisa vint le prendre et le présenter ; elle lui
SN-dém
tenait les mains pour l'empêcher de se cacher la figure.

Troisièmement, les SN démonstratifs se situent dans des extraits de commentaires du narrateur sur des histoires et des personnages du roman-fleuve. Dans ces extraits, le narrateur

ne maintient plus son attitude invisible dans la narration ; son parti pris subjectif envers les événements et les personnages est exprimé dans et par la prise de parole ; ses propos sont ainsi présents de manière indirecte. Par exemple, l'extrait [34] avec un SN démonstratif « ce petit être » est constitué de phrases exclamatives issues du narrateur et non d'un personnage du livre. 3 SN démonstratifs sur 10 sont concernés par ce troisième type de contexte. La rupture marquée par le SN démonstratif se situe entre le point de vue du narrateur et celui des personnages créés dans le roman-fleuve.

[34]Quelle surabondance de force, de joie, d'orgueil, en [**ce petit être**]! Quel trop-plein d'énergie ! Son
SN-dém

corps et son esprit sont toujours en mouvement, emportés dans une ronde qui tourne à perdre haleine.

En résumé, notons que les trois usages de SN démonstratifs s'inscrivent dans le contexte de la rupture de la prise de parole de manière directe et indirecte entre le narrateur et les personnages du roman-fleuve.

6.1.2.5. SN indéfinis : typologie de modes de référence et caractéristiques de divers aspects

Les SN indéfinis sont une des formes les moins nombreuses dans la chaîne « Jean-Christophe ». Ils ne concernent que huit maillons, équivalant à 0,2 % de la totalité de la chaîne. Les indices de chaque maillon sont présentés dans le tableau (6.10) ; leur contexte phrastique et textuel est exposé selon leur ordre d'apparition dans le roman-fleuve dans l'*Annexe 11 : SN indéfinis français dans la chaîne « Jean-Christophe »*.

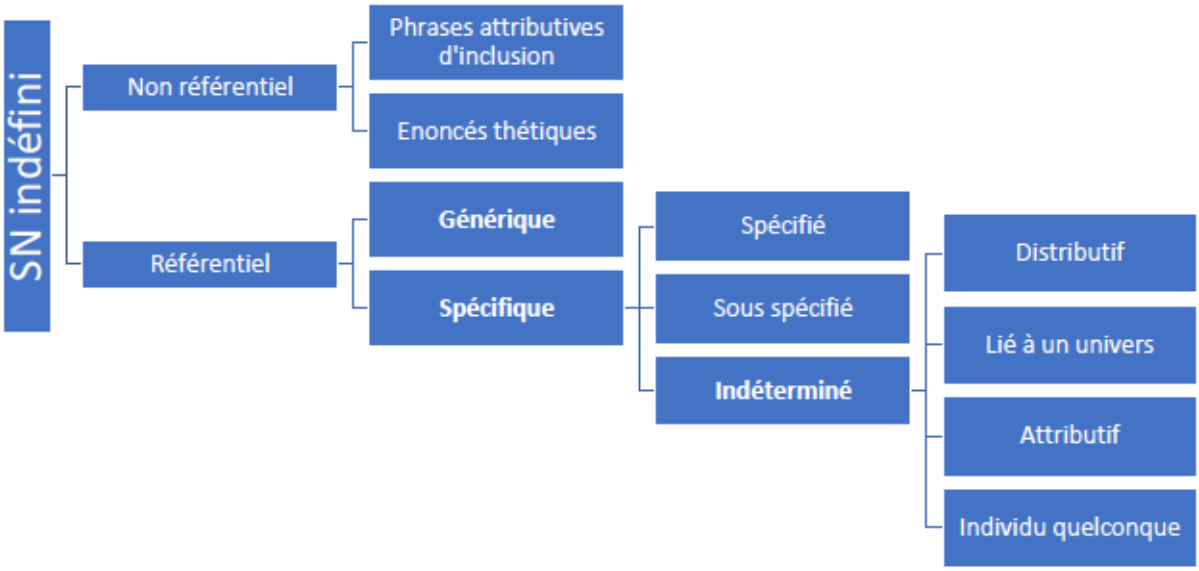
Tableau (6.10) : Caractéristiques discursives des SN indéfinis français dans la chaîne « Jean-Christophe »

S	E	Maillons	Fonction	Distance	Réf	Discours	Début
I	8	Un public complaisant	CDO	0	0	N	N
II	13	Un petit pauvre	A	0	1	N	N
	14	Un enfant	CDN	0	3	N	N
	20	Un petit garçon à qui chacun se croit le droit de faire des reproches	A	3	8	N	N
	22	Un petit garçon	A	0	0	N	N
		Un petit prodige	A	1	3	DD	N
III	33	Un mignon	A	1	0	N	PA
IV	50	Un enfant maladif et nerveux qui joue du violon à l'orchestre et écrit de médiocres concertos	A	1	0	N	N
Total / moyenne			0 S	0,8	1,9	1 D	1 PA

Les SN indéfinis sont introduits par le déterminant indéfini *un*. Ils constituent un groupe de déterminants plus nombreux que les génériques et les définis, concernant à la fois le singulier, le pluriel et le partitif, ainsi que tous ceux qui peuvent quantifier une entité ou un ensemble d'entités dénombrables ou massives (*cf*. Chen, 陈平, 1987, Charolles, 2002, Kleiber, 2005b). Les SN indéfinis ne sont pas en grand nombre dans le corpus ; les maillons sont très peu nombreux parce qu'une partie des SN indéfinis ne sont pas référentiels. Dans la

classification des SN indéfinis français de Charolles (2002), ce groupe d'expressions est bien subdivisé en *référentiels* et *non référentiels*. Les sous-types sont représentés dans la figure (6.1). Les SN indéfinis peuvent référer à une entité *générique* ou *spécifique*. Pour cette dernière, plusieurs manières favorisent les référents *indéterminés*, à savoir que la quantité d'individus de référent n'est pas fixe. Parmi toutes les catégories d'expressions référentielles que nous avons présentées (nom propre, SN défini, SN possessif, SN démonstratif), aucune ne permet de référer à une entité de quantité indéterminée. En raison de cette caractéristique, les SN indéfinis peuvent rarement être concernés dans les chaînes de référence puisque toutes les expressions de déterminants indéfinis ne réfèrent pas à la même quantité d'individus de référent. Le résultat du recensement correspond à cette caractéristique : peu de SN indéfinis sont maillons de la chaîne « Jean-Christophe ».

Figure (6.1) : Typologie des SN indéfinis français selon Charolles (2002)



Malgré la grande diversité des modes de référence des SN indéfinis, le seul déterminant indéfini des maillons du protagoniste est l'article indéfini masculin singulier « un ». Contrairement à la forme unique du déterminant, les têtes lexicales des SN indéfinis sont assez variables par rapport aux autres maillons nominaux, comme « public », « pauvre », « enfant » (2 fois), « garçon » (2 fois), « prodige » et « mignon ».

Au niveau syntaxique, selon Charolles (2002) entre autres, les SN indéfinis peuvent jouer le rôle d'introducteur de nouveaux référents dans un discours, c'est-à-dire que cette catégorie d'expressions peut donner lieu à un singleton, l'antécédent d'une anaphore ou un premier maillon de chaîne de référence. Contrairement aux noms propres et aux SN définis qui se situent le plus souvent en position de sujet de phrase, la place de complément d'objet lui est préférée pour les SN indéfinis.

> *Les SN indéfinis spécifiques ayant pour vocation d'introduire des référents nouveaux, ont une préférence pour les positions rhématiques plutôt que thématiques. Ce point a lui aussi été souligné par un grand nombre d'auteurs aussi bien à propos du français (Kleiber, 1981, 1995 ; Attal, 1997 ; Galmiche, 1984 ; Muller, 1999, etc.) que de l'anglais et d'autres langues. (Charolles, 2002 : 153)*

D'après notre observation, les SN indéfinis occupent principalement deux fonctions syntaxiques : attribut, comme « un petit pauvre » de l'extrait [35], et complément du nom, comme « (les yeux d') un enfant » de l'extrait [36] ; une seule exception à ce constat : « un public complaisant » de l'extrait [37].

[35]La fillette fit la moue et souffla à son frère – Christophe l'entendit, – que c'était [**un petit pauvre**].
SN-ind-attribut

[36]Après qu'il eut épuisé jusqu'à la dernière goutte l'incroyable provision de larmes qui tient dans les yeux d'[**un enfant**], il se sentit un peu soulagé. Il était las ; mais ses nerfs étaient trop tendus pour qu'il pût dormir.
SN-ind-CDN

[37]Le vieux avait une immense affection pour son petit-fils ; et ce lui était une joie de trouver en lui [**un public complaisant**].
SN-ind-exception

Ces proportions de fonctions syntaxiques ne ressemblent pas du tout à celles des autres expressions référentielles nominales.

Par rapport aux discours rapportés directs, les SN indéfinis sont peu utilisés dans les séquences dialogales ; un seul sur huit est repérable dans le discours rapporté direct. Relativement aux changements textuels, les SN indéfinis coïncident rarement avec le début d'une séquence textuelle pour marquer le changement typographique ou spatio-temporel : un seul SN indéfini sur huit se situe en début de paragraphe ; tous les autres ne se trouvent pas à la place initiale d'une séquence textuelle. Cette catégorie d'expressions ne sert pas non plus à redémarrer la chaîne après une longue distance. La distance moyenne est de 0,8 phrase entre un SN indéfini et son antécédent ; elle est beaucoup plus petite que celle des noms propres (2,9 phrases), des SN définis (3,1 phrases) et des SN possessifs (1,7 phrases). Le nombre de référents intervenus entre un SN indéfini et son antécédent est minime, à savoir 1,9 référents en moyenne ; pour une moitié de SN indéfinis (4 / 8), aucun référent n'intervient. Nous pouvons donc conclure que les SN indéfinis forment un groupe minimal de maillons dans l'intégralité de la chaîne, puisque c'est aussi un groupe d'expressions qui sont utilisées pour leur fonction qualitative et référentielle. Avec seulement huit maillons, le risque est grand que tous les résultats quantitatifs soient arbitraires et moins représentatifs par rapport aux autres maillons susmentionnés d'une quantité plus significative. C'est aussi pour cette raison que nous n'avons pas interprété outre mesure la signification de ces chiffres, afin d'éviter de nous engager dans une mauvaise direction.

6.1.2.6. SN sans déterminant : référence dans la narration et dans le discours rapporté direct

Nous parlons ici des SN sans déterminant pour désigner les expressions initiées directement par un nom commun, avec ou sans modifieur, dont la place initiale de déterminant est vide. Ce groupe d'expressions peut être formulé comme suit :

Ø – (modifieur(s)) – nom commun – (modifieur(s))

Selon ce critère morphologique, nous avons obtenu une toute petite quantité de huit SN sans déterminant dans la chaîne « Jean-Christophe », équivalant à 0,2 % de la totalité des maillons. Les caractéristiques de ces huit maillons apparaissent dans le tableau (6.11) suivant et sont présentés dans leur contexte dans l'*Annexe 12 : SN sans déterminant français dans la chaîne « Jean-Christophe »*.

Tableau (6.11) : Caractéristiques discursives des SN sans déterminant français dans la chaîne « Jean-Christophe »

S	E	Maillons	Fonction	Distance	Référents	Début	Discours
I	5	Petit fou	A	0	0	N	DD
II	14	Gamin	A	0	1	N	DD
	21	Gamin	A	0	0	N	DD
	22	Gamin	A	0	0	N	DD
III	30	Gamin	A	0	0	N	DD
		Petit sot	A	1	1	N	DD
	33	Petit	A	0	0	N	DD
IV	42	Pauvre petit chat	A	1	0	N	DD
Total / moyenne			0 S	0,3	0,2	0	8 D

En effet, cette catégorie n'est pas considérée comme expression référentielle dans l'échelle d'accessibilité d'Ariel (1990) pour le français, l'anglais et l'hébreu, ni dans la hiérarchie du donné référentiel de Gundel *et al*. (1993) pour l'anglais. En d'autres termes, ces auteurs ne considèrent pas les SN sans déterminant comme des expressions référentielles potentielles pour le français[17]. C'est pour cette raison qu'il est aisé d'en prévoir la rareté voire l'inexistence dans notre corpus.

Une des fonctions syntaxiques souvent occupées par les SN sans déterminant est attribut. Selon les remarques de Charolles (2002), il n'est pas garanti que les SN sans déterminant français à la fonction d'attribut de phrase puissent renvoyer à une entité :

> *Les expressions attributives font allusion à des êtres putatifs auxquels le locuteur n'a accès que par le biais de leurs attributs. [...] Ces usages s'opposent aux emplois référentiels où le SN fait allusion à une entité spécifique éventuellement sous-spécifiée. (Charolles, 2002 : 245)*

Pour cette raison, on peut anticiper le fait que l'attribut n'est pas une fonction référentielle principale des SN sans déterminant en français. Notre résultat correspond d'ailleurs à cette remarque : aucun SN sans déterminant n'a été obtenu à la place d'attribut de phrase. Cette catégorie de maillons occupe des fonctions syntaxiques très limitées : ils sont tous utilisés comme appellatifs dans les discours rapportés directs, comme le montrent les extraits [38] et [39].

[38]Sa mère se penche vers lui, et dit : « Qu'est-ce que tu as donc, [**petit fou**] ? »
SN-sans-dét

[39]Juste au moment où il descendait, il se heurta à son père qui rentrait.
– Que fais-tu là, [**gamin**] ? où vas-tu ? demanda Melchior.
SN-sans-dét

L'usage des SN sans déterminant peut être considéré comme indice des séquences dialogales, similaire à l'usage des déictiques que nous présenterons plus tard.

Résumons. Tous les SN sans déterminant sont utilisés dans des conditions presque identiques : ils sont appellatifs ; ils ont une petite distance inter-maillonnaire ; ils sont utilisés en contexte mono-référentiel ; ils ne coïncident pas avec les changements typographiques et

[17] Parce que pour ce qui est des appellatifs, ils travaillent sur des corpus écrits, manifestement sans séquence dialogale.

spatio-temporels ; ils sont utilisés dans les discours rapportés directs. Toutes ces conditions sont radicalement différentes de celles des autres maillons nominaux.

6.1.2.7. Bilan

Un bilan sur les maillons nominaux de la chaîne du protagoniste en français s'impose ici. Les remarques issues de la comparaison entre les différentes formes et les analyses de chiffres ont été présentées dans cette partie. Pour mieux comparer de manière systématique les caractéristiques de chaque catégorie de maillons, nous rassemblons les paramètres des six groupes dans le tableau (6.12) ci-dessous, afin de mettre en évidence des points similaires et des différences. Par le truchement de ce tableau, nous cherchons à établir un bilan pour obtenir les caractéristiques les plus significatives des maillons nominaux dans la chaîne « Jean-Christophe » en six colonnes : le nombre de maillons, le pourcentage de maillons en position de sujet phrastique, la distance inter-maillonnaire, le nombre de référents concurrents, le pourcentage de maillons en début de séquence textuelle et le pourcentage de maillons dans les séquences dialogales.

Tableau (6.12) : Comparaison des indices retenus des différentes catégories des maillons nominaux français dans la chaîne « Jean-Christophe »

Catégorie	Nbre	Sujet	Distance	Réf	Début	Discours
Nom propre	409	78 %	2,9	6,4	21 %	21 %
SN défini	140	57 %	3,1	7,1	19 %	18 %
SN possessif	38	8 %	1,7	5,7	5 %	34 %
SN démonstratif	10	50 %	0,8	2,3	10 %	40 %
SN indéfini	8	0	0,8	1,9	13 %	13 %
SN sans déterminant	8	0	0,3	0,2	0	100 %

- *Nombre*

La colonne *nombre* du tableau (6.12) correspond au nombre de maillons de cette catégorie dans la chaîne. Le nombre de noms propres dépasse largement celui de toutes les autres catégories nominales, étant le plus grand nombre de maillons nominaux. Avec les SN définis, ces deux groupes occupent les deux premières places, équivalant à 88 % de la chaîne. Les SN possessifs, démonstratifs, indéfinis et sans déterminant sont peu utilisés. En raison des faibles quantités de ces quatre catégories, les caractéristiques présentées par les proportions d'occurrences risquent d'être arbitraires et moins représentatives par rapport à celles des noms propres et des SN définis.

- *Sujet*

La colonne *sujet* représente le pourcentage de maillons à la fonction syntaxique de sujet dans la phrase. Étant donné le statut primordial du personnage Jean-Christophe dans le livre, les maillons de cette chaîne sont censés occuper la fonction de sujet dans le plus grand nombre de cas. Lorsque nous calculons le pourcentage de sujets de toutes les catégories, cela est confirmé : 66 % des maillons sont sujets. Mais d'importants écarts existent entre les

différentes catégories. D'après les chiffres obtenus, le nom propre constitue la seule catégorie dont la proportion de sujets dépasse largement celle des autres fonctions syntaxiques, soit 78 %. Le sujet ne concerne qu'un peu plus de la moitié des SN définis (57 %) ainsi que la moitié des SN démonstratifs (50 %). Les trois autres catégories présentent des proportions de sujets très basses : 13 % de SN indéfinis à 13 %, 10 % de SN démonstratifs, 5 % de SN possessifs et aucun SN sans déterminant.

- *Distance*

La colonne *distance* représente la distance inter-maillonnaire entre un maillon et son antécédent, en fonction du nombre de phrases.

Selon le modèle de continuité topicale de Givón (1983), les noms propres sont capables de redémarrer une chaîne de référence après une longue distance de plus de 20 phrases, et les SN définis peuvent maintenir la continuité d'une chaîne jusqu'à une distance de 10 – 20 phrases. Ce sont les deux seules catégories qui peuvent théoriquement être utilisées après une longue distance. Cela est confirmé par notre résultat : les distances des noms propres et des SN définis sont assez proches et atteignent respectivement 2,9 et 3,1 phrases en moyenne. Les noms propres sont utilisés 18 fois après une longue distance et les SN définis sept fois ; les autres formes ne sont jamais utilisées. Les petites distances moyennes sont celles des SN sans déterminant, des SN indéfinis et des SN démonstratifs, avec respectivement 0,3 phrase, 0,8 phrase et 0,8 phrase. En d'autres termes, ces trois types de maillons se situent assez souvent dans la même phrase que leur antécédent, permettant principalement de constituer des relations coréférentielles intra-phrastiques.

Par le moyen de l'approche inter-phrastique, Schnedecker (1997 : 70) constate que la saillance d'une chaîne peut être reflétée par la distance référentielle :

> *Un référent saillant (i.e. un topique continu) sera instancié par des expressions référentielles de haute accessibilité, qui se succéderont à très faible intervalle et sur une longue distance. A l'inverse, un référent non saillant (i.e. un topique discontinu) sera instancié très sporadiquement et très brièvement, ce qui nécessite l'emploi de marqueurs de faible accessibilité. (Schnedecker, 1997 : 70)*

La distance entre les maillons du protagoniste est donc censée être faible en raison de la saillance de ce personnage dans le roman-fleuve. Cependant, d'après les chiffres obtenus, nous constatons que toutes les distances moyennes ne sont pas aussi petites les unes que les autres. En effet, elles sont relativement grandes pour les noms propres et les SN définis en raison de la discontinuité de la chaîne. Les distances des quatre autres formes peuvent être considérées comme faibles : les moyennes de 0 – 1 phrase signifient que l'antécédent et le maillon en question sont dans la même phrase ; les moyennes de 1 – 2 phrases signifient que l'antécédent se situe dans la phrase précédente du maillon en question.

- *Référents*

La colonne *référents* (« réf ») présente le nombre de référents intervenus entre un maillon et son antécédent. Sachant que le roman-fleuve est un texte multi-référentiel, il est prévisible que d'autres référents se trouvent aussi dans le contexte du protagoniste : le nombre de référents n'est donc pas ici censé être zéro. Notre résultat montre globalement que ces nombres sont fortement influencés par la distance inter-maillonnaire ; autrement dit, plus

la distance est grande, plus les référents interférents sont nombreux. Par exemple, les noms propres et les SN définis correspondent au plus grand nombre de référents interférents, respectivement 6,4 et 7,1. Inversement, le nombre de ces référents exprimés par des SN sans déterminant est le moins élevé, avec seulement 0,2 référent. Par conséquent, il semble que la distance inter-maillonnaire soit le seul facteur qui influence le nombre de référents concurrents. La variation des autres facteurs ne produit pas de changement corrélatif au nombre de référents concurrents.

- *Début*

La colonne *début* contient le pourcentage de maillon qui se situe dans la première phrase d'une séquence textuelle (paragraphe, épisode et section). Nos résultats montrent que les noms propres et les SN définis sont le plus souvent utilisés au début des séquences textuelles, avec respectivement 21 % et 19 % des expressions. Ce sont les deux catégories qui coïncident le plus souvent avec des changements typographiques et spatio-temporels dans le déroulement textuel ; les quatre autres catégories présentent des nombres et des pourcentages moins significatifs, qui se situent entre 0 et 13 %. En d'autres termes, les SN possessifs, les SN démonstratifs, les SN indéfinis et les SN sans déterminant coïncident rarement avec les moments de changement.

- *Discours*

La colonne *discours* indique le pourcentage de maillons présents dans les séquences dialogales et épistolaire du roman-fleuve. Sachant que ces séquences non narratives concernent 8 % du corpus, la comparaison entre cette proportion et celle de chaque catégorie peut permettre de savoir si une catégorie est préférée dans les séquences narratives et non narratives. En effet, tous les maillons présentent une proportion supérieure à 10 % dans les séquences dialogales : les SN sans déterminant, avec la proportion la plus élevée de 100 % ; les SN indéfinis, avec la proportion la moins élevée de 13 %. En d'autres termes, les séquences dialogales favorisent l'usage des maillons des catégories nominales par rapport aux séquences narratives.

Après cette présentation comparative des usages de chaque type de maillons nominaux selon leur déterminant, nous enchaînons par leur examen selon leurs modifieurs. Après cela, nous étudierons l'usage des maillons non nominaux qui occupent des proportions beaucoup plus significatives dans la chaîne.

6.1.3. Modifieurs des maillons nominaux : longueur et restriction

Dans les expressions référentielles, la fonction primordiale des modifieurs éventuels est d'enrichir la structure syntaxique et d'ajouter des informations complémentaires aux caractéristiques des référents. Selon notre observation, les types de ces caractéristiques sont nombreux : qualités permanentes et inhérentes au référent, transformations et évolutions ontologiques ayant lieu pendant la narration, changements suscités par les différents points

de vue du narrateur et des personnages. Il n'est pas difficile de comprendre que tous les composants des modifieurs peuvent refléter ces changements, y compris la tête lexicale et tous les autres éléments dans les modifieurs. Nous avons déjà abordé la variation des têtes lexicales des maillons nominaux dans la partie précédente ; nous analyserons ici les effets engendrés par les modifieurs et les conditions discursives selon lesquelles ceux-ci seraient optionnels ou obligatoires.

Selon le critère formel, les modifieurs des maillons nominaux dans la chaîne du protagoniste se présentent sous les cinq catégories suivantes :

- les adjectifs (« le *pauvre* enfant ») ;
- les propositions subordonnées relatives (« l'enfant *qui pleure* ») ;
- les appositifs (« l'enfant, *fils de Melchior* ») ;
- les participes présents et passés (« l'enfant *se serrant contre le lit* ») ;
- les groupes prépositionnels (« l'enfant *aux cheveux longs* »).

D'après les résultats obtenus dans notre corpus, les maillons nominaux modifiés totalisent 79 occurrences dans la chaîne « Jean-Christophe ». Les expressions modifiées sont naturellement subdivisées par le groupe constitué des noms communs et celui des noms propres. Ces deux groupes concernent respectivement 43 et 36 maillons ; ils sont exposés selon l'ordre d'apparition dans le livre dans l'*Annexe 13 : Noms communs modifiés français dans la chaîne « Jean-Christophe »* et l'*Annexe 14 : Noms propres modifiés français dans la chaîne « Jean-Christophe »*.

Dans cette partie, nous nous interrogeons sur la relation entre l'apparition des noms communs modifiés et des noms propres modifiés dans la chaîne « Jean-Christophe » : toutes les catégories susmentionnées se situent-elles dans la chaîne et à quelle fréquence ? Comment sont-elles réparties dans le roman-fleuve et selon quels critères ? Quelles fonctions occupent-elles dans le texte i et dans quelles conditions textuelles spécifiques ? Les deux groupes de formes modifiées seront analysés selon leur longueur et selon leur fonction référentielle ; deux autres facteurs peuvent également entrer en jeu : leur position phrastique et leur accessibilité.

6.1.3.1. Noms communs modifiés

Nous parlons de *noms communs modifiés* pour désigner les expressions de la composition suivante :

(Déterminant) – (modifieur(s) 1) – nom commun – (modifieur(s) 2)

Le déterminant est mis entre parenthèses puisqu'il peut concerner des SN sans déterminant ; le modifieur peut être unique ou multiple, antéposé ou postposé. Dès que la composition de l'expression correspond à cette formule, elle est incluse dans le groupe des noms communs modifiés.

Par rapport aux SN sans modifieur, les SN modifiés sont tout d'abord marqués par leur longueur, qui est plus considérable et plus variable : mis à part le déterminant et la tête lexicale, les modifieurs peuvent compter un à plusieurs mots. Selon le critère de la longueur, ils peuvent être subdivisés en deux groupes : les *courts* sont composés d'adjectifs qualitatifs, puisqu'ils concernent dans la plupart des cas un seul mot ; les *longs* se constituent de toutes les autres formes composites, comme les subordonnées relatives et les groupes prépositionnels.

Les modifieurs dits *courts* concernent la grande majorité des cas dans la chaîne du protagoniste, à savoir 34 maillons sur 43 (79 %). « Petit » était catégorisé comme nom dans « le petit », maillon annoté comme SN défini sans modifieur, alors que ce mot joue le rôle de modifieur lorsque la tête lexicale est un autre nom. Dans ce cas-là, « petit » et « pauvre » sont les deux adjectifs le plus souvent utilisés pour qualifier le protagoniste ; ils ont respectivement 17 et 5 occurrences. Deux exemples de SN modifiés sont montrés dans les extraits [40] et [41] : « pauvre petit chat » est un SN sans déterminant modifié de deux adjectifs « pauvre » et « petit » antéposés au nom ; « son pianiste ordinaire » est un SN possessif modifié d'un adjectif « ordinaire » postposé au nom. Ce sont là des modifieurs courts.

[40]Christophe fut soulagé. Puis, à la pensée qu'il ne reverrait plus le cher grand-père, il pleura de nouveau, amèrement.
– [**Pauvre petit chat**] ! répétait Gottfried, regardant l'enfant avec commisération.

[41]Le grand-duc n'oubliait pas [**son pianiste ordinaire**] : non que la modique pension attribuée pour ce titre fût exactement payée, – il fallait toujours la réclamer ; – mais, de temps en temps, Christophe recevait l'ordre de se rendre au château, quand il y avait des invités de marque, ou bien quand il prenait fantaisie à Leurs Altesses de l'entendre.

Par rapport à la fonction qualitative relative aux têtes lexicales, les modifieurs expriment directement les points de vue du narrateur dans les séquences narratives et les points de vue des personnages dans les séquences dialogales. Cela signifie que les adjectifs peuvent être *descriptifs* pour qualifier les états physiques ou les caractéristiques ontologiques ; d'autres peuvent être *subjectifs* lorsqu'ils expriment des points de vue personnels en apportant des *jugements de valeur*. Les définitions suivantes des deux sous-types d'adjectifs en français sont proposées dans la *Grande Grammaire du français* :

> *On distingue les adjectifs descriptifs (*carré*), parfois appelés* objectifs *ou* classifiants*, et les adjectifs subjectifs (*passionnant*), ou* non classifiants*. Les premiers décrivent simplement des propriétés des entités, alors que les seconds font intervenir un locuteur. (Abeillé & Godard (éds), 2021 : 662)*

Cette typologie est appliquée aux adjectifs obtenus dans les maillons du protagoniste : selon le contexte spécifique de chaque occurrence, nous avons classé tous les adjectifs selon ces deux sous-types[18]. Par exemple, dans l'extrait [42], « cher » est *subjectif* car il exprime l'affection de la mère pour son enfant ; « petit » et « endormi » décrivent les états physiques et les traits ontologiques du personnage ; ils sont donc *qualitatifs*.

[42]Les trois cloches tranquilles continuaient à sonner la fête du lendemain. Louisa rêvait aussi, en les écoutant, à ses misères passées et à ce que serait plus tard [le **cher petit** enfant **endormi** auprès d'elle]. Elle était depuis des heures étendue dans son lit, lasse et endolorie.

Le type d'adjectif s'inscrit dans son contexte, c'est-à-dire que le même adjectif peut être employé comme subjectif ou qualitatif : lorsque le « petit » est qualitatif dans « le cher petit enfant » de l'extrait [42], cet adjectif est subjectif dans « mon petit poisson d'or » de l'extrait [43].

[43]La mère le presse contre elle, avec de petits mots :
« C'est fini, c'est fini, ne pleurons plus, mon jésus, [mon **petit** poisson d'or]…

[18] Certains auteurs proposent d'autres typologies des adjectifs. Les sous-types d'adjectifs sont ainsi dénommés autrement. Nous ne prenons que les deux étiquettes proposées par la *Grande Grammaire du Français* et n'avançons pas les manières de classer les adjectifs dans notre thèse.

Les modifieurs adjectivaux sont ainsi inventoriés par les deux sous-types dans le tableau (6.13) selon leur contexte : 56 % des modifieurs adjectivaux dans les maillons nominaux sont qualitatifs et 44 % sont subjectifs.

Tableau (6.13) : Adjectifs qualitatifs et subjectifs français dans les modifieurs des noms communs dans la chaîne « Jean-Christophe »

Adj	Adjectifs (nbre d'occurrences[19])	Nbre	Pourcentage
Qualitatifs	Inconscient, informe, petit (11), endormi, jeune (2), désœuvré, trop-plein, assis, honteux, heureux, blême, maladif, nerveux	24	56 %
Subjectifs	Pauvre (5), petit (6), misérable, cher, engourdi, complaisant, nouveau, malheureux, ordinaire (2)	19	44 %

Mis à part les modifieurs adjectivaux relativement courts, les autres sont dits *longs*, puisqu'ils sont souvent composés de plusieurs mots. Ces modifieurs concernent 13 maillons[20], comme le montrent les exemples dans l'extraits [44] : « l'enfant se serrant contre lui » composé d'un groupe verbal de participe présent ; « le petit, qui refusait de les regarder et de leur donner la main » composé d'une proposition subordonnée relative.

[44]Gottfried se leva et l'embrassa.
– Eh bien, mon petit, eh bien ? disait-il doucement.
– Ah ! oncle, oncle ! gémissait [**l'enfant se serrant contre lui**]. *M-mod*
– Pleure, disait Gottfried, pleure !
Les musiciens de l'orchestre éclataient de rire, et félicitaient [**le petit, qui refusait de les regarder et de leur donner la main**]. *M-mod* Melchior, l'oreille aux aguets, évaluait les acclamations qui ne s'arrêtaient point, et voulait ramener Christophe sur la scène.

En tant qu'un des types importants de modifieurs, les propositions subordonnées relatives, selon la *Grande Grammaire du Français*, se subdivisent en deux groupes en fonction de leur rôle dans les expressions référentielles : les *relatives restrictives*[21], qui servent à identifier les référents (sans elles les référents ne sont pas identifiables ou le sont difficilement par les lecteurs) ; les *relatives non restrictives*[22], au contraire, qui sont des modifieurs optionnels (sans elles les lecteurs n'ont pas de difficulté à identifier les référents[23]).

> *Une relative restrictive restreint la référence de l'antécédent. […] En revanche, une relative non restrictive ne joue pas de rôle dans la référence de l'antécédent, qui est fixée indépendamment. (Abeillé & Godard (éds), 2021 : 1473)*

Les propositions subordonnées relatives *restrictives* contribuent à qualifier les référents dans les maillons et sont obligatoires lors de l'identification du référent ; sans leur présence, il peut y avoir des risques d'ambiguïté et de sous-spécification du référent. Au

[19] Les adjectifs sans notification de nombre sont ceux dont le nombre d'occurrence et 1.

[20] Un maillon peut contenir en même temps plusieurs modifieurs courts et longs.

[21] Ou *relatives déterminatives* selon Dubois & Lagane (1973) et Galmiche (1989), entre autres.

[22] Ou *relatives appositives*, selon Dubois & Lagane (1973), entre autres ; ou *relatives descriptives*, selon Fuchs (1987), entre autres ; ou *relatives explicatives*, selon Galmiche (1989), Li (李先银, 2016) et Riegel *et al.* (2016 [1994]), entre autres.

[23] *Cf.* aussi Dubois & Lagane (1973), Teng (1981), Corblin (1987) et Liu (刘丹青, 2005), entre autres.

contraire, les *non restrictives* n'ont pas le même type d'effet pour l'interprétation du référent, dont l'identification n'est pas changée par leur absence. Ce dernier groupe de modifieurs peuvent également servir à ajouter des effets rhétoriques ou stylistiques dans la narration.

En effet, les modifieurs adjectivaux courts peuvent être regroupés selon ce critère de référence. Généralement, il semble que les modifieurs longs en subordonnées relatives ont la tendance inverse à être non restrictives lorsqu'ils sont utilisés comme composants des maillons de la chaîne, tandis que les adjectivaux manifestent les deux tendances : les qualitatifs tendent à être restrictifs puisqu'ils décrivent plutôt les caractéristiques physiques du référent, alors que les subjectifs inclinent à être appositifs, puisque, en principe, le jugement de valeur de quelqu'un ne devrait pas changer la nature du référent.

Pour vérifier notre remarque sur la base d'une première observation générale, nous poursuivons par un examen plus avancé en fonction des modifieurs détachés et attachés : le détachement marqué par une *virgule*, distinction formelle la plus superficielle (*cf*. Fuchs, 1989), fait partie des facteurs qui influencent la référence du maillon, en raison de la relation syntaxique plutôt lâche entre les deux parties. D'après l'observation du corpus, 5 modifieurs sur 43 sont détachés, comme le montrent les exemples dans les extraits [45] et [47]. Tous ces modifieurs sont non restrictifs selon le critère formel de la virgule et selon le critère *contextuel* (*cf*. aussi Fuchs, 1989) avec le test de suppression : les maillons sans modifieur permettent également de renvoyer au même référent dans le même contexte et la suppression du modifieur ne change pas la spécification référentielle dans la continuité textuelle :

[45]À la porte, Jean-Michel rencontra des gens de connaissance, et [**le petit, qui lui serrait la main très fort**], tant il avait peur de le perdre, ne comprenait pas comment ils pouvaient causer tranquillement et rire, en cet instant.
[46]*À la porte, Jean-Michel rencontra des gens de connaissance, et [**le petit**], tant il avait peur de le perdre, ne comprenait pas comment ils pouvaient causer tranquillement et rire, en cet instant.
[47]Mais comme Christophe tenait bon, Melchior n'insistait pas. Il se sentait étrangement intimidé devant les yeux sévères de [**cet enfant de quatorze ans, qui le jugeait**].
[48]*Mais comme Christophe tenait bon, Melchior n'insistait pas. Il se sentait étrangement intimidé devant les yeux sévères de [**cet enfant**].

En ce qui concerne ensuite les 38 maillons avec modifieurs attachés, le résultat de l'observation montre qu'ils peuvent tout de même être restrictifs ou non restrictifs. Le test de suppression permet de distinguer les deux types : 9 maillons sur 38 portent des modifieurs restrictifs, par exemple dans les extraits [49] et [51]. Après la suppression des modifieurs, la référence de « l'homme » dans les deux phrases n'est plus aussi claire qu'avant : ces modifieurs, y compris les adjectifs qualitatifs et toute autre forme qualitative, sont obligatoires pour identifier le référent.

[49]Pendant toute son enfance, ce fut le modèle vivant, sur lequel il eut les yeux fixés ; et c'est à son exemple que [**le petit homme de six ans**] décida, lui aussi, qu'il écrirait de la musique.
[50]*Pendant toute son enfance, ce fut le modèle vivant, sur lequel il eut les yeux fixés ; et c'est à son exemple que [**l'homme**] décida, lui aussi, qu'il écrirait de la musique.
[51]La sonate commença. [**Le petit homme**] la joua avec une sûreté imperturbable, la bouche serrée d'attention, les yeux fixés sur les touches, ses petites jambes pendantes le long de la chaise.
[52]*La sonate commença. [**l'homme**] la joua avec une sûreté imperturbable, la bouche serrée d'attention, les yeux fixés sur les touches, ses petites jambes pendantes le long de la chaise.

Les 29 modifieurs attachés restants (76 %) sont non restrictifs. Poursuivons par le même test de suppression sur ces maillons : « le cher petit enfant endormi auprès d'elle » de l'extrait [42] est raccourci en « l'enfant » de [53]* ; « l'enfant désœuvré » de l'extrait [43] est raccourci en « l'enfant » de [54]*. La référence n'est pas devenue moins claire avec les

expressions sans modifieur. Les modifieurs subjectifs, y compris les adjectifs subjectifs et toute autre forme exprimant un jugement de valeur, sont appositifs et n'influencent pas l'identification du référent.

[53]*Louisa rêvait aussi, en les écoutant, à ses misères passées et à ce que serait plus tard [**l'enfant**].
[54]*Les voyages qu'on fait plus tard, les grandes villes, les mers mouvantes, les paysages de rêves, les figures aimées, ne se gravent pas dans l'âme avec la justesse infaillible de ces promenades d'enfance, ou du simple coin de jardin tous les jours entrevu par la fenêtre, à travers la buée de vapeur que fait sur la vitre la petite bouche collée de [**l'enfant**] …

En guise de bilan provisoire, nous pouvons dire que les noms communs modifiés concernent une petite proportion de maillons dans la chaîne « Jean-Christophe » avec 43 occurrences. Ils peuvent se composer de modifieurs courts ou longs selon diverses formes qui permettent partiellement de spécifier le référent ou seulement de compléter la narration. La longueur des modifieurs présente une variété allant d'un seul mot à plusieurs, antéposés et postposés au nom. Nous allons voir dans la partie suivante si les modifieurs des noms propres présentent des différences ou des points communs.

6.1.3.2. Noms propres modifiés

La notion de *nom propre modifié* (Gary-Prieur & Noailly, 2019) inclut non seulement tous les types de modifieurs susmentionnés des noms communs, comme l'adjectif qualitatif (« Christophe assoupi » de l'extrait [55]), mais également les déterminants facultatifs. Par exemple, « le petit Christophe » de l'extrait [56] est modifié par un déterminant défini facultatif, « le », et par un adjectif qualitatif, « petit ». Contrairement aux noms communs, le déterminant peut modifier à lui seul un nom propre, comme « son Christophe » de l'extrait [57]. Dès que le maillon contient un déterminant ou d'autres types de modifieurs, il est considéré comme un nom propre modifié.

[55]Mais le suprême bonheur est lorsque maman vient, qu'elle prend la main de [**Christophe assoupi**], *npr-adj*
et que, penchée sur lui, à sa demande, elle chante à mi-voix une vieille chanson, dont les mots ne veulent rien dire.
[56]Mais qu'importait sans doute à la force inconnue qui l'avait rapproché de la servante aux cheveux de lin ? Il avait rempli son rôle ; et [**le petit Jean-Christophe**] venait de prendre pied sur cette terre, *dét-adj-npr*
où le poussait son destin.
[57]Elle n'y voyait pas malice et se réjouissait de pouvoir, grâce à cet argent, épargner un peu de peine à [**son Christophe**] et ajouter un plat au maigre souper.
dét-npr

Dans tous les cas, lorsque le nom propre n'est pas le composant unique de l'expression référentielle, les autres modifieurs et le nom propre constituent ensemble un *nom propre modifié*. Cette catégorie totalise 36 maillons, nombre légèrement moins élevé que celui des noms communs modifiés. Nous examinons ce groupe dans cette partie selon leur longueur et leur fonction référentielle.

Tout comme les noms communs modifiés, une partie des noms propres sont modifiés par des adjectifs. Avec les déterminants, ils sont dits modifieurs *courts* et concernent 22

maillons[24], dont 19 sont modifiés par des adjectifs et 3 par des déterminants. Le bilan des adjectifs se trouve dans le tableau (6.14). Au contraire des adjectifs des noms communs modifiés qui sont, pour une moitié, subjectifs, et pour l'autre, qualitatifs, les adjectifs dans les noms propres modifiés sont majoritairement subjectifs (90 %), comme « bon » de l'extrait [58], par exemple.

[58][**Le bon Christophe**] se laissait toujours prendre ; il avait un tel besoin d'être aimé qu'un mot affectueux suffisait pour désarmer sa rancune.

Au lieu de décrire les traits concrets du personnage, la plupart des adjectifs qui modifient les noms propres présentent des jugements de valeur du narrateur ou des personnages.

Tableau (6.14) : Modifieurs adjectivaux des noms propres français dans la chaîne « Jean-Christophe »

	Adjectifs (nbre d'occurrences)	Nbre	Pourcentage
Qualitatifs	Assoupi, très touché	2	10 %
Subjectifs	Petit (13), pauvre, bon (2), cher, honnête	18	90 %

Par rapport aux modifieurs longs, la proportion de ceux des noms propres est plus élevée que ceux des noms communs : 14 maillons sur 36 (39 %) sont modifiés par des éléments longs, dont les formes sont moins variables que celles des noms communs ; 12 sont des propositions subordonnées relatives, comme le montre l'extrait [59] ; les 2 qui restent concernent les groupes adjectivaux, comme le montre l'extrait [60].

[59]Le vieux, voyant pleurer son petit-fils, lui avait dit avec gravité qu'il valait bien la peine de souffrir un peu pour le plus bel art et le plus noble qui fût donné aux hommes, pour leur consolation et pour leur gloire. Et [**Christophe, qui était reconnaissant à grand-père de ce qu'il lui parlait comme à un homme**]*npr-mod-long*, avait été secrètement touché par cette naïve parole qui s'accordait avec son stoïcisme enfantin et son orgueil naissant.

[60]Après un moment, grand-père, qui était assis à côté de Christophe et lui tournait les pages, lui demanda quelle était cette musique. [**Christophe, trop absorbé par son jeu pour distinguer ce qu'il jouait**]*npr-mod-long*, répondit qu'il n'en savait rien.

Pour étudier le rôle référentiel que jouent les modifieurs dans les noms propres de la chaîne « Jean-Christophe », nous effectuons des tests en passant par les mêmes deux étapes que pour les noms communs.

La première ne concerne que les modifieurs détachés marqués par une virgule, étant donné qu'ils sont censés être non restrictifs puisque le nom propre seul peut référer de manière autonome sans susciter de questions d'ambiguïté sémantique. Les modifieurs détachés concernent 14 maillons (39 %). Avec ces derniers, nous comparons les extraits originels par le biais d'un premier test de suppression et d'un second test de modification au niveau de la structure syntaxique, comme le montrent les deux groupes de test [61] – [63] et [64] – [66]. Dans le premier, la suppression des modifieurs n'influence pas l'identification du référent des noms propres ; ils sont donc non restrictifs, mais cela réduit effectivement les

[24] Exceptionnellement, il y a 2 maillons modifiés des groupes adjectivaux qui sont considérés également comme modifieurs *longs*.

informations spécifiques dans la narration. Dans le second test, la séparation en deux phrases indépendantes permet également d'avoir la même référence, mais la micro-structure du discours est complètement différente et il risquerait d'y avoir des redondances ou des ambiguïtés avec les autres référents mentionnés dans les modifieurs. Pour ces raisons-là, les modifieurs détachés sont non restrictifs du point de vue référentiel ; leur usage permet d'éviter des répétitions et des ambiguïtés éventuelles et de diversifier les structures syntaxiques.

[61]Le vieux, voyant pleurer son petit-fils, lui avait dit avec gravité qu'il valait bien la peine de souffrir un peu pour le plus bel art et le plus noble qui fût donné aux hommes, pour leur consolation et pour leur gloire. Et [**Christophe, qui était reconnaissant à grand-père de ce qu'il lui parlait comme à un homme**], avait été secrètement touché par cette naïve parole qui s'accordait avec son stoïcisme enfantin et son orgueil naissant.

[62]*(…) Et [**Christophe**], avait été secrètement touché par cette naïve parole qui s'accordait avec son stoïcisme enfantin et son orgueil naissant.

[63]*(…) Et [**Christophe**] était reconnaissant à grand-père de ce qu'il lui parlait comme à un homme. [**Christophe / Il**] avait été secrètement touché par cette naïve parole qui s'accordait avec son stoïcisme enfantin et son orgueil naissant.

[64]Après un moment, grand-père, qui était assis à côté de Christophe et lui tournait les pages, lui demanda quelle était cette musique. [**Christophe, trop absorbé par son jeu pour distinguer ce qu'il jouait**], répondit qu'il n'en savait rien.

[65]*(…) [**Christophe**], répondit qu'il n'en savait rien.

[66]*(…) [**Christophe**] était trop absorbé par son jeu pour distinguer ce qu'il jouait. [**Christophe / Il**] répondit qu'il n'en savait rien.

La deuxième étape des deux tests concerne les autres maillons qui portent directement des modifieurs attachés, comme le montrent les deux groupes des extraits [67] et [70]. Il semble que la conclusion soit la même que celle des modifieurs détachés : la suppression ne produit pas d'effet relativement à la référence, mais c'est plutôt au niveau de la narration qu'une partie des informations sémantiques sont omises ; le changement de structure phrastique apporte aussi des répétitions ou des ambiguïtés.

[67]Christophe avait de bons poings et la conscience de son droit : il faisait marcher rondement ses cadets. Ils n'en faisaient pas moins de lui ce qu'ils voulaient ; ils abusaient de sa crédulité, ils lui tendaient des panneaux, où il ne manquait jamais de tomber ; ils lui extorquaient de l'argent, mentaient impudemment, et se moquaient de lui derrière son dos. [**Le bon Christophe**] se laissait toujours prendre ; il avait un tel besoin d'être aimé qu'un mot affectueux suffisait pour désarmer sa rancune.

[68]*(…) [**Christophe**] se laissait toujours prendre ; (…)

[69]*(…) [**Christophe**] était bon. [**Christophe / Il**] se laissait toujours prendre ; (…)

[70]Louisa coucha l'enfant. Il voulut qu'elle restât près de son lit, à lui tenir la main. Louisa passa une partie de la nuit, assise au chevet de [**Christophe qui avait la fièvre**]. L'ivrogne ronflait sur le carreau.

[71]*(…) Louisa passa une partie de la nuit, assise au chevet de [**Christophe**]. L'ivrogne ronflait sur le carreau.

[72]*(…) Louisa passa une partie de la nuit, assise au chevet de [**Christophe**]. [**Christophe / Il**] avait la fièvre. L'ivrogne ronflait sur le carreau.

Une conclusion provisoire s'impose : les modifieurs courts sont plus souvent utilisés avec les noms propres que les modifieurs longs. Tous les modifieurs des noms propres sont non restrictifs, puisque le seul nom propre sans modifieur, « Christophe », suffit pour désigner le personnage sans aucune ambiguïté. Dans ce cas-là, les modifieurs ajoutent des caractéristiques qualitatives ou descriptives relatives à l'état physique et sentimental du personnage dans les contextes spécifiques au gré de l'intrigue du livre. Après la suppression des modifieurs, ni les expressions ni les phrases ne sont incomplètes au niveau informatif sur les états ou les changements du personnage. Les lecteurs n'ont aucune difficulté à identifier

ce dernier ; mais la réduction ou l'omission des informations peut éventuellement changer la continuité textuelle.

Figure (6.2) : Maillons nominaux modifiés français dans la chaîne « Jean-Christophe »

6.1.4. Divers maillons non nominaux

Nous commençons par une présentation quantitative générale des maillons non nominaux et notons que les quatre formes non nominales le plus souvent utilisées dans la chaîne « Jean-Christophe », avec des fréquences assez éloignées, sont les *pronoms personnels de troisième personne* (62 %), les *déterminants possessifs* (16 %), les *déictiques* (4 %) et les *zéros* (4 %). Ces quatre formes mises à part, tous les autres maillons non nominaux[25] totalisent 0,1 %. Contrairement aux maillons nominaux, pour lesquels un recensement de chaque maillon est faisable, les maillons non nominaux occupent 84 % de la chaîne, c'est-à-dire que leurs occurrences sont presque omniprésentes dans le texte et qu'il n'est pas possible de trier ces expressions par SACR et CRViewer pour recenser directement leur fonction syntaxique. Par

[25] Ce groupe comprend les pronoms démonstratifs, les pronoms indéfinis et les SN sans nom. Ils sont très peu utilisés dans la chaîne « Jean-Christophe », avec seulement sept maillons au total. Nous considérons cette quantité peu significative, peu représentative et fort aléatoire, et nous n'analysons donc pas ce groupe de maillons dans notre thèse.

conséquent, il nous arrive d'illustrer et d'analyser l'usage des quatre groupes de maillons non nominaux de manière qualitative par des exemples ou par des échantillons d'exemples.

6.1.4.1. Les longues suites des maillons non nominaux

Dans la section précédente, nous avons obtenu des suites de maillons nominaux dont la longueur ne dépasse pas 9 maillons. Contrairement à ces suites, celles des maillons non nominaux atteignent facilement 20 maillons et plus. Les suites les plus longues, à savoir toutes celles de plus de 40 maillons, se situent souvent à la fin des épisodes. Par exemple, l'extrait [73] est issu de la fin de l'épisode 7, où 55 maillons non nominaux se succèdent.

[73]Si la baguette était petite, [**Christophe**] *M-nominal* se faisait chef d'orchestre ; [**il**] *M-non-nominal-1* était le chef, et [**il**] *M-non-nominal-2* était l'orchestre ; [**il**] *M-non-nominal-3* dirigeait, et [**il**] *M-non-nominal-4* chantait ; et ensuite, [**il**] *M-non-nominal-5* saluait les buissons, dont le vent agitait les petites têtes vertes. […]
Au milieu de ces jeux, [**il**] *M-non-nominal-65* avait des instants de rêvasserie étrange et de complet oubli. Tout ce qui [**l'**] *M-non-nominal-66*] entourait s'effaçait, [**il**] *M-non-nominal-67* ne savait plus ce qu'[**il**] *M-non-nominal-68* faisait, [**il**] *M-non-nominal-69* ne se souvenait même plus de [**lui-même**] *M-non-nominal-70*. Cela [**le**] *M-non-nominal-71* prenait à l'improviste. En marchant, en montant l'escalier, un vide soudain s'ouvrait… Il semblait qu'[**il**] *M-non-nominal-72* ne pensât plus à rien. Quand [**il**] *M-non-nominal-73* revenait à [**lui**] *M-non-nominal-74*, [**il**] *M-non-nominal-75* avait un étourdissement, en se retrouvant à la même place, dans l'obscur escalier. C'était comme s'[**il**] *M-non-nominal-76* avait vécu toute une vie, – l'espace de quelques marches.

Cette composition ne nous surprend pas puisque de nombreuses études, présentées dans le *Chapitre 2*, montrent déjà que les maillons non nominaux maintiennent la continuité de la chaîne de référence, à condition qu'il n'y ait pas d'autres personnages ou d'autres référents qui interviennent ou soient potentiellement concurrents de la chaîne. Se pose donc la question du nombre de référents et de chaînes dans le contexte des suites. Combettes (1986) mentionne une régularité d'ordre discursif : il faudrait qu'aucun autre référent concurrent n'existe dans le même contexte pour que la forme pronominale ne crée pas d'ambiguïté référentielle.

> *L'utilisation d'un pronom-substitut n'est pas à interpréter comme signe d'économie, ni comme renvoi à tel ou tel antécédent, mais comme indice qu'il n'y a pas d'autre sujet-thème à introduire dans le texte ; même analyse en ce qui concerne la réception : le lecteur ne se demande pas à chaque occurrence de « il » ou de « elle » quel est le groupe remplacé ; il continue, intuitivement et naturellement, pourrait-on dire, ses enchaînements sur le même thème. (Combettes, 1986 : 81)*

Selon notre observation, les suites de maillons non nominaux ne sont pas seulement utilisées dans le contexte mono-référentiel : les expressions d'autres référents peuvent également intervenir, en constituant éventuellement un contexte multi-référentiel. Par exemple, l'extrait [73] contient des maillons de la chaîne « Jean-Christophe » dans l'épisode 3 du corpus. La chaîne de cet extrait est composée d'un SN défini « l'enfant » à la position initiale et d'une succession de 15 maillons non nominaux, constituée principalement de pronoms personnels de troisième personne et de déterminants possessifs. Cependant, l'extrait [74]

renferme aussi les expressions coréférentielles d'un autre personnage, « la mère », et la chaîne d'un référent abstrait, « la souffrance », avec huit maillons. Ce sont là, d'après la définition de Schnedecker (1992), des référents *challengers* à la saillance du protagoniste. Or, les pronoms et les déterminants possessifs réussissent à garder la place primordiale du protagoniste parce que ces référents n'ont pas le même genre quand ils sont repris par les catégories non nominales. Les lecteurs n'accordent pas « elle » à un personnage masculin ; donc l'usage des maillons non nominaux ne crée pas d'ambiguïté entre le référent *leader* et les *challengers*.

[74]Dans le lit, près de [la mère], [**l'enfant**] s'agitait de nouveau. [Une souffrance inconnue]
personnage-secondaire *M-nominal* *réf-abstrait-M-1*

montait du fond de [**son**] être. [**Il**] se raidit contre [elle]. [**Il**] tordit [**son**] corps, [**il**]
M-non-nominal-1 *M-non-nominal-2* *réf-abstrait-M-2* *M-non-nominal-3* *M-non-nominal-4* *M-non-nominal-5*

serra les poings, [**il**] fronça les sourcils. [La douleur] grandissait, tranquille, sûre de [sa] force.
M-non-nominal-6 *réf-abstrait-M-3* *réf-abstrait-M-4*

[**Il**] ne savait pas ce qu'[elle] était, ni jusqu'où [elle] allait. Elle [**lui**] paraissait immense, et ne devoir
M-non-nominal-7 *réf-abstrait-M-5* *réf-abstrait-M-6* *M-non-nominal-8*

jamais prendre fin. Et [**il**] se mit à crier lamentablement. [[**Sa**] mère] [**le**] caressa
M-non-nominal-9 *M-non-nominal-10* *personnage-secondaire* *M-non-nominal-11*

avec de douces mains. Déjà [la souffrance] devenait moins aiguë. Mais [**il**] continuait de pleurer ;
référent-abstrait-M-7 *M-non-nominal-12*

car [**il**] [la] sentait toujours près de [**lui**], en [**lui**].
M-non-nominal-13 *réf-abstrait-M-8* *M-non-nominal-14* *M-non-nominal-15*

Par ailleurs, il se peut également que l'emploi de chaque forme de maillon non nominal dépende d'autres facteurs dans le contexte ; nous examinerons donc respectivement les exemples de chaque forme de maillon pour en dégager quelques régularités.

6.1.4.2. Pronoms déictiques « je » et « tu » : séquences dialogales *vs* séquences narratives

Les pronoms de première et deuxième personnes (« je », « tu ») sont classés dans un groupe autre que celui de troisième personne étant donné que ce dernier ainsi que les deux premiers ne présentent pas le même degré d'accessibilité pour les lecteurs : le référent des pronoms déictiques est plus accessible que les pronoms de troisième personne puisque les premiers sont présents directement dans la situation d'énonciation (*cf.* Charolles, 2002), tandis que les derniers ne sont utilisés que dans le contexte linguistique.

Le résultat de notre annotation révèle que les expressions déictiques sont composées de deux types : les pronoms personnels déictiques de première et deuxième personne (« je », « tu ») et les expressions spatio-temporelles déictiques (« ici », « maintenant »). En ce qui concerne les maillons de la chaîne « Jean-Christophe », seuls sont concernés : « je » et « tu », qui réfèrent au protagoniste, impliquant que celui-ci est en même temps un participant au dialogue (*cf.* Bhat, 2004). Ces deux pronoms déictiques présentent 171 occurrences dans la chaîne « Jean-Christophe », soit 4 % des composants.

L'usage de « je » et « tu » peut être l'indice des séquences dialogales du roman-fleuve, comme le montre l'extrait [75], puisque tous les pronoms déictiques obtenus sont utilisés dans les discours rapportés directs des personnages, dont un des interlocuteurs est le protagoniste.

[75]Christophe se jeta à son cou, et sanglota :
– Papa, mon cher papa ! [**Je**] ne te méprise pas ! [**Je**] suis si malheureux !
déictique-1 *déictique-2*

Ils pleuraient bruyamment. Melchior se lamentait :

– Ce n'est pas ma faute. Je ne suis pourtant pas méchant.
Il promettait de ne plus boire. Christophe hochait la tête, d'un air de doute ; et Melchior convenait qu'il ne pouvait pas résister, quand il avait de l'argent dans les mains. Christophe réfléchit, et dit :
– Sais-tu, papa, il faudrait…
Il s'arrêta.
– Quoi donc ?
– [**J'**] ai honte…
déictique-3
– Pour qui ? demanda naïvement Melchior.
– Pour toi.

6.1.4.3. Zéro : diverses contraintes syntaxico-discursives

L'usage des zéros concerne 154 maillons, soit 4 % des maillons dans la chaîne « Jean-Christophe ». Par rapport aux trois autres catégories non nominales qui maintiennent la continuité référentielle, les maillons zéro sont moins nombreux étant donné que leur usage connaît plus de contraintes que celles des autres catégories. En ce qui concerne les règles grammaticales et syntaxiques, les zéros français se situent dans les propositions coordonnées avec plusieurs verbes prédicatifs. Le sujet de la proposition initiale peut être un SN ou un pronom, alors que, dès la deuxième proposition coordonnée, le même sujet peut être exprimé ou non. Dans ce dernier cas, il peut donc y avoir des zéros, comme le montre l'extrait [76] :

[76]Quand ils louaient ainsi une œuvre qu'[il] aimait, il [lui] semblait qu'on [l']outrageait [lui-même]. [Il] se raidissait, [**Ø**] blêmissait, [**Ø**] prenait un air glacial, [**Ø**] affectait de ne pas s'intéresser à la
zéro-1 *zéro-2* *zéro-3*
musique ; [il] l'eût haïe, si c'eût été possible.

Selon l'approche cognitive, Ariel (1990 : 80-83) expose deux principes à suivre pour distinguer le degré de saillance référentielle d'une expression, lesquels aident à comprendre l'usage particulier du zéro par rapport aux autres formes explicites. Le premier concerne la quantité d'informations que l'expression renferme : plus l'expression est enrichie par des informations, moins le référent est accessible.

> *Definite descriptions encode a lower degree of accessibility than demonstratives, pronouns and zeros, because they are far more informative. (Ariel, 2013 : 4)*

Le second principe se rapporte aux caractéristiques d'atténuation phonétique : plus une expression est atténuée ou concise, plus accessible est le référent.

> *Phonetic attenuation is the third criterion for associating degree of accessibility with specific forms. Stressed pronouns, nonstressed pronouns, cliticized pronouns and person verbal agreement markers are equally informative and rigid in Hebrew. But they are not equally attenuated. The more attenuated forms (the agreement markers) encode a higher degree of accessibility than the less attenuated forms (stressed pronouns), with the other forms constituting intermediate accessibility markers. (Ariel, 2013 : 4)*

Selon les deux principes en question, l'usage de zéro représente le plus haut degré d'accessibilité du référent, avec le lien le plus fort avec le contexte.

> *High Accessibility Markers are mainly those which the "geographic" definition would claim to be text-dependent. Thus pronouns and gaps belong in this category, though, these markers are also appropriate when referring to extra-textual information. (Ariel, 1990 : 56)*

Du point de vue discursif, en fonction du nombre de référents saillants, le roman-fleuve se constitue *a priori* d'un contexte multi-référentiel, ce qui implique une condition défavorable pour l'emploi des zéros. Selon le tableau (6.2) à la première section de ce chapitre qui montre la répartition des chaînes en épisodes, seuls 4 d'entre eux sont mono-référentiels : il s'agit des épisodes 7, 24, 48 et 50, dans lesquels la chaîne du protagoniste constitue la seule chaîne dominante de l'épisode, contrairement au cas de cohabitation entre deux ou plusieurs chaînes concurrentes dans les 46 épisodes restants.

Le recensement des maillons indique que les zéros n'occupent que 2 % de la chaîne dans les 4 épisodes mono-référentiels, soit une proportion encore plus réduite que celle de la chaîne dans la totalité du corpus qui est de 4 %. Pourtant, les pronoms personnels de troisième personne et les déterminants possessifs constituent la majorité absolue de la chaîne, avec 91 % des maillons, nombre encore plus élevé que les 84 % de ces deux catégories dans l'intégralité de la chaîne. Par conséquent, il semble que le contexte mono-référentiel ne soit pas une condition *sine qua non* qui favorise l'usage des zéros en français.

Tableau (6.15) : Composition de la chaîne du protagoniste dans les épisodes mono-référentiels

Catégorie du maillon	Pourcentage (nbre)
Nom propre	4 % (12)
SN défini	2 % (5)
SN démonstratif	0,3 % (1)
SN indéfini	0,3 % (1)
Pronom personnel de troisième personne & déterminant possessif	91 % (270)
Déictique	0,3 % (1)
Zéro	2 % (6)

En raison de la grande quantité de maillons non nominaux, il nous était impossible de résumer les tendances et les caractéristiques de ces catégories par l'examen de toutes les occurrences au cas par cas. Les exemples ont néanmoins permis d'illustrer quelques usages de nature qualitative.

6.2. Chaînes des trois personnages *secondaires* en français

Comme le montre le tableau (6.2) dans l'introduction de ce chapitre, les chaînes les plus longues et les plus réparties du roman-fleuve sont celles des personnages : celle du protagoniste, que nous avons présentée dans la partie précédente, et celles des personnages *secondaires*, que nous abordons à présent. Parmi toutes les chaînes des personnages, aucune ne présente de longueur et de répartition semblables à celles de la chaîne « Jean-Christophe », qui comprend 4366 maillons et couvre 49 épisodes, soit 98 % de l'intégralité du corpus. Outre cette chaîne, qui comporte le plus grand nombre de maillons et la distribution la plus large, trois autres sont repérées dont la distribution s'étend aux quatre sections du roman-fleuve dans le corpus, mais avec des discontinuités dans certains épisodes. Ces trois chaînes semblent partager trois points communs :

- elles ont des longueurs relativement proches de celles des autres chaînes du corpus, contenant entre 500 et 1000 maillons ;
- elles occupent des proportions similaires d'environ une moitié du corpus ;
- leur premier maillon apparaît dans le premier épisode de la première section du roman-fleuve, presqu'en même temps que celui du protagoniste.

Le tableau (6.16) suivant montre les deux premiers points communs : la chaîne « Louisa » (mère du protagoniste) comprend 511 maillons et couvre 28 épisodes, soit 56 % du corpus ; la chaîne « Melchior » (père du protagoniste) contient 788 maillons et couvre 32 épisodes, ce qui représente 64 % du corpus ; la chaîne « Jean-Michel » (grand-père du protagoniste) comporte 917 maillons et couvre 27 épisodes, soit 54 % du corpus. Ces trois chaînes présentent des longueurs et des répartitions relativement proches en mesure du nombre de maillons et d'épisodes occupés, par rapport aux autres chaînes.

Tableau (6.16) : Nombre de maillons et répartition des chaînes des personnages secondaires en français

Nom de chaîne	Nbre de maillons	Répartition (nbre d'épisodes)	Répartition (proportion)
« Louisa »	511	28	56 %
« Melchior »	788	32	64 %
« Jean-Michel »	917	27	54 %

En raison de ces points communs et des grandes différences avec la chaîne du protagoniste et les autres chaînes du corpus, ces trois chaînes sont considérées comme *secondaires*, car renvoyant à trois personnages *secondaires*, à savoir « Louisa », mère du protagoniste, « Melchior », son père et « Jean-Michel », son grand-père. Leurs maillons sont exposés dans les *Annexes 15 – 17*[26]. Outre les trois points communs que nous venons de souligner, ces trois chaînes présentent aussi des similitudes et des différences au niveau de leur composition. Le tableau (6.17) synthétise leurs groupes de maillons selon le critère morpho-syntaxique.

Tableau (6.17) : Compositions des chaînes des personnages secondaires

Composant	« Louisa »	« Melchior »	« Jean-Michel »
Noms propres	14 % (75)	18 % (145)	5 % (50)
SN définis	2 % (12)	1 % (10)	13 % (115)
SN possessifs	12 % (60)	7 % (54)	2 % (19)
SN sans déterminant	2 % (11)	2 % (12)	10 % (96)
SN indéfinis	1 % (6)	0,3 % (2)	0,3 % (3)
SN démonstratifs	0	0,1 % (1)	0,1 % (1)
Pronoms personnels	42 % (212)	41 % (400)	42 % (345)
Déterminants possessifs	10 % (52)	14 % (110)	16 % (149)
Déictiques	9 % (45)	2 % (19)	4 % (39)
Zéro	7 % (36)	4 % (35)	7 % (61)

[26] Voir l'*Annexe 15 : Maillons français dans la chaîne « Louisa »*, l'*Annexe 16 : Maillons français dans la chaîne « Melchior »* et l'*Annexe 17 : Maillons français dans la chaîne « Jean-Michel »*.

Dans cette partie, nous analyserons ces trois chaînes comme un seul groupe en raison des points qu'elles partagent, et aussi comme trois chaînes distinctives étant donné les différences significatives obtenues par la comparaison.

6.2.1. Premier maillon : introduction des personnages secondaires

L'ordre d'apparition des référents implique la saillance des référents (*cf.* Schnedecker, 1997, 2019). Par exemple, Sanford *et al.* (1988 : 44, cités par Schnedecker, 1997 : 74) signalent que les personnages principaux ont tendance à être mentionnés avant les personnages secondaires. Gernsbacher & Hargreaves (1998 : 699, cités par Schnedecker, 1997 : 75) observent un *effet de primauté* entre les référents, à savoir que les personnages (ou d'autres types de référents) introduits en première mention sont plus facilement mémorisés par les lecteurs que les suivants. Par conséquent, les personnages principaux sont souvent introduits avant les personnages secondaires.

Les deux personnages, « Jean-Michel » et « Louisa », apparaissent dès le deuxième paragraphe du livre, sous forme de SN définis « le vieux » et « la mère », comme le montre l'extrait [77].

[77]Le grondement du fleuve monte derrière la maison. La pluie bat les carreaux depuis le commencement du jour. Une buée d'eau ruisselle sur la vitre au coin fêlé. Le jour jaunâtre s'éteint. Il fait tiède et fade dans la chambre.
[Le nouveau-né] s'agite dans [son] berceau. Bien que [**le vieux**] ait laissé, pour entrer, [ses] sabots
protagoniste-1 *protagoniste-2* *secondaire-JM-1* *secondaire-JM-2*
à la porte, [son] pas a fait craquer le plancher : [l'enfant] commence à geindre. [**La mère**] se
secondaire-JM-3 *protagoniste-3* *secondaire-L-1*
penche hors de [son] lit, afin de [le] rassurer ; et [le grand-père] allume la lampe en tâtonnant,
secondaire-L-2 *protagoniste-4* *secondaire-JM-4*
pour que [le petit] n'ait pas peur de la nuit. La flamme éclaire la figure rouge [du vieux
protagoniste-5 *secondaire-JM-5*
Jean-Michel], [sa] barbe blanche et rude, [son] air bourru et [ses] yeux vifs. [Il] vient près du
secondaire-JM-6 *secondaire-JM-7* *secondaire-JM-8* *secondaire-JM-9*
berceau. [Son] manteau sent le mouillé ; [il] traîne en marchant [ses] gros chaussons bleus.
secondaire-JM-10 *secondaire-JM-11* *secondaire-JM-12*
[Louisa] [lui] fait signe de ne pas [s'] approcher. [Elle] est d'[un blond presque blanc] ; [ses]
secondaire-L-3 *secondaire-JM-13* *secondaire-JM-14* *secondaire-L-4* *secondaire-L-5* *secondaire-L-6*
traits sont tirés ; [sa] douce figure mouton est marquée de taches de rousseur ; [elle] a des lèvres
secondaire-L-7 *secondaire-L-8*
pâles et grosses, qui ne parviennent pas à se rejoindre et qui sourient avec timidité ; [elle] couve
secondaire-L-9
[l'enfant] des yeux – des yeux très bleus, très vagues, où la prunelle est un point
protagoniste-6
tout petit, mais infiniment tendre.

Le troisième personnage, « Melchior », apparaît quelques paragraphes plus loin, au milieu du premier épisode, sous la forme du SN possessif « ton mari », dans un dialogue entre les deux autres personnages secondaires, comme le montre l'extrait [78].

[78][Il] se tut un instant, méditant s'[il] ne conviendrait pas de développer cette pensée ; mais [il]
secondaire-JM-15 *secondaire-JM-16* *secondaire-JM-17*
ne trouva rien de plus à dire ; et, après un silence, [il] reprit d'un ton irrité :
secondaire-JM-18
– Comment se fait-il que [[**ton**] **mari**] ne soit pas ici ?
secondaire-L-10 *secondaire-M-1*

– [Je] crois qu'[il] est au théâtre, dit timidement [Louisa]. [Il] a répétition.
secondaire-L-11 *secondaire-M-2* *secondaire-L-12* *secondaire-M-3*

Certes, le premier maillon des trois personnages surgit après le protagoniste ; cependant, étant donné la longueur de ce roman-fleuve, ce retard de quelques phrases et paragraphes est peu significatif : ces trois personnages peuvent être considérés comme se situant au début de l'intégralité du roman-fleuve. Avec le déroulement de l'histoire cité dans les extraits [77] et [78], la saillance des deux premiers personnages secondaires semble même dépasser celle du protagoniste en raison du nombre de maillons supérieur à ceux de la chaîne de ce dernier. Au niveau syntaxique, le premier maillon des trois chaînes joue le rôle de sujet de phrase, tout comme le premier maillon du protagoniste. Donc, à ce niveau, il s'avère également que les personnages secondaires sont introduits avec le même degré de saillance que celui du protagoniste.

En raison de ces trois statuts similaires au niveau de la localisation dans le texte, de la longueur et de la fonction syntaxique, le protagoniste et les trois personnages secondaires semblent aussi saillants les uns que les autres au début du livre : la saillance *locale* des référents secondaires peut égaler ou même dépasser celle du protagoniste selon les extraits. Ce que nous définissons comme *secondaires* et *protagoniste* concerne seulement la saillance *globale* des référents dans l'intégralité du corpus.

Lorsqu'on parle de la notion de *saillance linguistique*, les psycholinguistes proposent des échelles graduelles en lien avec les formes morpho-syntaxiques. Par exemple, Sanford *et al*. (1988, cité par Schnedecker, 1997) proposent une hypothèse relative à l'influence de la *nature grammaticale de l'expression référentielle* sur la saillance, en indiquant que la forme de l'expression est un facteur décisif. Les noms propres sont particulièrement marqués comme introducteurs potentiels de référents plus importants que ceux des noms communs :

> *Les personnages principaux ont tendance à être introduits au moyen de noms propres tandis que les personnages secondaires sont souvent introduits par des descriptions de rôle (par exemple : le garçon, la serveuse, etc.). (Sanford* et al., *1988, cité par Schnedecker, 1997 : 50)*

Cette remarque peut expliquer partiellement ce que nous constatons dans notre corpus : les trois personnages secondaires sont certes introduits par des SN comportant des noms communs au lieu de l'être par des noms propres ; mais rappelons que le protagoniste est également initié par un SN de nom commun, « le nouveau-né ». De ce point de vue, les quatre personnages présentent également le même degré de saillance au début du livre.

6.2.2. Noms propres et SN définis : différences suscitées par l'identité familiale

Les proportions de noms propres dans les chaînes « Louisa » et « Melchior » sont assez proches, respectivement 14 % et 18 %, alors que celle de la chaîne « Jean-Michel » n'atteint que 5 %, avec 50 maillons concernés, proportion et quantité relativement moins élevées que les deux premiers. Ce troisième personnage secondaire tend à être désigné par des noms communs qui reflètent son statut social ou qui lui confèrent des traits qualitatifs, contrairement aux deux premiers qui sont nommés plus souvent de manière neutre par leur

prénom sans exprimer aucune caractéristique. Quelques exemples de noms propres des trois chaînes sont visibles dans l'extrait suivant[27].

[79][Elle] était domestique, quand [elle] avait épousé, à la surprise de tous, et surtout à la [sienne],
L-1 *L-2* *L-3*

[**Melchior Krafft, le fils de** [**Jean-Michel**]]. Les Krafft étaient sans fortune, mais considérés dans
JM-1-npr *M-1-npr*

la petite ville rhénane, où [le vieux] s'était établi, il y avait presque un demi-siècle. Ils étaient
JM-2-SN-déf

musiciens de père en fils et connus des musiciens de tout le pays, entre Cologne et Mannheim. [**Melchior**] était violon au Hof-Theater ; et [**Jean-Michel**] avait dirigé naguère les concerts du
M-2-npr *JM-3-npr*

grand-duc. [Le vieillard] fut profondément humilié du mariage de [**Melchior**] ; [il] bâtissait de grands
JM-4-SN-déf *M-3-npr* *JM-5*

espoirs sur [[son] fils] ; [il] eût voulu en faire l'homme éminent qu'[il] n'avait pu être [lui-même].
JM-6 *M-4-SN-poss* *JM-7* *JM-8* *JM-9*

Ce coup de tête ruinait [ses] ambitions. Aussi avait-[il] tempêté d'abord et couvert de malédictions
JM-10 *JM-11*

[**Melchior**] et [**Louisa**]. Mais, comme [il] était [un brave homme], [il] avait pardonné à [[sa]
M-5-npr *L-4-npr* *JM-12* *JM-13-SN-ind* *JM-14* *JM-15*

bru], dès qu'[il] avait appris à [la] mieux connaître ; et même, [il] s'était pris pour [elle] d'une affection
L-5-SN-poss *JM-16* *L-6* *JM-17* *L-7*

paternelle, qui se traduisait le plus souvent par des rebuffades.
Nul ne pouvait comprendre ce qui avait poussé [**Melchior**] à ce mariage, – [**Melchior**] moins que
M-6-npr *M-7-npr*

personne. Ce n'était certes pas la beauté de [**Louisa**]. Rien en [elle] n'était fait pour séduire :
L-8-npr *L-9*

[elle] était petite, pâlotte et frêle ; et [elle] faisait [un singulier contraste avec [**Melchior**] et
L-10 *L-11* *L-12-SN-ind* *M-8-npr*

[**Jean-Michel**]], tous deux hauts et larges, des colosses à la figure rouge, au poing solide, mangeant
JM-18-npr

bien, buvant sec, aimant rire, et faisant grand bruit. [Elle] semblait écrasée par eux ; on ne [la]
L-13 *L-14*

remarquait guère ; et [elle] cherchait à s'effacer encore plus. Si [**Melchior**] avait eu bon cœur, on
L-15 *M-9-npr*

eût pu croire qu'[il] avait préféré à tout autre avantage la simple bonté de [**Louisa**] ; mais [il] était
M-10 *L-16-npr* *M-11*

[l'homme le plus vain]. Qu'un garçon de [son] espèce, assez beau et ne l'ignorant pas, très fat, non
M-12-SN-déf *M-13*

sans talent, et pouvant prétendre à quelque riche parti, capable même – qui sait ? – de tourner la tête à une de [ses] élèves bourgeoises, ainsi qu'[il] s'en vantait, eût été brusquement choisir [une
M-14 *M-15* *L-17-SN-ind*

fille du peuple, pauvre, sans éducation, sans beauté, qui ne [lui] avait fait aucune avance] …
M-16

on eût dit une gageure !

Contrairement aux noms propres, les proportions de SN définis dans les trois chaînes secondaires montrent des différences inverses de celles des premiers : les chaînes « Louisa » et « Melchior » présentent des proportions relativement proches et très peu élevées, 2 % pour la chaîne « Louisa » et 1 % pour la chaîne « Melchior », avec respectivement 12 et 10 maillons[28] ; mais, en revanche, la chaîne « Jean-Michel » affiche une proportion beaucoup plus importante de 13 %, équivalant à 118 maillons : ce personnage est donc susceptible d'être plus

[27] Cet extrait contient également d'autres maillons nominaux et non nominaux des trois chaînes. Ils sont tous annotés entre crochets, cela pour éviter les citations redondantes dans les analyses suivantes.

[28] Voir l'*Annexe 18 : SN définis français dans la chaîne « Louisa »* et l'*Annexe 19 : SN définis français dans la chaîne « Melchior »*.

souvent désigné par « le vieux » que par d'autres désignations, comme « le grand-père (du protagoniste) », « le père (du père du protagoniste » et « le beau-père (de la mère du protagoniste) ». Nous ne citons pas d'autre extrait pour montrer les SN définis des trois chaînes puisque quelques exemples sont déjà présents dans l'extrait [79], comme « le vieux » (« JM-2-SN-déf ») et « l'homme le plus vain » (« M-12-SN-déf »).

S'il existe des régularités éventuelles pour expliquer ces différences entre noms propres et SN définis dans les chaînes secondaires, il nous semble qu'elles sont suscitées par la relation entre les personnages : étant donné leur statut familial de parents, « Louisa » et « Melchior » sont plus souvent mentionnés par leur prénom, alors que le grand-père « Jean-Michel » l'est moins souvent mais plutôt par son lien familial. Le SN défini est utilisé pour marquer le respect et la politesse par rapport à la génération des enfants et à celle des petit-enfants.

6.2.3. SN possessifs : reflet de divers degrés de dépendance et de relation familiale

À la différence des deux groupes de maillons examinés, les proportions des SN possessifs sont largement différentes entre les trois chaînes : elles s'élèvent à 12 % dans la chaîne « Louisa », 7 % dans la chaîne « Melchior » et 2 % seulement dans la chaîne « Jean-Michel[29] ». Quelques exemples sont montrés dans l'extrait [79], comme « son fils » (« M-4-SN-poss ») et « sa bru » (« L-5-SN-poss »). Il nous semble que ces écarts sont dus aux différents degrés de relation entre les personnages. Ces degrés sont liés non seulement à la hiérarchie générationnelle, mais également à l'inégalité du statut familial : la domination familiale du grand-père (par rapport à son fils et à son petit-fils) et celle du mari (par rapport à sa femme) sont reflétées par les SN possessifs relativement peu utilisés, puisque ces personnages de statut supérieur sont moins souvent mentionnés en relation avec les autres. Dans cette hiérarchie, le grand-père « Jean-Michel » est sans doute le plus souvent présenté comme repère de la relation familiale et comme personne de statut indépendant par rapport aux autres personnages. Le père « Melchior » occupe la place intermédiaire : plus dépendant que le grand-père et moins dépendant que la mère. Cette dernière, « Louisa », est celle qui est le plus susceptible d'être dépendante des autres personnages : son beau-père, son mari et son fils.

6.2.4. SN sans déterminant : appellatifs dans les séquences dialogales

La différence entre les SN sans déterminant des trois chaînes ressemble à celle des SN définis : ce groupe de maillons est peu important dans les chaînes « Louisa » et « Melchior » (seulement 2 % dans les deux chaînes), mais ils sont beaucoup plus nombreux dans la chaîne « Jean-Michel », totalisant 10 %. Quelques exemples de SN sans déterminant de la chaîne « Jean-Michel » sont montrés dans l'extrait [80].

---

[29] Voir les 19 SN possessifs de la chaîne « Jean-Michel » dans l'*Annexe 20 : SN possessifs français dans la chaîne « Jean-Michel »*.

[80][Il] hésita un instant, puis [Ø] demanda d'un ton plus bas, un peu honteux :
JM-1 *JM-2*

– Est-ce qu'il a… de nouveau ?

– Non, [**père**], non, [**père**], dit précipitamment Louisa.
JM-3-SN-sans-dét *JM-4-SN-sans-dét*

[Le vieux] la regarda ; elle évita [son] regard.
JM-5 *JM-6*

– Ce n'est pas vrai, tu mens.

Elle pleura silencieusement.

– Bon Dieu ! cria [le vieillard]JM-7, en donnant un coup de pied au foyer. Le tisonnier tomba bruyamment. La mère et l'enfant tressaillirent.

– [**Père**], je [vous] en prie, dit Louisa, il va pleurer.
JM-8-sans-dét *JM-9*

D'après l'observation des extraits contenant des SN sans déterminant dans les chaînes « Louisa » (11 occurrences) et « Melchior » (12 occurrences) dans l'*Annexe 21 : SN sans déterminant français dans la chaîne « Louisa »* et l'*Annexe 22 : SN sans déterminant français dans la chaîne « Melchior »*, une grande partie des occurrences sont localisées dans les séquences dialogales : 7 SN sans déterminant sur 11 de la chaîne « Louisa » et 10 sur 13 de la chaîne « Melchior » se trouvent dans des dialogues, avec la fonction syntaxique d'appellatif. L'extrait [80] montre ainsi l'emploi le plus prototypique des SN sans déterminant dans les chaînes secondaires.

En ce qui concerne les occurrences dans les séquences narratives à proprement parler, les SN sans déterminant et avec déterminant réservent respectivement des têtes lexicales spécifiques : les premiers sont constitués de « maman » et « papa » seuls (extrait [81]), alors que les seconds comportent des noms comme « mère » et « père », qui peuvent être utilisés avec des déterminants comme « la mère », « ma mère » ou « cette mère », au lieu d'être appellatifs sans déterminant.

[81]Alors il rit de plus belle, et peut-être même il se force à rire, parce qu'il a un public. [**Maman**] prend
SN-sans-dét

un air sévère, et met un doigt sur sa bouche, pour qu'il ne réveille pas le père ; mais ses yeux fatigués rient malgré elle. Ils chuchotent ensemble… Brusquement, un grognement furieux du père. Ils tressautent tous deux. [**Maman**] tourne précipitamment le dos comme une petite fille coupable,
SN-sans-dét

elle fait semblant de dormir.

6.2.5. SN indéfinis et SN démonstratifs : maillons de quantités non significatives

Les SN indéfinis et les SN démonstratifs sont inclus dans le même groupe parce que ces deux catégories présentent de très petites quantités et des proportions peu significatives pour les trois chaînes. Les SN indéfinis concernent au total 11 maillons[30], dont 6 sont dans la chaîne « Louisa », 2 dans la chaîne « Melchior » et 3 dans la chaîne « Jean-Michel ». Aucun groupe n'atteint 1 % de maillons dans aucune chaîne. Les SN démonstratifs constituent également un très petit groupe de maillons : ils sont absents dans la chaîne « Louisa » et totalisent 2 maillons dans les deux autres chaînes secondaires (extraits [82] et [83]), dont l'un se trouve dans la

[30] Voir les SN indéfinis dans leur contexte dans l'*Annexe 23 : SN indéfinis français dans les chaînes des personnages secondaires.*

chaîne « Melchior », l'autre dans la chaîne « Jean-Michel ». Nous comparerons ces deux groupes avec ceux du chinois pour obtenir des conclusions éventuelles dans le cadre de l'approche franco-chinoise.

[82]Jean-Michel continua d'une voix plus sourde, avec des éclats de colère :
– Qu'ai-je fait au bon Dieu pour avoir [cet ivrogne de fils] ? C'est bien la peine d'avoir vécu comme
M-SN-dém
j'ai vécu, de m'être privé de tout !… Mais toi, toi, tu n'es donc pas capable de l'empêcher ? Car enfin, sacrebleu ! c'est ton rôle. Si tu le retenais au logis !…

[83]Sous sa direction, l'orchestre de la Cour acquit une petite célébrité dans les pays rhénans, où Jean-Michel était devenu légendaire par sa stature athlétique et par ses accès de colère. Il ne pouvait se maîtriser malgré tous ses efforts : car [**cet homme violent**] était au fond timide et craignait de se
JM-SN-dém
compromettre ; il aimait le décorum et redoutait l'opinion.

6.2.6. Redémarrage des chaînes

Les chaînes des personnages secondaires présentent des discontinuités dans certains épisodes du corpus. Ces épisodes sans maillons peuvent être considérés comme de longues distances, après lesquelles les chaînes sont redémarrées. Ce sont des endroits importants dans la structure textuelle puisque les maillons de redémarrage se chargent de réactiver les anciennes informations sur le même référent et d'en ajouter de nouvelles pour la progression du texte. Comme le montre le tableau (6.18), chaque chaîne secondaire présente quelques ruptures dans le corpus : 9 pour la chaîne « Louisa », 12 pour la chaîne « Melchior » et 10 pour la chaîne « Jean-Michel ». Toutes les catégories de maillons ne sont pas observées à la place de redémarrage : les trois catégories préférées des trois chaînes sont les noms propres (14 occurrences), les SN sans déterminant (8 occurrences) et les SN possessifs (6 occurrences).

Tableau (6.18) : Maillons à la place de redémarrage des chaînes des personnages secondaires

Chaîne	« Louisa »	« Melchior »	« Jean-Michel »
Nombre de ruptures	9	12	10
Noms propres	**3**	**9**	**2**
SN défini	1	0	1
SN possessifs	**4**	**1**	**1**
SN démonstratifs	0	0	0
SN indéfinis	0	0	0
SN sans déterminant	**1**	**1**	**6**
Non nominaux	0	1	0

Les catégories préférées pour chaque chaîne varient. Ce sont les noms propres et les SN possessifs qui redémarrent la plupart du temps dans la chaîne « Louisa » ; les noms propres dans la chaîne « Melchior » et les SN sans déterminant dans la chaîne « Jean-Michel ». Ces catégories ne sont pas seulement les préférées dans chaque chaîne au moment du redémarrage : elles présentent également des proportions plus importantes parmi tous les maillons nominaux de chaque chaîne. On trouve toutefois une exception dans la chaîne « Jean-Michel » : les SN définis sont plus fréquemment utilisés et leur proportion dépasse celle des

SN sans déterminant ; mais ce sont ces derniers qui occupent principalement la place de redémarrage de la chaîne.

6.2.7. Comparaison des chaînes des personnages secondaires avec celle du protagoniste

L'observation de ce groupe de trois chaînes nous a conduit à les considérer comme secondaires non seulement en raison de la différence quantitative et de la répartition proportionnelle des maillons par rapport à la chaîne du protagoniste, mais également pour des raisons sémantiques, syntaxiques et discursives.

Certaines catégories de maillons présentent des proportions assez similaires entre les deux groupes de chaînes. Par exemple, le pourcentage de noms propres s'élève à 11 % dans la chaîne du protagoniste et à 12 % en moyenne dans les chaînes secondaires ; c'est donc la catégorie nominale le plus souvent utilisée dans les quatre chaînes. Les SN définis sont aussi proches dans les deux groupes : 3 % dans la chaîne du protagoniste et 5 % dans les chaînes secondaires. Deux groupes de petites quantités sont aussi semblables : les SN indéfinis et les SN démonstratifs n'atteignent pas 1 % dans les deux groupes ; ce sont ainsi les deux catégories qui ne conviennent ni pour le démarrage et le redémarrage de la chaîne, ni pour la maintenance de sa continuité.

D'autres catégories présentent des différences remarquables en fonction des deux groupes : il s'agit des SN possessifs, des SN sans déterminant et des catégories non nominales.

Les SN possessifs reflètent le degré d'association et d'indépendance des personnages relativement aux autres référents. De ce point de vue, le protagoniste est beaucoup moins associé et beaucoup plus indépendant, avec seulement 1 % de SN possessifs (voir le tableau (6.3) en plus haut), par rapport aux personnages secondaires, avec respectivement 12 %, 7 % et 2 %. Plus la proportion de SN possessifs est élevée, plus le référent est associé aux autres référents et dépendant d'eux. Donc, selon cet indice, l'échelle d'association des personnages peut être formulée de la manière suivante, avec une proportion de possessifs ascendante :

Jean-Christophe (1 %) < Jean-Michel (2 %) < Melchior (7 %) < Louisa (12 %)

En ce qui concerne les SN sans déterminant, cette catégorie occupe seulement 0,2 % des maillons du protagoniste, mais 5 % en moyenne des maillons des secondaires. D'après l'observation des occurrences, cette dernière proportion beaucoup plus élevée est due à l'usage des appellatifs dans les séquences dialogales : le protagoniste est plus souvent appelé par son prénom par les autres interlocuteurs ; mais les personnages secondaires apparaissent plus souvent sous forme de SN sans déterminant, comme « maman », « papa » et « grand-père ».

Enfin, les proportions des maillons non nominaux sont très proches dans les chaînes des personnages secondaires : elles s'élèvent respectivement à 68 % dans la chaîne « Louisa », 71 % dans la chaîne « Melchior » et 69 % dans la chaîne « Jean-Michel ». Les chaînes des personnages secondaires et celle du protagoniste sont composées d'une grande moitié de maillons non nominaux ; mais la moyenne des trois chaînes se situe à 69 % : elle est beaucoup moins élevée que la proportion des maillons non nominaux de la chaîne du protagoniste, qui atteint 84 %.

Tableau (6.19) : Comparaison entre la moyenne des chaînes des personnages secondaires et la chaîne du protagoniste

	Protagoniste	Moyenne des trois personnages secondaires
Nbre	4366	738,7
Répartition	98 % (49 épisodes)	58 % (29 épisodes)
Noms propres	11 %	12 %
SN définis	3 %	5 %
SN possessifs	**1 %**	**7 %**
SN sans déterminant	**0,2 %**	**5 %**
SN indéfinis	0,2 %	0,5 %
SN démonstratifs	0,2 %	0,1 %
Non nominales	**84 %**	**69 %**

Après avoir analysé les chaînes du protagoniste et des trois personnages secondaires dans le roman-fleuve, nous poursuivons par l'examen des chaînes des personnages *épisodiques* et *éphémères*.

6.3. Chaînes des personnages *épisodiques* et *éphémères*

Outre le protagoniste et les trois personnages secondaires qui se présentent en chaînes de référence *globales*, le roman-fleuve comporte encore d'autres personnages et d'autres référents non humains dont les expressions constituent des chaînes de référence *locales* ; ces dernières sont subdivisées en chaînes des personnages *épisodiques*, des personnages *éphémères* et des référents non humains. En plus d'avoir la fonction essentielle de maintenir la continuité référentielle du récit, ces chaînes jouent également un rôle dans la divulgation de nouvelles informations. Dans cette partie, nous présenterons les chaînes des personnages épisodiques et éphémères, dont la saillance est inférieure aux personnages principaux et supérieure aux référents non humains. Nous nous demanderons quelles sont leurs particularités par rapport aux chaînes globales dans la continuité référentielle du roman-fleuve.

6.3.1. Des chaînes *épisodiques* aux chaînes *éphémères*

Les chaînes locales des personnages sont subdivisées en *épisodiques* et en *éphémères* en raison des différences de longueur et de répartition, ainsi que de l'importance qu'elles revêtent dans l'intégralité du roman-fleuve.

Les chaînes des personnages épisodiques se situent dans un ou plusieurs épisodes[31] et ne comportent pas forcément un grand nombre de maillons par rapport à celles du protagoniste et des secondaires. La plus longue est celle du personnage nommé « Gottfried »

[31] Voir les détails de ce groupe de chaînes dans l'*Annexe 25 : Maillons français dans les chaînes des personnages épisodiques*.

avec 287 maillons ; la plus courte (10 maillons) est une chaîne de *fusion*[32] de trois personnages déjà mentionnés, « Jean-Christophe & Ernest & Rodolphe ». La distribution des chaînes épisodiques est variable dans le corpus : la plus répandue est celle de la fusion de deux personnages secondaires, « Melchior & Louisa », qui couvre 8 épisodes ; plusieurs chaînes épisodiques se distribuent seulement dans un ou deux épisodes : leurs détails sont présentés dans l'*Annexe 24 : Chaînes des personnages épisodiques en français*[33]. Presque tous les épisodes du corpus comportent une ou deux chaînes de ces personnages ; c'est la raison pour laquelle ces dernières sont appelées *épisodiques*.

À la différence des chaînes épisodiques, celles des personnages *éphémères* présentent des longueurs et des répartitions beaucoup plus réduites. La plupart d'entre elles ne comprennent que 3 ou 4 maillons dans un seul épisode[34] voire deux. Par exemple, la chaîne « Jean-Michel & Louisa » est une fusion de deux personnages secondaires déjà mentionnés : elle apparaît seulement dans l'épisode 1 de la section I et comporte 4 maillons ; celle du nouveau personnage « Ottilie » a 7 maillons et on ne la trouve qu'à l'épisode 11 de la section II. C'est en raison de leur apparition éphémère dans l'intégralité du roman-fleuve que ces chaînes sont ainsi appelées.

Contrairement aux chaînes globales, ces chaînes locales renvoyant à des humains épisodiques et éphémères diffèrent les unes des autres du point de vue de la répartition et de la longueur ; mais lorsqu'il s'agit de la contribution à la continuité textuelle, elles présentent des fonctions plus ou moins similaires : elles garantissent la cohérence locale dans le cadre épisodique ou dans quelques épisodes ; néanmoins, elles paraissent relativement peu importantes en raison de leur répartition locale et peu significatives pour l'intégralité du roman-fleuve.

6.3.2. Chaînes épisodiques et éphémères selon les sous-types sémantiques

L'observation des chaînes obtenues révèle que les personnages épisodiques et éphémères présentent une grande variété sémantique : ils peuvent être classés en cinq sous-types d'humains, comme le montre le tableau (6.20) : « humain générique ou indéfini » (44 %), « personnage spécifique » (31 %), « personnage de fusion » (16 %), « humain collectif » (5 %) et « humain historique » (2 %).

[32] Voir la définition de ce mode de cohabitation des chaînes de Schnedecker (1992, 1997, 2021) dans la section *3.2.2.1. Chaînes de référence dans divers genres en français* du *Chapitre 3*.

[33] Leur répartition étant assez réduite, elle n'est pas montrée sous forme de tableau.

[34] Voir les détails de ce groupe de chaînes dans l'*Annexe 26 : Maillons français dans les chaînes des personnages éphémères*.

Tableau (6.20) : Sous-types d'humains dans les chaînes des personnages épisodiques et éphémères

Type d'humain	Nbre	Pourcentage
Humain générique ou indéfini	44	44 %
Personnage spécifique	31	31 %
Personnage de fusion	16	16 %
Humain collectif	5	5 %
Humain historique	2	2 %

Les chaînes des sous-types d'humains dans ce groupe ne partagent pas de points communs. Celles des humains génériques et indéfinis occupent la plus grande proportion, alors que les humains historiques sont assez marginaux en totalisant trois chaînes. En dehors de ces différences, nous remarquons les propriétés suivantes qui sont manifestées par quelques sous-types d'humains.

6.3.2.1. Désignation par les noms propres des sous-types d'humains

Les humains génériques et indéfinis sont rarement désignés par les noms propres puisqu'il est naturel que les référents comme « les enfants » et « les gens dans l'église » soient évoqués comme tels sans recourir à une dénomination. En revanche, les personnages spécifiques du roman-fleuve sont censés présenter le cas contraire : sur la base des dénominations du protagoniste et des personnages secondaires, d'autres noms propres peuvent enrichir le roman-fleuve. Pourtant, tous les humains spécifiques ne sont pas dénommés par des noms propres. D'après l'observation des chaînes repérées, seuls le sont, totalement ou partiellement, deux types d'humains épisodiques et éphémères : « personnage spécifique » et « humain historique ».

Au sujet des humains historiques, le nom propre semble être le choix le plus univoque pour identifier clairement ces personnages qui sont censés être reconnus par les lecteurs francophones grâce à leurs connaissances encyclopédiques ou à leur culture générale. Comme on peut le voir dans les extraits [84] et [85], « Napoléon » et « Mozart » ne sont pas les seuls humains historiques mentionnés dans le livre : d'autres sont également présents, mais ces derniers ne sont que des singletons qui ne sont pas réévoqués dans la suite du récit.

[84]Il parlait de [**Régulus**], d'[**Arminius**], des chasseurs de [**Lützow**], de [**Kœrner**] et de [**Frédéric Stabs**], celui qui voulait tuer [**l'empereur Napoléon**]. Sa figure rayonnait, en rapportant des traits d'héroïsme inouïs.

[85]Aux murs, des versets encadrés, et de mauvais chromos de [**Mozart**] et de [**Beethoven**]. Un petit piano dans un coin, un violoncelle dans l'autre ; des rayons de livres pêle-mêle, des pipes accrochées, et, sur la fenêtre, des pots de géraniums.

Qu'une expression soit singleton ou maillon de chaîne, tous les humains historiques sont désignés par leurs noms propres français. Nous verrons, dans la section *7.3. Chaînes d'autres humains : épisodiques, éphémères et indéfinis* du *Chapitre 7*, que, dans la version traduite du livre, les mêmes humains historiques ne sont pas tous censés être identifiés par les lecteurs sinophones en raison des écarts culturels entre chaque communauté linguistique ; ces humains peuvent être désignés autrement en chinois.

À la différence des humains historiques, ainsi que du protagoniste et des personnages secondaires qui sont dénommés et mentionnés majoritairement par leur prénom, les personnages épisodiques et éphémères désignés par les noms propres ne concernent que 12 chaînes sur 31. Que ce soit avec ou sans dénomination, la désignation par le nom propre peut être, d'un côté, le choix de narrateur, et, d'un autre côté, un indice qui reflète éventuellement la saillance du personnage en question. En effet, ceux qui sont dénommés par les noms propres sont plus saillants (extrait [86]) que ceux qui le sont de diverses autres façons : le statut social (« grand-duc » et « princesse »), la profession (« voiturier » et « valet », comme le montre l'extrait [87]), le genre et l'âge (« dame » et « garçon ») et la relation familiale ou amicale (« voisine » et « père de Clara »). Ces noms communs au rôle de désignateur indiquent en même temps l'identité qualitative la plus importante de ces personnages.

[86]Excellent musicien, il s'était fait promptement apprécier dans un pays où tous sont musiciens. Il y avait pris racine en épousant, à quarante ans passés, [**Clara Sartorius, la fille du maître de chapelle du prince**]*spécifique-dénommé-1*, qui lui transmit sa charge. [**Clara**]*spécifique-dénommé-2* était une Allemande placide qui avait deux passions : la cuisine et la musique. [**Elle**]*spécifique-dénommé-3* eut pour [**son**]*spécifique-dénommé-4* mari un culte qu'égalait seul celui qu'[**elle**]*spécifique-dénommé-5* avait pour [**son**]*spécifique-dénommé-6* père. Jean-Michel n'admirait pas moins [**sa femme**]*spécifique-dénommé-7*.

[87]Ce jour-là, sa mère lui avait mis ses habits les plus propres, de vieux habits donnés, dont l'ingénieuse patience de Louisa avait su tirer parti. Il alla la rejoindre, comme elle le lui avait dit, dans la maison où elle travaillait. Il était intimidé, à l'idée d'entrer seul. [**Un valet**]*spécifique-non-dénommé-1* flânait sous le porche ; [**il**]*spécifique-non-dénommé-2* arrêta l'enfant et [**Ø**]*spécifique-non-dénommé-3* lui demanda d'un ton protecteur ce qu'il venait faire.

La différence de saillance est aussi reflétée par la longueur des deux chaînes : celle du personnage dénommé par un nom propre de l'extrait [86] est plus longue que celle du personnage non dénommé de l'extrait [87].

6.3.2.2. Introduction et redémarrage

Contrairement à celles du protagoniste et des secondaires, les chaînes épisodiques et éphémères sont courtes et nombreuses, et elles se localisent dans presque tous les épisodes du corpus. Il nous paraît donc intéressant d'étudier les catégories préférées aux places d'introduction et de réactivation de ces chaînes locales des référents moins importants par rapport aux personnages globaux.

Tableau (6.21) : Maillons de démarrage et de redémarrage des chaînes épisodiques et éphémères

Réf	Humain générique ou indéfini	Personnage spécifique	Personnage de fusion	Humain collectif
Noms propres	0	20	3	0
SN définis	8	8	1	6
SN démonstratifs	5	0	0	0
SN possessifs	6	2	10	1
SN indéfinis	15	11	0	1
SN sans déterminant	6	1	1	0
Non nominaux	10	1	21	4
Total	50	43	36	12

Établir des points communs entre chaque type d'humain et entre chaque forme de maillon en les comparant est une tâche difficile car chaque groupe présente de nombreuses propriétés spécifiques. Par exemple, les chaînes des humains génériques et indéfinis sont initiées et réactivées par toutes les catégories de noms communs : ce groupe est désigné selon les identités sociale, professionnelle et familiale, au lieu d'avoir une dénomination en nom propre. En revanche, les noms propres sont la forme dominante pour initier et réactiver les chaînes des personnages spécifiques, à savoir que ces derniers sont désignés par leur prénom ou leur nom de famille afin d'éviter d'éventuelles ambiguïtés suscitées par les noms communs classifiants. Les SN indéfinis sont également fréquemment utilisés dans les chaînes des humains spécifiques ainsi que dans celles des génériques et indéfinis, alors que cette catégorie est rarement repérée dans les chaînes d'autres types d'humains. Les chaînes de fusion sont majoritairement démarrées et redémarrées par les non nominaux et les SN possessifs. Les premiers indiquent que les référents inclus dans ces fusions sont déjà respectivement mentionnés dans le contexte ; les seconds impliquent que l'apparition de ces chaînes est fortement dépendante d'autres référents. Ils établissent une relation d'association et les catégories non nominales sont elles-mêmes des anaphoriques, comme le montre l'exemple de la fusion des deux personnages individuels (« Christophe » et « l'oncle ») en un seul groupe d'individus (« ils ») :

[88][Christophe] revenait, triste que [l'oncle] n'eût pas compris. Mais, comme [**ils**] arrivaient à la
personnage-1 *personnage-2* *fusion-redémarrage*
maison, [Gottfried] [lui] dit : « D'autres soirs, si [tu] veux, [nous] irons encore entendre la musique
personnage-2 *personnage-1* *personnage-1* *fusion-reprise*
du bon Dieu, et [je] [te] chanterai d'autres chansons. » Et quand [Christophe] [l'] embrassa,
personnage-2 *personnage-1* *personnage-1* *personnage-2*
plein de reconnaissance, en [lui] disant bonsoir, [il] vit bien que [l'oncle] avait compris.
personnage-1 *personnage-1* *personnage-2*

Depuis lors, [ils] allaient souvent se promener ensemble, le soir ; et [ils] marchaient sans causer,
fusion-reprise *fusion-reprise*
le long du fleuve, ou à travers les champs.

6.3.3. Composition des chaînes épisodiques et éphémères

Après avoir observé les détails de chaque sous-groupe de chaînes épisodiques et éphémères, nous pouvons proposer un bilan de la composition de ce groupe de chaînes selon

les critères morpho-syntaxiques. Comme le montre le tableau (6.22), certaines formes présentent des points communs dans les deux groupes, comme la proportion de SN définis (identique à 12 %) et celle de SN possessifs (respectivement à 4 % et 5 %). Les SN démonstratifs et les SN sans déterminant sont rarement utilisés dans ces chaînes ; ce sont deux formes qui ne sont pas censées garantir la continuité référentielle.

Tableau (6.22) : Composants des chaînes épisodiques et éphémères

	Chaînes épisodiques (1361 maillons)	Chaînes éphémères (242 maillons)
Noms propres	12 %	5 %
SN définis	12 %	12 %
SN démonstratifs	0,6 %	2 %
SN possessifs	4 %	5 %
SN indéfinis	3 %	14 %
SN sans déterminant	1 %	2 %
Zéro	7 %	9 %
Autres non nominaux	60 %	51 %

Les différences entre les chaînes épisodiques et les éphémères sont constatables dans quatre groupes de composants : les noms propres, les SN indéfinis, les catégories non nominales explicites et les zéros. Les noms propres sont plus couramment utilisés dans les chaînes épisodiques (12 %) que dans les éphémères (5 %) : les personnages épisodiques sont plus probablement dénommés par des noms propres en comparaison des personnages éphémères. Les SN indéfinis sont peu récurrents dans les chaînes épisodiques (3 %) mais beaucoup plus réguliers dans les éphémères (14 %). Les maillons non nominaux présentent une grande moitié dans les deux groupes de chaînes, respectivement à 67 % pour les épisodiques et à 60 % pour les éphémères. Pourtant, les zéros sont plus régulièrement utilisés dans les éphémères (9 %) que dans les épisodiques (7 %), cela pour trois raisons éventuelles. Premièrement, les zéros sont censées exprimer les référents d'un très haut degré de saillance ; mais cela ne semble pas suffire à expliquer cette différence puisque les personnages éphémères ne présentent pas de plus haut degré de saillance par rapport aux épisodiques. Deuxièmement, les zéros sont utilisés régulièrement après une toute petite distance de leur antécédent, ce dernier se situant fréquemment dans la même phrase que celle des premiers, qui sont préférables aux formes explicites ; cela explique la haute proportion de zéros dans les chaînes éphémères. Troisièmement, les zéros sont aussi utilisés dans un contexte mono-référentiel lorsqu'il n'y a pas de référents concurrents dans le contexte proche qui susciteraient des risques d'ambiguïté.

6.3.4. Chaînes vraisemblables de « on » : référents équivoques

La référence de « on » est complexe. Il nous est fréquemment arrivé pendant l'annotation de notre corpus que « on » soit annoté pour la première fois d'une manière intuitive, et d'une autre manière lors de la rectification de l'annotation. En effet, l'antécédent de « on » est très difficile à identifier, en raison, d'un côté, des nombreux référents auxquels ce pronom peut renvoyer, et, d'un autre côté, de l'influence des éléments contextuels qui permettent le cas échéant diverses interprétations référentielles acceptables. Nous nous

demandons ainsi de quelle façon il est possible de fixer le référent de « on » et sous quelle condition il peut être considéré comme constituant des chaînes des référents univoques. L'essentiel pour nous est d'obtenir des facteurs réguliers et éventuels dans notre corpus qui peuvent servir à interpréter le référent de « on » de manière univoque ou équivoque.

Avant de parler des chaînes « on », il convient de se demander si les pronoms « on » peuvent être coréférentiels entre eux. Cette question comporte trois aspects. Premièrement, « on » est défini comme un pronom personnel puisque son référent est toujours humain. Deuxièmement, « on » est un pronom indéfini puisque l'on ne sait pas combien d'individus sont impliqués. Cette propriété permet à ce pronom de renvoyer à des *groupes flous* (*cf.* Landragin & Tanguy, 2014), c'est-à-dire que l'on ne sait pas quels individus font partie de ces groupes et quels individus ne le font pas.

> *L'important, c'est, d'une part, qu'il peut jouer un rôle de pronom personnel, valant pour je, tu, il, nous, vous, ils, d'autre part, qu'il peut jouer un rôle de pronom indéfini, valant soit pour quelqu'un, des gens (emploi épisodique), soit pour chacun, les gens (emploi générique), et ce, en suivant un ensemble de contraintes syntaxiques et sémantiques qui lui sont propres. (Landragin & Tanguy, 2014 : 100)*

Troisièmement, les choses nous semblent encore plus compliquées lorsqu'il y a des « on » qui se succèdent en constituant selon toute vraisemblance des chaînes. En réalité, « on » est assez récurrent dans notre corpus : en plus de ceux qui sont composants des chaînes déjà présentées, nous avons récupéré 199 occurrences de « on » et ses variants répartis dans 27 épisodes (voir l'*Annexe 27 : Succession des pronoms « on » en français*).

La situation la plus évidente pour saisir le référent de « on » est quand le contexte contient un antécédent univoque. Par exemple, dans les deux extraits [89] et [91] suivants, un individu spécifique, « sa mère », et un groupe d'individus « du monde », sont mentionnés dans le contexte proche. Ils servent donc d'antécédents des « on » qui suivent. Cette coréférence peut être affirmée par le test de remplacement de « on » par « sa mère » et « les gens dans le jardin », comme le montrent les tests [91]* et [92]* ; les autres remplacements ne sont pas pertinents pour ces extraits.

[89]Certains jours, il profite de ce que [sa mère] a le dos tourné, pour sortir de la maison. D'abord, [**on**] court après lui, [**on**] le rattrape. Puis, [**on**] s'habitue à le laisser aller seul, pourvu qu'il ne s'éloigne pas trop.

[90]*Certains jours, il profite de ce que [sa mère] a le dos tourné, pour sortir de la maison. D'abord, [**sa mère**] court après lui, [**elle**] le rattrape. Puis, [**elle**] s'habitue à le laisser aller seul, pourvu qu'il ne s'éloigne pas trop.

[91]Un homme, qui passait sur la route, arrêta l'enfant. Christophe était hors d'état de parler ; il montra la maison ; l'homme y entra, et Christophe le suivit. D'autres avaient entendu ses cris et arrivaient des maisons voisines. Bientôt le jardin fut plein [de monde]. [**On**] marchait sur les fleurs, [**on**] se penchait autour du vieux, [**on**] parlait tous à la fois.

[92]*… Bientôt le jardin fut plein [de monde]. [**Les gens dans le jardin**] marchait sur les fleurs, [**ils**] se penchaient autour du vieux, [**ils**] parlaient tous à la fois.

Pourtant, il est aussi possible que le contexte contienne un antécédent ; mais le référent de « on » est toujours équivoque : « on » peut être interprété comme coréférentiel de cet antécédent ou comme « les gens génériques », cela étant le cas de l'extrait [93], avec lequel les deux tests de remplacement [94]* et [95]*, « Christophe » ou par « les gens / les humains », semblent acceptables :

[93][Christophe] réfléchissait : il pensait que quand [**on**] est grand, [**on**] ne s'étonne plus de rien, [**on**] est fort, [**on**] connaît tout.

[94]*Christophe réfléchissait : il pensait que quand [**il**] est grand, [**il**] ne s'étonne plus de rien, [**il**] est fort, [**il**] connaît tout.

[95]*Christophe réfléchissait : il pensait que quand [**les gens**] sont grands, [**ils**] ne s'étonnent plus de rien, [**ils**] sont forts, [**ils**] connaissent tout.

Vu les trois extraits [89], [91] et [93] susmentionnés, il semble que la présence explicite de l'antécédent dans le contexte peut servir de référent univoque ou d'un des référents équivoques : il ne conditionne pas l'interprétation unique de « on ».

Une situation encore moins évidente est celle où on ne trouve aucun antécédent. Cela est le cas des deux extraits [96] et [98] suivants, où « on » peut avoir une des trois interprétations dans chaque succession d'occurrences :

i) les gens génériques, le monde, qu'il est possible de remplacer par « les gens » ou « tout le monde » ;

ii) un groupe d'individus de nombre indéfini, on-ne-sait-pas-qui, qu'on peut remplacer par « des gens », « certaines personnes », « quelqu'un » ;

iii) un seul individu, un individu quelconque, on-ne-sait-pas-qui, possiblement remplaçables par « quelqu'un ».

Selon l'acceptabilité des tests de remplacements, il est possible de fixer une interprétation plus appropriée et pertinente.

[96]Plusieurs fois, [**on**] avait tenté de le fixer quelque part, de lui acheter un petit fonds, un bazar, une mercerie. Mais il ne pouvait s'y faire : une nuit il se levait, mettait la clef sous la porte, et repartait avec son ballot. [**On**] restait des mois sans le voir. Puis il reparaissait : un soir, [**on**] entendait gratter à l'entrée ; la porte s'entre-bâillait, et la petite tête chauve, poliment découverte, se montrait avec ses bons yeux et son sourire timide.

[97]*Plusieurs fois, [**les gens du village**] avaient tenté de le fixer quelque part, de lui acheter un petit fonds, un bazar, une mercerie. Mais il ne pouvait s'y faire : une nuit il se levait, mettait la clef sous la porte, et repartait avec son ballot. [**Les gens du village**] restait des mois sans le voir. Puis il reparaissait : un soir, [**les gens du village**] entendait gratter à l'entrée ; la porte s'entre-bâillait, et la petite tête chauve, poliment découverte, se montrait avec ses bons yeux et son sourire timide.

[98]En ce sens, [**on**] peut dire que la misère est un maître, non seulement de pensée, mais de style ; elle apprend la sobriété à l'esprit, comme au corps. Quand le temps est compté et les paroles mesurées, [**on**] ne dit rien de trop et [**on**] prend l'habitude de ne penser que l'essentiel. Ainsi [**on**] vit double, ayant moins de temps pour vivre.

[99]*En ce sens, [**tout le monde**] peut dire que la misère est un maître, non seulement de pensée, mais de style ; elle apprend la sobriété à l'esprit, comme au corps. Quand le temps est compté et les paroles mesurées, [**tout le monde**] ne dit rien de trop et [**on**] prend l'habitude de ne penser que l'essentiel. Ainsi [**tout le monde**] vit double, ayant moins de temps pour vivre.

Dans la dernière situation, dont la complication est encore plus grande, les « on » d'une succession ne sont pas coréférentiels, puisque tous les « on » ne peuvent pas être remplacés par la même expression :

[100] Mais il n'était pas convaincu ; il ne se rassurait que quand les musiciens s'installaient. Encore craignait-il jusqu'au dernier moment que le rideau se levât, et que [**l'on**] annonçât, comme [**on**] le fit un soir, un changement de spectacle. […] [**On**] s'agitait derrière le rideau, [**on**] entendait un bruit de voix et de pas précipités. C'était un accident, un malheur imprévu ?… Le silence se rétablissait. Le chef d'orchestre était à son poste. Tout semblait enfin prêt… [**On**] ne commençait pas !

[101] *Mais il n'était pas convaincu ; il ne se rassurait que quand les musiciens s'installaient. Encore craignait-il jusqu'au dernier moment que le rideau se levât, et que [**quelqu'un**] annonçât, comme [**il**] le fit un soir, un changement de spectacle. […] [**Les musiciens**] s'agitaient derrière le rideau, [**les gens / Christophe**] entendaient un bruit de voix et de pas précipités. C'était un accident, un malheur imprévu ?… Le silence se rétablissait. Le chef d'orchestre était à son poste. Tout semblait enfin prêt… [**Les musiciens**] ne commençaient pas !

Par conséquent, toutes les suites de « on » que nous avons récupérées dans l'*Annexe 27* ne constituent pas de vraies chaînes en raison de leurs interprétations équivoques et de leur relation non coréférentielle entre eux.

6.4. Chaînes des référents *non humains*

Dans les annotations des types sémantiques des référents, nous avons choisi le terme *humain* pour désigner l'ensemble de tous les types d'humains mentionnés dans le roman-fleuve. Bien que le nombre d'expressions référentielles des humains s'élève à 84 % dans l'intégralité du corpus en français, les 16 % d'expressions renvoyant aux référents *non humains* ne sont pas négligeables et présentent des particularités bien distinctes de celles des premières, selon deux aspects. D'une part, ces expressions désignent des référents d'une grande diversité de catégories sémantiques. Comme le montre le tableau (6.1) au début de ce chapitre, 2 % renvoient aux lieux, 9 % aux autres référents concrets, 0,7 % aux temps et 5 % aux autres référents abstraits. On y trouve des référents de classes génériques et des référents individuels spécifiques ; ces référents peuvent exister dans le monde extra-linguistique ou seulement dans le cadre fictif du roman-fleuve ; ils peuvent être autonomes ou étroitement liés aux autres référents. D'autre part, la plupart des expressions de ces référents ne constituent pas des chaînes : elles sont des singletons qui se répartissent de manière éparse dans tout le corpus. Le nombre de chaînes des référents non humains est ainsi beaucoup plus réduit[35] ; malgré leur faible nombre, nous cherchons à synthétiser quelques particularités.

6.4.1. Chaînes du « fleuve » et de « la musique »

L'observation de ce groupe de chaînes des référents non humains nous conduit à remarquer deux (groupes de) chaînes qui nous semblent particulièrement significatifs et importants du point de vue de la continuité textuelle : la chaîne du fleuve « Rhin » et les chaînes de « la musique » et des notions musicales.

- Chaîne du fleuve « Rhin »

Le Rhin est un référent spécifique dans le monde réel. Son importance dans *Jean-Christophe* est commentée par Leblond (2010) selon l'approche littéraire : en tant que fleuve, la dénomination du genre *roman-fleuve* a vu le jour grâce à la parution de ce livre. Le Rhin est considéré comme une métaphore de l'écoulement temporel de la vie du protagoniste :

> *Reflet poétique de la durée, le fleuve représente la conscience ou l'âme, en termes plus rollandiens. Alors que le Rhin est le compagnon de Christophe, les deux cours se déroulant en parallèle et se croisant périodiquement pour se rejoindre à la fin, la rivière est une donnée immanente à l'héroïne féminine de Romain Rolland. L'eau courante est liée à une définition quasi mystique de la féminité, qui ne se*

[35] Voir les détails dans l'*Annexe 28 : Chaînes des référents non humains en français.*

cantonne pas à la maternité, mais met en scène passions et révolte contre un ordre social oppresseur, sur un mode résolument épidictique. (Leblond, 2010 : 456)

L'importance du fleuve se manifeste par la longueur et la distribution de sa chaîne : elle totalise 50 maillons répartis dans 5 épisodes. Malgré cette longueur peu significative par rapport à celle des chaînes des personnages principaux, elle se situe à la première place de toutes celles des non humains.

- Chaînes de « la musique » et des notions qui y sont relatives

Sachant que l'histoire de l'œuvre se déroule au gré de la vie d'un musicien, la « musique » et les notions musicales constituent le deuxième (groupe de) référents non humains importants, comme le remarque Leblond (2010) :

> *Jean-Christophe fait ainsi cohabiter lecture musicale et lecture verbale. L'irruption ponctuelle du fragment de partition fait signe vers une lecture musicale élargie au roman. Rolland avertit son lecteur à l'ouverture de « ce vaste poème en prose qui ne tenait aucun compte des obstacles matériels et brisait délibérément avec toutes les conventions admises dans le monde littéraire français » (*Jean-Christophe*, p. XIV.). Il exige ainsi la lecture tabulaire et globalisante que requiert le poème. La lecture du roman-fleuve sera double, de même que l'écoute linéaire d'une œuvre musicale peut être doublée par la lecture de la partition. (Leblond, 2010 : 510)*

À la différence de ceux de la chaîne « Rhin », dont toutes les occurrences sont regroupées dans quelques épisodes, les maillons de la chaîne du référent générique « musique » sont encore plus épars et moins nombreux : 50 au total, localisés dans 7 épisodes, et constituant ainsi la deuxième plus longue chaîne des référents non humains. Malgré cette longueur, ces maillons n'apparaissent jamais comme topiques discursifs locaux. La distance inter-maillonnaire est très importante : les critères employés pour celles des humains permettent à peine de définir cette chaîne comme telle, puisque la quasi-totalité des maillons sont utilisés pour la redémarrer après une grande distance de plus de 10 propositions :

[102] Ce n'est pas, comme ces sages paroles auraient pu le faire croire, que Melchior se préoccupât de défendre l'enfant contre l'exaltation dangereuse d'un orgueil prématuré. Il devait se charger de démontrer promptement le contraire. Mais, n'ayant jamais eu lui-même aucune idée à exprimer en [**musique**], ni le moindre besoin d'en exprimer aucune, il en était arrivé, dans son infatuation de virtuose, à considérer la composition comme une chose secondaire, à laquelle l'art de l'exécutant donnait seul tout son prix. Il n'était certes pas insensible aux enthousiasmes suscités par les grands compositeurs, comme Hassler ; il avait pour ces ovations le respect qu'il éprouvait toujours pour le succès, – mêlé secrètement d'un peu de jalousie, car il lui semblait que ces applaudissements lui étaient dérobés. Mais il savait par expérience que les succès des grands virtuoses ne sont pas moins bruyants, qu'ils sont même plus personnels et plus fertiles en conséquences agréables et flatteuses. Il affectait de rendre un profond hommage au génie des maîtres musiciens ; mais il avait plaisir à raconter d'eux des anecdotes ridicules, qui donnaient de leur intelligence et de leurs mœurs une triste opinion. Il plaçait le virtuose au sommet de l'échelle artistique : car, disait-il, il est bien connu que la langue est la plus noble partie du corps ; et que serait la pensée sans la parole ? que serait [**la musique**] sans l'exécutant ?

Un groupe de référents génériques et spécifiques entourent celui-ci : des musiques, des voix, des chansons, des sons, des bruits, des œuvres, des compositions et des airs.

[103] Mais le suprême bonheur est lorsque maman vient, qu'elle prend la main de Christophe assoupi, et que, penchée sur lui, à sa demande, elle chante à mi-voix [**une vieille chanson, dont les mots ne veulent rien dire**]. Le père trouve [**cette musique**] stupide ; mais Christophe ne s'[**en**] lasse pas. Il retient son souffle ; il a envie de rire et de pleurer ; son cœur est ivre.

Plus que la « musique » générique, c'est ce groupe de notions musicales qui est fondamental dans le livre.

6.4.2. Composants selon la catégorie grammaticale

D'après la longueur et la répartition des chaînes non humaines, les référents concernés sont moins saillants que les humains. Pourtant, le nombre de maillons n'est pas le seul facteur décisif en ce qui concerne la saillance référentielle. On peut donc s'attendre à ce que ces chaînes soient composées de formes moins marquées. Selon le recensement des composants des chaînes non humaines, comme le montre le tableau (6.23), d'autres facteurs influencent éventuellement leur composition.

Tableau (6.23) : Composants des chaînes des référents non humains

Nom de chaîne (nbre de maillons)	« Rhin » (46)	« Musique » (40)	Lieux (85)	Nocturnes (37)	Parties du corps (143)
Noms propres	4	0	0	0	0
SN définis	21	14	58	29	76
SN possessifs	0	0	12	0	32
SN démonstratifs	0	1	0	0	4
SN indéfinis	0	6	1	2	11
SN sans déterminant	0	9	4	2	3
Non nominaux	21	10	10	4	17

De manière générale, toutes les catégories de maillons sont présentes dans les chaînes des référents non humains ; cependant, les proportions sont fortement différentes les unes des autres : ce n'est pas en raison de leur catégorie sémantique non humaine que ces chaînes partagent des points communs. Excepté la chaîne « Rhin », aucun référent n'est désigné par des noms propres, qui ne sont pas non plus la catégorie préférée de celle-ci. Les autres composants principaux sont les SN définis et les maillons non nominaux : au contraire de ces derniers, les premiers sont aussi les composants qui dominent toutes les autres chaînes non humaines. Les SN possessifs n'apparaissent que dans les chaînes des lieux et des parties du corps : ces référents sont *a priori* associés aux autres référents en ce qui concerne la relation possessive ou inaliénable, ce qui indique que ceux-ci sont ainsi relativement autonomes. Les SN démonstratifs sont en général les moins utilisés dans toutes les chaînes non humaines, tout comme dans celles des humains, car cette catégorie n'est pas apte à garantir la continuité des chaînes de référence. Les SN indéfinis et les SN sans déterminant présentent quelques occurrences dans les chaînes des notions musicales, des repères spatio-temporels et des parties du corps : ils sont normalement utilisés dans trois cas : i) pour introduire des référents inconnus ; ii) comme fonction syntaxique des compléments du nom et iii) pour désigner le nombre indéterminé des référents.

Nous avons donc présenté et analysé les chaînes de toutes les catégories de référents, ce qui nous a permis d'observer des performances très différentes les unes des autres dans le cadre d'un roman-fleuve. Dans la prochaine partie, nous proposerons une synthèse de toutes ces chaînes en français comme résultat et bilan de notre étude de ce long texte narratif.

6.5. Bilan

Dans ce chapitre, nous avons essayé d'analyser, de manière quantitative et qualitative, les chaînes de référence significatives qui jouent un rôle primordial dans la continuité textuelle de *Jean-Christophe*. Les analyses ont été effectuées à deux niveaux : au niveau global en ce qui concerne la réalisation des chaînes dans l'intégralité du corpus ; au niveau local afin de synthétiser leurs spécificités dans chaque type de séquences textuelles. La prise en considération de ce dernier niveau est aussi importante que celle du premier étant donné qu'il est difficile d'étudier toutes les chaînes ayant des fins objectives dans l'intégralité d'un roman-fleuve de dix volumes.

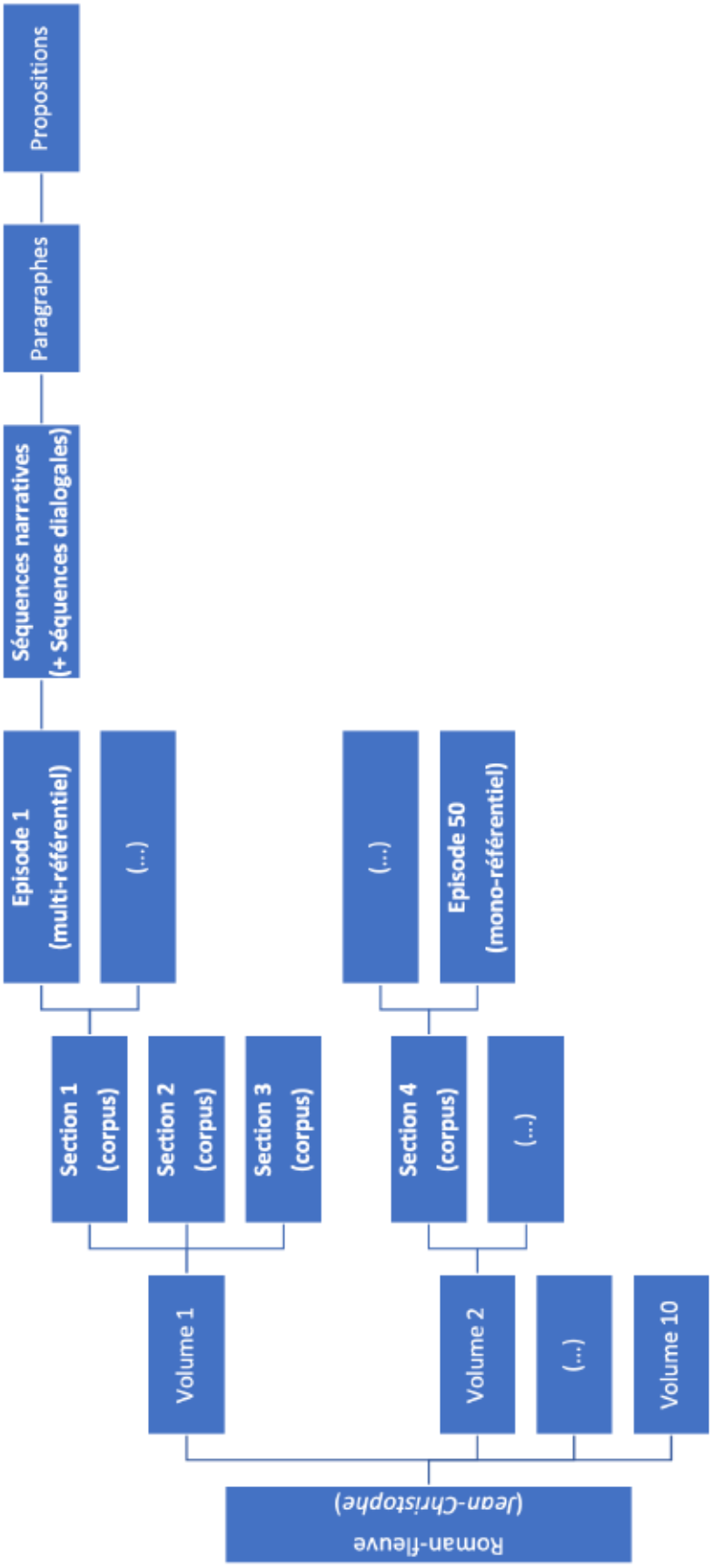
Le niveau local peut être défini selon plusieurs critères. Premièrement, nous avons découpé les quatre premières sections de l'œuvre en 50 épisodes. Ces derniers constituent des unités textuelles supérieures aux paragraphes et aux propositions, mais inférieures aux sections, aux volumes et au roman-fleuve lui-même. La hiérarchie de ces unités textuelles est visible dans la figure (6.2) de plus haut ; la taille de ces épisodes a beaucoup facilité le dépouillement des chaînes de référence par les outils informatiques. Deuxièmement, ce roman-fleuve est, à strictement parler, un composite de séquences narratives et de séquences dialogales représentées par des tours de parole en forme de discours rapporté direct ; ces dernières sont souvent repérées à l'intérieur des épisodes. Troisièmement, les épisodes du roman-fleuve peuvent être définis comme séquences mono-référentielles et multi-référentielles selon le nombre de référents saillants, bien que le genre de roman-fleuve soit considéré comme multi-référentiel dans son intégralité.

Dans la structure textuelle de ce roman-fleuve, les chaînes de référence sont considérées comme des moyens essentiels à la cohérence textuelle à tous niveaux : les chaînes globales, comme celles du protagoniste et des secondaires, permettent la cohésion de toutes les unités textuelles, des volumes aux propositions ; les abondantes chaînes locales, celles des personnages épisodiques et éphémères et des référents non humains qui ne se localisent pas dans toutes les sections ni dans tous les épisodes, garantissent l'unité des petites entités textuelles, comme les épisodes, les paragraphes et les propositions.

En fin de compte, à partir de ces deux niveaux, global et local, nous proposons un bilan des régularités et des spécificités des chaînes en français dans ce long texte narratif appartenant au genre de roman-fleuve selon diverses approches.

À partir de la figure (5.2) du *Chapitre 5*, nous essayons d'affiner les caractéristiques de chaque unité textuelle du corpus *Jean-Christophe* par la figure (6.4) suivante. Selon les repères typographiques, le roman-fleuve est divisé en volumes et en sections. Le premier volume et le premier tiers du deuxième volume sont découpés par nous en épisodes. Selon le nombre de référents concurrents, chaque épisode peut être qualifié comme mono-référentiel ou multi-référentiel ; selon les constituants, chaque épisode peut être constitué par des séquences narratives et des séquences dialogales. Ce sont ainsi ces unités textuelles qui nous permettent d'analyser les chaînes de référence selon les spécificités du contexte.

Figure (6.3) : Caractéristiques des unités textuelles du corpus *Jean-Christophe*



6.5.1. Chaînes de référence au niveau global du corpus : catégorie sémantique, répartition, longueur et composants

Les humains sont censés occuper, dans les textes narratifs, la place centrale : leur chaîne est ainsi logiquement saillante à la fois au niveau global et local. Cependant, les autres référents accessoires qui entourent la vie des humains principaux peuvent également jouer un rôle plus ou moins crucial dans le déroulement du récit. C'est pour cette raison que les chaînes de tous les types de référents sont regroupées et analysées selon l'ordre de leur importance : celle du protagoniste à la première place, celles des personnages secondaires à la deuxième, celles des personnages épisodiques et éphémères à la troisième et enfin celles des référents non humains à la dernière. D'après les recensements des chaînes françaises dans ce chapitre, quatre critères reflètent primordialement la saillance du référent, comme le montre le tableau (6.24) : la catégorie sémantique de celui-ci, la continuité de la chaîne, sa longueur ainsi que sa composition. En fonction des variations de ces quatre critères, le degré d'importance des référents diminue ou augmente.

Tableau (6.24) : Synthèse de quatre groupes de chaînes de référence

	Catégorie	Répartition	Longueur	Composition dominante
Protagoniste	Humains	98 %	4366	84 % non nominaux 11 % noms propres
Secondaires	Humains	58 %	500 – 950	69 % non nominaux 12 % noms propres
Épisodiques et éphémères	Humains	2 – 20 %	10 – 300	66 % non nominaux 12 % noms propres
Non humains	Non humains	2 – 10 %	3 – 50	18 % non nominaux 56 % SN définis

La répartition et la longueur des chaînes confirment que les humains sont les référents primordiaux par rapport aux référents non humains dans ce roman-fleuve ; qui plus est, les différences sont très importantes entre les groupes humains et non humains. De manière générale, les chaînes du protagoniste et des personnages secondaires présentent des répartitions dans la globalité du corpus, c'est-à-dire dans toutes les sections et dans une grande partie des épisodes (en moyenne 58 %). Les quelques ruptures qui surviennent au milieu du roman-fleuve n'empêchent pas leur rôle dominant dans la continuité textuelle globale, puisque ces personnages sont toujours récurrents après les ruptures en question, même si quelques distances dépassent plus de 10 ou 20 propositions, ou même quelques épisodes. Malgré la dénomination par le nom propre « Jean-Christophe » à partir du titre du livre, la chaîne du protagoniste ne peut pas garantir la continuité et la progression du roman-fleuve à elle seule, mais nécessite l'accompagnement d'autres chaînes globales de personnages secondaires. En revanche, à cause de leurs toutes petites distributions et de leurs longueurs, les chaînes locales des personnages épisodiques et éphémères ainsi que celles des référents non humains jouent un rôle peu significatif pour la continuité globale de l'œuvre.

Comme le montrent de nombreuses études, entre autres cognitives, la saillance des référents est étroitement liée à la forme morpho-syntaxique de l'expression[36]. Ariel (1990) et Cornish (1999), notamment, indiquent que l'usage du pronom personnel de troisième personne suppose un haut degré de saillance du personnage dans la représentation mentale du locuteur. La proportion de ce pronom diminue donc théoriquement entre les quatre groupes de chaînes. Cette régularité est confirmée par nos chiffres : les maillons non nominaux, dont une partie majoritaire concerne les pronoms personnels de troisième personne, présentent des proportions en ordre descendant du protagoniste (84 %), en passant par les personnages secondaires (69 %), les personnages épisodiques et éphémères (66 %) jusqu'aux référents non humains (18 %). L'écart entre les groupes des humains et des non humains est particulièrement significatif : cela confirme les degrés de saillance complètement différents de ces référents dans le livre. D'un autre côté, la catégorie nominale privilégiée de chaque groupe de chaînes est aussi recensée comme présentant de grandes différences. Pour les trois groupes d'humains, ce sont toujours les noms propres qui dépassent largement les autres catégories nominales : les humains tendent à être dénommés et ensuite repris directement par leur nom et leur prénom. À l'inverse, ce sont les SN définis qui occupent la plupart des maillons des chaînes de référents non humains, sans dénomination spécifique par un nom propre. Bref, les chaînes des humains et des non humains ne se réalisent pas du tout de la même manière dans ce roman-fleuve.

6.5.2. Chaînes de référence au premier niveau local : la première chaîne et la chaîne la plus longue dans chaque épisode

Bien que la chaîne du protagoniste « Jean-Christophe » soit la première chaîne d'humain qui apparaît dans le livre et bien qu'elle soit la chaîne la plus longue au niveau global, au niveau local cette chaîne n'est pas toujours dominante dans tous les épisodes ; dans certains d'entre eux prédominent en effet d'autres référents humains et non humains.

Dans les épisodes où les maillons de « Jean-Christophe » sont peu nombreux, il existe souvent une ou plusieurs chaînes d'autres référents qui comportent un grand nombre de maillons. Par exemple, selon les chiffres montrés dans le tableau (6.2), la chaîne du protagoniste est très courte, avec 3 maillons, ou absente, dans les épisodes 2 et 11, où les autres personnages sont devenus localement les principaux : leur nombre de maillons dépasse largement celui de la chaîne « Jean-Christophe ». Ainsi, la chaîne de « Melchior » est dominante dans l'épisode 2 avec 56 maillons ; celle de « Jean-Michel » est la plus saillante dans l'épisode 11 avec 118 maillons. Excepté cette observation quantitative relative au nombre de maillons, deux chaînes de chaque épisode attirent particulièrement notre attention : il semble qu'il existe une coïncidence entre la première chaîne et la chaîne la plus longue, à savoir la chaîne dominante dans chaque épisode. Pour confirmer notre hypothèse, ces deux chaînes sont recensées, avec leur nom et leur nombre de maillons, dans le tableau (6.25).

[36] *Cf.* Chastain (1975), Corblin (1995), Schnedecker (2005), Landragin (2011), Schnedecker & Longo (2012) et Glikman *et al.* (2014), entre autres.

Tableau (6.25) : Coïncidence entre la première chaîne, la chaîne du protagoniste, et la chaîne la plus longue dans chaque épisode

S	E	Premier personnage (nbre de maillons)	Personnage de la chaîne la plus longue (nbre de maillons)
I	1	« Jean-Christophe » (49)	« Louisa » (91)
	2	« Louisa » (28)	« Melchior » (55)
	3	« Louisa » (17)	« Jean-Christophe » (3)
	4	**« Jean-Christophe » (21)**	**« Jean-Christophe » (21)**
	5	**« Jean-Christophe » (84)**	**« Jean-Christophe » (84)**
	6	**« Jean-Christophe » (63)**	**« Jean-Christophe » (63)**
	7	**« Jean-Christophe » (83)**	**« Jean-Christophe » (83)**
	8	**« Jean-Michel » (87)**	**« Jean-Michel » (87)**
	9	« Jean-Michel » (33)	« Jean-Christophe » (92)
	10	**« Jean-Christophe » (47)**	**« Jean-Christophe » (47)**
II	11	**« Jean-Michel » (118)**	**« Jean-Michel » (118)**
	12	**« Jean-Christophe » (56)**	**« Jean-Christophe » (56)**
	13	« Louisa » (42)	« Jean-Christophe » (211)
	14	**« Jean-Christophe » (147)**	**« Jean-Christophe » (147)**
	15	**« Jean-Christophe » (70)**	**« Jean-Christophe » (70)**
	16	**« Jean-Christophe » (128)**	**« Jean-Christophe » (128)**
	17	**« Jean-Christophe » (55)**	**« Jean-Christophe » (55)**
	18	**« Jean-Christophe » (172)**	**« Jean-Christophe » (172)**
	19	**« Jean-Christophe » (75)**	**« Jean-Christophe » (75)**
	20	« Fritz » (8)	« Jean-Christophe » (121)
	21	« Jean-Michel » (6)	« Jean-Christophe » (109)
	22	« Melchior » (19)	« Jean-Christophe » (128)
	23	**« Jean-Christophe » (114)**	**« Jean-Christophe » (114)**
	24	« On » (5)	« Jean-Christophe » (44)
III	25	**« Jean-Christophe » (110)**	**« Jean-Christophe » (110)**
	26	« Jean-Michel & Jean-Christophe » (10)	« Jean-Christophe » (64)
	27	**« Jean-Christophe » (43)**	**« Jean-Christophe » (43)**
	28	« Jean-Michel » (16)	« Jean-Christophe » (85)
	29	« Hassler » (3)	« Jean-Christophe » (45)
	30	« Jean-Michel » (77)	« Jean-Christophe » (143)
	31	**« Jean-Christophe » (34)**	**« Jean-Christophe » (34)**
	32	« Gottfried » (100)	« Jean-Christophe » (124)
	33	« Gottfried » (57)	« Jean-Christophe » (107)
	34	**« Jean-Christophe » (39)**	**« Jean-Christophe » (39)**
	35	« Coiffeur » (3)	« Jean-Christophe » (219)
	36	« Melchior » (12)	« Jean-Christophe » (99)
IV	37	**« Jean-Christophe » (107)**	**« Jean-Christophe » (107)**
	38	**« Jean-Christophe » (49)**	**« Jean-Christophe » (49)**
	39	**« Jean-Christophe » (45)**	**« Jean-Christophe » (45)**
	40	**« Jean-Christophe » (50)**	**« Jean-Christophe » (50)**
	41	« Melchior » (18)	« Jean-Michel » (128)
	42	« Jean-Michel » (74)	« Jean-Christophe » (147)
	43	« Jean-Michel » (5)	« Jean-Christophe » (28)
	44	« Jean-Christophe » (55)	« Melchior » (75)
	45	**« Jean-Christophe » (78)**	**« Jean-Christophe » (78)**
	46	« Louisa » (13)	« Jean-Christophe » (133)
	47	« Melchior » (43)	« Jean-Christophe » (95)
	48	**« Jean-Christophe » (76)**	**« Jean-Christophe » (76)**
	49	« Ernst & Rodolphe » (19)	« Jean-Christophe » (106)
	50	**« Jean-Christophe » (80)**	**« Jean-Christophe » (80)**

Total	« Jean-Christophe » : 27 (54 %)	« Jean-Christophe » : 44 (88 %)

Il n'est pas surprenant de constater que la chaîne du protagoniste est récurrente pour les deux types, au vu de sa longueur de 4366 maillons, soit en moyenne 1091,5 maillons par section et 87,3 par épisode. Le protagoniste apparaît comme premier personnage dans 27 épisodes sur 50, ce qui équivaut à 54 % du corpus ; sa chaîne est la plus longue dans 44 épisodes sur 50, soit 86 % du corpus. Le premier personnage et le personnage dominant coïncident dans 26 épisodes (mis en gras dans le tableau (6.25)), à savoir 52 % du corpus, alors que ce personnage d'épisode dominant n'est pas forcément le protagoniste du livre. Selon ces chiffres, il semble que l'importance du nombre de maillons de la chaîne dépasse celle de l'ordre d'apparition dans le texte.

En ce qui concerne la place du premier personnage, le premier maillon ne tarde pas à faire son apparition au début de chaque épisode : il apparaît souvent au moment de la fixation du cadre spatio-temporel, c'est-à-dire que son occurrence coïncide avec des expressions de lieu et de temps. Par exemple, dans les épisodes où le protagoniste est le premier personnage, la première occurrence est repérable dès les premières lignes de l'épisode (et non de section) en question ; cela concerne 23 épisodes sur 27. Par conséquent, ce ne sont pas les expressions spatio-temporelles qui définissent à elles seules le début d'une intrigue ; l'usage d'expressions personnelles est aussi très approprié pour marquer une nouvelle étape narrative dans le cadre d'un long récit.

6.5.3. Chaînes de référence au deuxième niveau local : épisodes mono-référentiels et multi-référentiels

Après avoir découpé les sections du roman-fleuve en épisodes, il est aussi plausible de définir ceux-ci par le nombre de référents saillants contenus dans chaque unité textuelle. Selon ce dernier facteur, les épisodes peuvent être mono-référentiels ou multi-référentiels. Comme le constatent et l'affirment Schnedecker (1992) et Xu (徐赳赳, 2014), le nombre de référents saillants, ou le nombre de topiques, est un des facteurs qui influencent la longueur des chaînes de référence ; or, d'après notre observation, ce facteur a également d'autres effets sur les chaînes. L'ensemble du roman-fleuve étant défini comme un genre multi-référentiel, les épisodes peuvent malgré tout être des séquences textuelles mono-référentielles ou multi-référentielles. Selon les études sur les chaînes dans les genres mono-référentiels (*cf*. Masseron & Schnedecker, 1995 ; Jenkins, 2002 ; Oberlé *et al*., 2018 ; Guo, 2021), les chaînes issues des deux types de situation référentielle ne sont pas censées être les mêmes. Il paraît ainsi intéressant d'exploiter les régularités distinctes des chaînes en les rapportant au nombre de référents concurrents dans les différents genres de discours.

Selon la répartition des chaînes montrée dans le tableau (6.2) au début de ce chapitre, 4 épisodes sur 50 peuvent être définis comme mono-référentiels, dont la seule chaîne dominante est celle du protagoniste, sans compétiteur dans le même épisode. Il s'agit des épisodes 7 (avec 86 maillons), 24 (avec 47 maillons), 48 (avec 82 maillons) et 50 (avec 81 maillons). Dans ces épisodes, les autres référents peuvent être nombreux, mais ils sont soit des simples singletons, soit des chaînes très courtes, sans pouvoir être compétiteurs de

saillance du référent dominant. Comme le montre l'extrait [104], la chaîne du protagoniste est entourée de multiples référents singletons :

[104] [**Il**] est à l'église avec grand-père. [**Il**] s'ennuie. [**Il**] n'est pas très à [**son**] aise. On [**lui**] défend de remuer, et les gens disent ensemble des mots qu'[**il**] ne comprend pas, et puis se taisent ensemble. Ils ont tous une figure solennelle et morose. [**Il**] les regarde, intimidé. La vieille Lina, la voisine, assise à côté de [**lui**], a pris un air méchant ; à des moments, [**il**] ne reconnaît même plus [**son**] grand-père. [**Il**] a un peu peur. Puis [**il**] s'habitue, et [**il**] cherche à se désennuyer par tous les moyens dont [**il**] dispose. [**Il**] se balance, [**il**] se tord le cou pour regarder au plafond, [**il**] fait des grimaces, [**il**] tire grand-père par son habit, [**il**] étudie les pailles de [**sa**] chaise, [**il**] tâche d'y faire un trou avec [**ses**] doigts, [**il**] écoute les cris d'oiseaux, [**il**] bâille à se décrocher la mâchoire.

Nous avons effectué une comparaison entre la chaîne « Jean-Christophe » dans les épisodes mono-référentiels et dans tout le corpus, ce dernier étant majoritairement multi-référentiel.

La longueur moyenne est de 53,8 maillons dans les épisodes mono-référentiels ; elle est largement inférieure à celle du corpus, qui est de 89,1 maillons par épisode. Par contre, la densité référentielle de ces épisodes, obtenue en divisant le nombre de maillons dans les épisodes mono-référentiels par leur nombre de mots, est de 0,078. Ce nombre est en revanche légèrement supérieur à celui du corpus, qui est 0,071.

$$0{,}078\ (densité\ référentielle) = \frac{296\ (\text{nombre de maillons})}{3816\ (\text{nombre de mots})}$$

D'après l'observation de l'extrait [104], dont la chaîne comporte 100 % de maillons non nominaux, il semble que la composition des chaînes des épisodes mono-référentiels est différente de celle des multi-référentiels. D'après le recensement des composants, 94 % des maillons des épisodes mono-référentiels sont non nominaux, contrairement à 84 % dans tout le corpus. Par contre, ce contexte ne motive pas beaucoup l'usage des zéros : seuls 2 % des maillons sont concernés, contre une proportion de 4 %, la plus élevée, de cette catégorie dans tout le corpus. L'emploi de zéro n'est pas favorisé par la diminution du nombre de référents saillants. Quant aux maillons nominaux, les noms propres sont toujours privilégiés dans les épisodes mono-référentiels et dans la globalité du corpus, mais ils se présentent beaucoup moins couramment dans ces derniers, à 4 % seulement, contre 11 % dans l'intégralité du corpus.

Tous les critères et chiffres sus-présentés sont synthétisés dans le tableau (6.26).

Tableau (6.26) : Chaînes de référence dans les épisodes mono-référentiels

Texte	Épisodes mono-référentiels	Corpus
Nbre de chaînes importantes	1	4
Nbre d'épisodes	4	50
Longueur moyenne de la chaîne du protagoniste	53,8	89,1
Densité référentielle	0,078	0,071
Composants remarquables	94 % de non nominaux 2 % de zéros 4 % de noms propres	84 % de non nominaux 4 % de zéros 11 % de noms propres

6.5.4. Chaînes de référence au troisième niveau local : séquences narratives et dialogales à l'intérieur des épisodes

Selon la définition d'Adam (1987) de la notion de *séquence*, un épisode peut comprendre un seul ou plusieurs types de séquences textuelles. Deux types de séquences textuelles sont identifiés dans notre corpus : les narratives et les dialogales[37]. Les premières constituent la partie primordiale du roman-fleuve, alors que les dernières ne sont pas négligeables en tant que constituants importants de l'œuvre, totalisant 8 % du corpus en nombre de mots. Toutes les séquences dialogales sont présentées dans l'*Annexe 6*.

Comme nous l'avons vu avec la définition de la section *3.2.3.1. Séquences narratives et dialogales dans le genre narratif* du *Chapitre 3*, les séquences dialogales sont constituées, d'une part, par les énoncés échangés entre deux ou plusieurs personnages du livre, sous forme de discours rapportés directs entre guillemets, avec des changements de lignes réguliers, et, d'autre part, par les phrases contextuelles antéposées, interposées et postposées aux énoncés. Les antéposées, comme « il dit, », « il crie, », ont pour fonction d'initier le dialogue ; les interposées, comme « fit le vieux », « selon son habitude », marquent les changements de prise de parole entre les participants ; les postposées, comme « ils ne parlent plus », « tous les deux se taisent », servent à clore la séquence dialogale.

Les 43 séquences dialogales totalisent 4939 mots et comportent 1169 maillons de diverses chaînes de divers référents humains et non humains. La catégorie sémantique privilégiée de ces séquences ne manifeste pas de différence avec celle du corpus entier. Toutes les séquences sont multi-référentielles puisque chacune d'elles est dominée par deux personnages ou plus, dont la chaîne correspond à chaque participant au dialogue. La densité de ces maillons est de 0,23, c'est-à-dire que 23 mots sont constituants des chaînes tous les cent mots. Ce chiffre paraît très élevé puisque la densité est seulement de 0,16 dans le corpus entier.

$$0{,}23\ (densité\ des\ séquences\ dialogales) = \frac{1169\ (\text{nombre de maillons})}{4939\ (\text{nombre de mots})}$$

$$0{,}16\ (densité\ du\ corpus) = \frac{9687\ (\text{nombre de maillons})}{61862\ (\text{nombre de mots})}$$

Les résultats de tous les critères de la comparaison en question sont synthétisés dans le tableau (6.27).

Tableau (6.27) : Chaînes de référence dans les séquences dialogales et dans le corpus

	Séquences dialogales	Corpus
Occupation du corpus	8 %	100 %
Macro-structure	Discours rapportés directs, sauts de lignes	Paragraphes
Référents des chaînes principales	Humains, non humains	Humains, non humains
Densité de maillons	0,23	0,16
Nombre de chaînes saillantes	Multi-référentiels (participants de dialogue)	Mono-référentiel et multi-référentiel

[37] Une seule séquence épistolaire est repérée. Vu cette quantité minimale, nous discutons des chaînes de ce type avec celles des séquences dialogales.

En ce qui concerne les composants privilégiés, la comparaison du tableau (6.28) permet d'obtenir quelques remarques sur les points communs et les différences entre l'emploi de chaque catégorie de maillon dans les séquences dialogales et dans le corpus intégral.

Tableau (6.28) : Composants des chaînes de référence dans les séquences dialogales et dans le corpus

Texte	Séquences dialogales (1169 maillons)	Corpus (9687 maillons)
Noms propres	14 %	9 %
SN définis	7 %	7 %
SN sans déterminant	3 %	2 %
SN démonstratifs	0,3 %	0,9 %
SN indéfinis	0,8 %	1 %
SN possessifs	4 %	2 %
Déictiques	32 %	3 %
Zéro	7 %	4 %
Autres catégories non nominales	31 %	58 %

À l'exception des noms propres, qui sont plus souvent utilisés dans les séquences dialogales (14 %) que dans le corpus (9 %), la fréquence de l'emploi des autres formes nominales est identique ou similaire dans les deux types de texte, cela impliquant des usages semblables de ces formes dans les chaînes de référence. Pourtant, des écarts se manifestent en ce qui concerne les catégories non nominales. Premièrement, les déictiques concernent 32 % de maillons dans les séquences dialogales, mais seulement 4 % dans tout le corpus ; deuxièmement, les zéros sont légèrement plus élevés (7 %) que celles dans le corpus (4 %) ; troisièmement, les autres catégories non nominales, y compris majoritairement les pronoms personnels de troisième personne et les déterminants possessifs, concernent 31 % de maillons dans les séquences dialogales, ce qui est largement inférieur à ces deux catégories dans le corpus entier (58 %). Ces trois différences permettent d'affirmer que les zéros et les pronoms de première et de deuxième personnes sont les composants privilégiés des séquences dialogales, alors que ce sont les pronoms de troisième personne et les déterminants possessifs qui le sont dans les séquences narratives dans la majorité du corpus.

6.5.5. Caractérisation globale des chaînes de référence dans le de roman-fleuve

Étant donné que de nombreuses études montrent déjà que les chaînes de référence peuvent permettre de distinguer les genres (*cf*. Longo & Todirascu, 2012, 2013 ; Schnedecker, 2014, 2017), nous avons essayé de proposer une caractérisation globale des chaînes de référence dans le genre de roman-fleuve au fondement de l'observation du corpus *Jean-Christophe*. Dans la macro-structure, ce genre de texte est un composite de diverses séquences textuelles selon trois critères.

En premier lieu, l'écrivain propose une partition du roman-fleuve en volumes, en sections numérotées et en paragraphes. Les sections sont marquées par des étoiles ou d'autres signes selon les éditions. Nous avons choisi d'ajouter un niveau d'unité textuelle appelée

épisode inférieur aux sections et supérieur aux paragraphes, afin de faciliter le traitement informatique et d'obtenir des intrigues marquées par les changements spatio-temporels au fil de la narration. La place d'apparition et la longueur des chaînes dans chaque épisode constituent les facteurs primordiaux qui déterminent la saillance du référent au niveau épisodique du corpus.

En deuxième lieu, selon la distinction de Schnedecker (1992, 1997) entre le contexte mono-référentiel et multi-référentiel, ce roman-fleuve se définit comme multi-référentiel dans son intégralité ; néanmoins, cela n'empêche pas qu'il comporte des extraits (parfois longs) mono-référentiels. La densité référentielle dans les épisodes mono-référentiels est plus élevée que celle du corpus ; les maillons non nominaux composent la partie majoritaire des épisodes mono-référentiels ; les épisodes multi-référentiels sont dominés par une ou plusieurs chaînes saillantes selon le nombre de maillons.

En troisième lieu, selon la définition de *séquence* d'Adam (1987), ce roman-fleuve est un composé de séquences narratives, à proprement parler, et de séquences dialogales. Les chaînes des séquences dialogales sont marquées par les composants privilégiés que sont les déictiques et les zéros, contrairement à celles des séquences narratives qui se composent principalement de pronoms de troisième personne.

Pour les trois raisons susmentionnées, un roman-fleuve est un genre plus complexe que les autres genres narratifs courts. Deux manières de le caractériser sont plausibles. D'un côté, du point de vue global, les chaînes de ce genre présentent des particularités spécifiques aux autres genres narratifs courts ; d'un autre côté, du point de vue local, lesdites chaînes ont des caractéristiques assez distinctes qui dépendent entièrement de leur type de séquences textuelles.

Sur la base des explorations de tous les phénomènes effectuées de la manière la plus fine possible dans la partie française du corpus, nous nous posons la question de savoir si les mêmes résultats se retrouvent aussi en chinois. Avec l'aide des études antérieures dans l'état de la question, nous chercherons, dans le chapitre suivant, d'abord à exploiter les chaînes de référence en chinois, et ensuite à formuler des hypothèses éventuelles sur les points communs et les différences qui pourront se manifester entre les deux langues.

Chapitre 7 : Propriétés des chaînes de référence dans un roman-fleuve traduit en chinois : approches contrastives franco-chinoises

Notre corpus *Jean-Christophe* est composé de deux parties : les extraits du roman en français et leurs textes parallèles traduits en chinois. Étant donné cette relation entre les textes en deux langues, certaines caractéristiques des chaînes de référence sont censées être similaires, comme le nombre de chaînes, les types de référents, les types de contextes mono-référentiels et multi-référentiels, la longueur de chaque chaîne et le coefficient de stabilité. D'autres peuvent être différentes, comme la catégorie privilégiée des maillons et celle du premier maillon, la fonction syntaxique privilégiée des maillons et les types de modifieurs des maillons. Nous avons présenté dans le chapitre précédent les chaînes de référence françaises ; dans ce deuxième chapitre consacré aux résultats, nous présentons les caractéristiques des chaînes de référence parallèles chinoises, par comparaison avec celles du français.

Notre première remarque relative aux chaînes de référence concerne les plus longues dans l'ensemble du corpus. Il est prévisible que ces chaînes soient les mêmes que celles du français en raison de la relation de traduction. Le résultat obtenu correspond en effet à cette hypothèse : les chaînes les plus longues sont les mêmes que celles du français, à savoir celles du protagoniste et de quelques personnages secondaires. La répartition et le nombre de maillons de ces chaînes sont présentés dans le tableau (7.1)[1]. Les cases sont marquées par le chiffre de maillons contenus dans chaque épisode ; celles qui restent vides ne contiennent pas de maillon ou ne contiennent que 1 ou 2 maillons.

Tableau (7.1) : Répartition des chaînes des personnages en chinois

		JC	L	M	JM	Duc	ML	ER	G
S I	1	48	76	26	73	-	-	-	-
	2	-	30	61	15	-	-	-	-
	3	36	17	4	26	-	-	-	-
	4	10	16	3	9	-	-	-	-
	5	78	8	7	-	-	-	-	-
	6	70	5	-	5	-	-	-	-
	7	94	-	-	-	-	-	-	-
	8	33	-	-	88	-	-	-	-
	9	95	-	-	32	-	-	-	-
	10	40	9	-	-	-	-	-	-
S II	11	-	-	69	123	4	-	-	-
	12	46	10	7	-	-	6	22	-
	13	185	42	-	-	-	-	-	-
	14	131	20	14	-	-	13	-	-
	15	58	32	19	-	-	-	-	-
	16	107	33	115	-	-	-	-	-
	17	45	-	3	-	-	-	-	-
	18	134	5	4	20	-	-	-	-
	19	68	37	-	-	-	-	-	-
	20	137	19	11	-	-	4	-	-
	21	85	6	19	5	-	-	-	-

[1] Les notes des abréviations de ce tableau sont les mêmes que celles du tableau (6.2) du *Chapitre 6*.

	22	123	3	25	-	-	-	-	-
	23	112	-	38	-	-	4	-	-
	24	34	-	-	-	-	-	-	-
S III	25	91	-	-	27	-	-	-	-
	26	66	-	-	27	-	-	-	-
	27	44	-	3	-	6	-	-	-
	28	76	-	-	18	3	-	-	-
	29	43	4	-	-	-	-	-	-
	30	136	-	24	77	-	-	-	-
	31	34	7	4	8	-	-	-	72
	32	130	-	-	-	-	-	-	102
	33	95	-	18	20	4	-	-	51
	34	45	-	20	10	7	-	-	-
	35	217	-	29	28	17	-	-	-
	36	81	4	14	8	-	6	-	-
S IV	37	92	-	-	4	5	-	-	-
	38	44	7	4	24	-	-	-	-
	39	36	-	4	21	-	-	-	-
	40	60	-	-	-	-	-	-	16
	41	-	9	12	105	-	-	-	-
	42	147	19	-	69	-	-	8	31
	43	32	6	8	7	-	-	-	4
	44	56	6	66	10	-	-	-	-
	45	72	8	75	-	-	-	13	-
	46	134	14	22	-	-	-	-	-
	47	89	11	37	-	-	-	3	-
	48	64	-	-	-	-	-	-	-
	49	83	9	9	-	-	-	20	-
	50	81	-	-	-	-	-	-	-
Nbre de maillons (épisodes)		3817 (47)	472 (29)	774 (32)	859 (26)	46 (7)	33 (5)	66 (5)	276 (6)

Tout comme en français, la chaîne la plus longue et la plus répandue est celle du protagoniste 'Christophe' *Kelisiduofu* 克利斯朵夫 (« JC »). Elle est composée de 3817 maillons qui répartissent dans presque tout le corpus, à savoir es quatre sections et 47 sur 50 épisodes. Cette chaîne sera présentée, analysée et comparée avec celle du français dans la première partie de ce chapitre. Ensuite, les trois autres chaînes longues et importantes sont celles des trois personnages secondaires 'Louisa' *Luyisha* 鲁意莎 (« L »), 'Melchior' *Manxiwo* 曼希沃 (« M ») et 'Jean-Michel' *Yuehan-Mixier* 约翰·米希尔 (« JM »). Elles sont identiques à celles du français. Elles se situent également dans toutes les quatre sections du corpus et dans environ la moitié des épisodes du corpus (entre 26 et 32 épisodes sur 50 selon les personnages). Nous procéderons de la même manière pour analyser ces trois chaînes. Outre les chaînes susmentionnées, celles des personnages *épisodiques* et *éphémères* (comme « Duc », « ML », « ER » et « G »[2]) paraissent également non négligeables. Elles sont présentes dans moins de 10 épisodes sur 50 ; elles ne possèdent qu'un petit nombre de maillons par rapport aux chaînes du protagoniste et des personnages secondaires. Ces deux groupes de chaînes seront

[2] Tous les personnages ne figurent pas dans le tableau (7.1) en raison de leur nombre trop élevé. Voir la répartition de tous les personnages épisodiques et éphémères dans les *Annexes 50* et *52*.

examinés en détails dans un troisième temps. Enfin, nous terminerons par les analyses sur les chaînes des référents non humains.

7.1. Chaîne du protagoniste 'Christophe' *Kelisiduofu* 克利斯朵夫

La chaîne du protagoniste 'Christophe' *Kelisiduofu* 克利斯朵夫 est constituée de 3817 maillons dans l'intégralité du corpus chinois. Tous les maillons sont listés avec les gloses françaises dans l'*Annexe 29 : Maillons chinois dans la chaîne 'Christophe'* Kelisiduofu *克利斯朵夫*.

La densité référentielle est de 0,042 ; elle est moins élevée que celle du français, à 0,071 : seuls 4,2 caractères sont concernés dans la chaîne tous les cent caractères chinois.

$$0{,}042\ (densité\ référentielle) = \frac{3817\ (nombre\ de\ maillons)}{89858\ (\text{nombre de caractères chinois})}$$

Cette densité relativement basse explique en quelque sorte pourquoi les chaînes sont moins étudiées en chinois : sans parler des zéros invisibles qui concernent une proportion significative dans les chaînes, les maillons explicites et implicites ne concernent qu'une petite proportion de caractères, par rapport à ceux du français. Les maillons visibles concernent encore une plus petite proportion de texte.

Les groupes parallèles de maillons nominaux et non nominaux du français sont repris pour ceux du chinois, malgré quelques modifications mineures : les nominaux sont subdivisés en *nom propre*, *SN possessif*, *SN démonstratif*, *SN indéfini* et *SN sans déterminant* ; les non nominaux le sont en *pronom personnel*, *déterminant possessif*, *déictique*, *zéro* et *autres*. Le nombre ainsi que la proportion de chaque catégorie sont synthétisés dans le tableau (7.2).

Tableau (7.2) : Catégories des maillons chinois dans la chaîne du protagoniste

Maillon	Catégorie	Pourcentage (nbre) de maillons	
Catégories nominales	Nom propre	11 % (410)	17 %
	SN possessif	0,1 % (3)	
	SN démonstratif	0,3 % (13)	
	SN indéfini	0,1 % (4)	
	SN sans déterminant	6 % (210)	
Catégories non nominales	Pronom personnel	44 % (1676)	83 %
	Déterminant possessif	8 % (305)	
	Déictique	4 % (161)	
	Zéro	27 % (1032)	
	Autres	0,1 % (3)	

Étant donné que les textes chinois sont traduits à partir des textes français, il est prévisible que la proportion de chaque catégorie soit plus ou moins semblable entre les deux langues. Cette supposition permet d'expliquer les proportions similaires de plusieurs catégories entre le chinois et le français. Par exemple, celle des maillons nominaux est de 17% en chinois et de 16 % en français, et respectivement de 83 % et de 84 % pour les non nominaux.

Ces deux types de maillons manifestent donc des fréquences d'occurrences très proches dans la chaîne du protagoniste dans les deux langues.

Le tableau comparatif (7.3) révèle que plusieurs groupes de maillons nominaux chinois présentent des proportions assez proches. 11 % de noms propres chinois égalent à la proportion identique de noms propres français ; 0,1 % de SN possessifs chinois correspondent à la petite quantité de 1 % de SN possessifs français, malgré cette différence de moins de 1 % ; 0,3 % de SN démonstratifs existent en chinois, similairement à 0,2 % en français ; les SN indéfinis sont aussi très peu nombreux dans cette chaîne dans les deux langues, à savoir 0,1 % en chinois et 0,2 % en français. En ce qui concerne les maillons non nominaux, les déictiques manifestent la même proportion de 4 % dans les deux langues.

Les autres groupes de maillons que ceux susmentionnés montrent cependant des proportions qui contredisent notre supposition selon laquelle des proportions similaires existent entre les textes en relation de traduction. Il convient par conséquent de s'intéresser à présent à ces dissimilitudes éventuelles entre les chaînes du même personnage dans les deux langues. Cinq différences significatives sont constatées :

- le groupe de SN définis, qui concerne 3 % de composants dans la chaîne en français, n'existe pas en chinois[3] ;
- les SN sans déterminant constituent 6 % de la chaîne avec 210 occurrences en chinois, alors que seuls 8 maillons sont récupérés dans la chaîne française ;
- les pronoms personnels de troisième personne concernent 44% de la chaîne chinoise, contrairement à une proportion plus élevée de 60 % en français ;
- les déterminants possessifs constituent 8 % de la chaîne chinoise, ce qui représente seulement la moitié de celle du français : 16 % ;
- les zéros sont récupérés 1032 fois, soit 27 % de la chaîne chinoise, alors que leur proportion en français atteint seulement 4 %.

Tableau (7.3) : Proportions dissimilaires des maillons chinois et français dans la chaîne du protagoniste

Langue	Ch	Fr
SN définis	-	3 % (139)
SN sans déterminant	6 % (210)	0,2 % (8)
Pronoms personnels de troisième personne	44 % (1676)	60 % (2643)
Déterminants possessifs	8 % (305)	16 % (720)
Zéros	27 % (1032)	4 % (161)

À partir de ce bilan de ces proportions, nous présenterons et analyserons dans cette partie les raisons et les interprétations de ces résultats en prenant des exemples *ad hoc* qui représentent les cas prototypiques et les cas exceptionnels, dans le but de chercher quelles sont les conditions spécifiques pour tel usage de maillon en chinois par comparaison avec celles du français. Pour les catégories qui présentent des faibles quantités de maillons (3 – 40 occurrences), les analyses quantitatives et qualitatives seront réalisées sur la totalité de toutes

[3] Le chinois est dépourvu d'articles définis, comme le turc et certaines langues slaves (*cf.* Anderson & Keenan, 1985, cités par Ariel, 1990 : 52). Si les articles définis et indéfinis sont obligatoires en français, ils ne le sont pas en chinois (*cf.* Dong, 董秀芳, 2010 : 19) ni en japonais (*cf.* Gil, 1990 : 385).

les occurrences, avec la démonstration par des exemples ; pour celles qui en contiennent une grande quantité (≥ 40 occurrences), les analyses se réaliseront sur l'ensemble d'occurrences ou sur quelques échantillons, que nous expliciterons selon les spécificités des maillon et éventuellement des chaînes. Ces échantillons doivent se situer dans des épisodes consécutifs choisis de manière aléatoire dans différentes sections du roman afin d'éviter des répétitions et les biais éventuels des résultats.

En l'absence de définition reconnue dans le domaine de la linguistique chinoise, les chaînes de référence chinoises sont interprétées de différentes manières en fonction des exemples des notions proches ou annexes, comme nous l'avons montré dans les chapitres des deux premières parties. Avec ces définitions fondamentales et primordiales, nous rappelons la composition d'une chaîne de référence chinoise en citant le début de la chaîne du protagoniste dans le roman qui contient quatre maillons annotés entre crochets :

[1] **[初生的 婴儿]** 在 摇篮 里 扭动。老人 进来 虽然 把 木靴 脱 在 门
chushengde yinger zai yaolan li niudong laoren jinlai suiran ba muxue tuo zai men
nouveau-né bébé PREP berceau dedans s'agiter vieux-homme entrer malgré BA sabot laisser PREP porte
protagoniste-1

外， 走路 的 时候 地板 还是 格格的响， **[孩子]** 哼啊嗐啊 的 哭 了。母亲
wai zoulu de shihou diban haishi gegedexiang haizi hengahaia de ku le muqin
extérieur marcher DE temps plancher encore craquer enfant ONOMA DE pleurer LE mère
protagoniste-2

从 床 上 探出 身子 抚慰 **[他]**； 祖父 摸索着 点起灯来， 免得 **[孩子]**
cong chuang shang tanchu shenzi fuwei ta zufu mosuozhe dianqidenglai miande haizi
par lit dessus sortir corps assurer 3P grand-père tâtonnant allumer-lampe évitant enfant
protagoniste-3 *protagoniste-4*

在 黑夜 里 害怕。
zai heiye li haipa
PREP nuit dedant avoir-peur

'**[Le nouveau-né]** s'agite dans **[son]** berceau. Bien que le vieux ait laissé, pour entrer, ses sabots à la porte, son pas a fait craquer le plancher ; **[l'enfant]** commence à geindre. La mère se penche hors de **[son]** lit, afin de **[le]** rassurer ; et le grand-père allume la lampe en tâtonnant, pour que **[le petit]** n'ait pas peur de la nuit.'

Cet extrait de chaîne contient 4 maillons de catégories nominales et non nominales, occupant des fonctions syntaxiques diverses. Le premier, 'nouveau-né bébé' *chushengde-yinger* 初生的婴儿, est un SN sans déterminant, modifié par un adjectif, en position de sujet de la phrase. Ce maillon est ainsi le premier de la chaîne du protagoniste dans tout le roman ; c'est donc un SN sans déterminant modifié chinois qui démarre la chaîne, à la différence d'un SN défini sans modifieur français. Le deuxième maillon, 'enfant' *haizi* 孩子, est un SN sans déterminant et sans modifieur, en position de sujet de la phrase. Cette expression nominale est coréférentielle et anaphorique au premier maillon 'nouveau-né bébé' *chushengde-yinger* 初生的婴儿. Le troisième maillon, '3P' *ta* 他, est un pronom personnel masculin de troisième personne au singulier, en position de complément d'objet de la phrase, anaphorique. Le quatrième maillon, 'enfant' *haizi* 孩子, est aussi un SN sans déterminant sans modifieur. Il est en position de sujet de la phrase, constituant une cataphore avec le troisième maillon '3P' *ta* 他.

Nous pouvons donc noter provisoirement que cette chaîne échantillonnée de 4 maillons se constitue de 3 SN sans déterminant composants majoritaires de la chaîne, tandis que la proposition de cette catégorie dans la totalité de la chaîne ne montre pas forcément la même tendance, puisqu'elle n'en concerne que 6 %. En d'autres termes, les premiers maillons sont plutôt en positions particulières avec une grande densité de SN sans déterminant. Ce

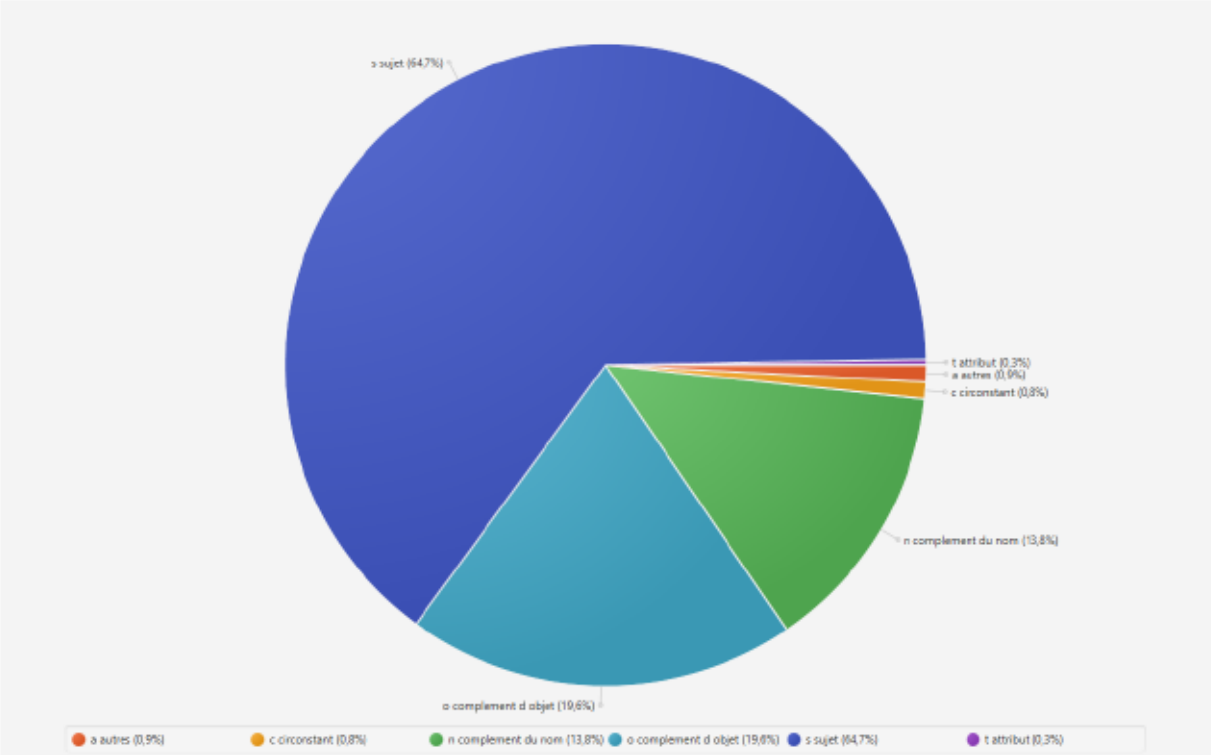
constat nous incite à continuer l'observation des influences éventuelles liées à des facteurs syntaxiques et discursifs sur la composition de la chaîne en chinois.

En ce qui concerne les fonctions syntaxiques observées dans l'extrait [1], 3 maillons du protagoniste sur 4 sont sujets et un seul est complément d'objet de la phrase. Dans les études traditionnelles de la référence en chinois, de nombreux auteurs ont remarqué que les sujets ont tendance à référer aux entités définies, alors que les compléments d'objet réfèrent aux entités indéfinies (*cf*. Wan, 完权, 2016 : 273), parce que les sujets sont censés mentionner des entités déjà connues, contrairement aux compléments d'objet qui en introduisent normalement de nouvelles dans les constructions syntaxiques prototypiques SVO, malgré quelques structures exceptionnelles[4]. C'est la raison principale pour laquelle le sujet est la fonction préférée des maillons du protagoniste. Cela admis, revenons à l'observation des chiffres du corpus. Dans la globalité de la chaîne, la proportion de chaque fonction syntaxique est présentée dans la figure (7.1) issue du calculeur CRViewer : en accord avec notre hypothèse, la plus grande proportion de maillons (65%, soit environ deux tiers) sont sujets de phrase ; le dernier tiers est réparti parmi l'ensemble des autres fonctions : 20 % de compléments d'objet à la deuxième place, 14 % de compléments du nom à la troisième place, 0,8 % de circonstants, 0,3 % d'attributs et 0,3 % d'autres fonctions. L'extrait [1] reflète approximativement la même composition que celle de la chaîne dans le corpus entier : les deux composants principaux sont sujets et compléments d'objet ; les sujets se trouvent à la première place avec une grande moitié de composants, et les compléments d'objet à la seconde ; les autres fonctions ne présentent pas de proportions significatives. Selon l'approche cognitive, les référents exprimés à la place de sujet sont théoriquement plus *centraux* que ceux des autres fonctions syntaxiques[5]. Par conséquent, les maillons de la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫, avec 3 occurrences au sujet dans l'extrait [1], peuvent être à une place plus centrale que celle des référents dont les expressions engagent d'autres fonctions syntaxiques, comme 'berceau' *yaolan* 摇篮 et 'sabots' *muxue* 木靴, qui ne sont pas sujets de phrase.

[4] Trois exceptions sont significatives selon Wan (完权, 2016) : (i) les phrases existentielles ; (ii) les phrases co-topicales avec les constructions en *gei* 给 et *ba* 把 et (iii) les conversations actualisées. En chinois, ces trois types de constructions sont particulièrement utilisées, soit pour introduire un nouveau référent, soit pour mettre en relief un référent qui ne se trouvait pas en focus d'attention.

[5] Voir la définition des notions de *centrage*, *centre* et *central* dans la section *3.2.1.2. Chaînes de référence dans les contextes mono-référentiel et multi-référentiel* du *Chapitre 3*.

Figure (7.1) : Fonctions syntaxiques des maillons chinois dans la chaîne du protagoniste



En termes de nombre de référents, tout comme en français, *Jean-Christophe* en chinois peut être défini comme un texte multi-référentiel, constitué majoritairement de séquences multi-référentielles et de quelques séquences mono-référentielles. Cela nous permet d'observer la manière dont le nombre de référents influence la composition des chaînes de référence en chinois.

Après cette présentation générale de la chaîne du protagoniste en chinois, nous examinons ses composants selon les mêmes critères formels que nous avons appliqués pour celle du français, c'est-à-dire selon la tête lexicale, le déterminant et le modifieur pour les maillons nominaux, et ensuite selon la forme explicite ou implicite des maillons non nominaux.

7.1.1. Maillons nominaux selon la catégorie du nom et le déterminant

Selon la catégorie du nom, nous avons obtenus des maillons constitués de noms propres et de noms communs. Ceux-ci sont marqués, ou non, par divers déterminants, comme les démonstratifs, les possessifs et les indéfinis. Selon ces deux critères, nous examinerons les maillons nominaux répartis en cinq groupes : les noms propres, les SN sans déterminant, les SN démonstratifs, les SN possessifs et les SN indéfinis.

7.1.1.1. Noms propres : interprétations similaires et différentes selon les approches comparatives

Avec un total de 410 maillons, ce qui équivaut à 11 % de la chaîne, le nom propre est la catégorie nominale la plus présente ; cette proportion est identique à celle du français, correspondant à un nombre un peu plus élevé de 503 maillons. Cette identité proportionnelle des deux langues permet de présumer que les noms propres sont probablement et majoritairement traduits de manière fidèle, sous les mêmes conditions syntaxiques et discursives.

Globalement, l'interprétation des noms propres chinois ressemble beaucoup à celle du français ; les points similaires se réduisent à trois aspects principaux.

En premier lieu, la prédominance des noms propres s'accorde avec l'hypothèse cognitive relative à l'importance du référent : les noms propres chinois sont aussi privilégiés par rapport aux autres catégories lors de l'introduction des référents les plus importants. C'est la première remarque que nous avons obtenue sur les noms propres chinois.

En deuxième lieu, au vu l'équivalence des noms propres dans les deux langues, notre hypothèse porte sur la correspondance traductive entre les noms propres français et chinois : ces derniers sont présumés être majoritairement traduits de manière fidèle en chinois. L'observation des exemples dans le corpus confirme qu'ils correspondent à la forme des noms propres français d'une manière plus fidèle que quand il s'agit des autres formes des noms communs et des catégories non nominales. Par exemple, les deux noms propres français dans l'extrait [2] sont traduits par deux noms propres chinois dans l'extrait parallèle [3], sans modification morphologique. Si nous les comparons avec les autres expressions référentielles, on en remarque4 dont la catégorie n'est pas la même : le pronom démonstratif français « cela » (« ER-1 ») est exprimé par le SN démonstratif chinois 'ce CL sens' *zhe-ge-yisi* 这个意思 ; le SN défini français « le vieux » (« ER-2 ») est traduit par le SN sans déterminant chinois 'vieux-homme' *laoren* 老人 ; le SN indéfini français « une histoire » (« ER-3 ») correspond au SN sans déterminant chinois 'histoire' *gushi* 故事, et le même type de changement se produit avec le SN défini français « l'enfant » (« ER-4 ») traduit par le SN sans déterminant chinois 'enfant' *haizi* 孩子.

[2] Grand-père toussotait. [**Christophe**]$_{npr\text{-}1}$ savait bien ce que [cela]$_{ER\text{-}1}$ voulait dire. [Le vieux]$_{ER\text{-}2}$ brûlait d'envie de raconter [une histoire]$_{ER\text{-}3}$; mais il voulait que [l'enfant]$_{ER\text{-}4}$ la lui demandât. [**Christophe**]$_{npr\text{-}2}$ n'y manquait pas. Ils s'entendaient ensemble.

[3]

祖父	咳	了	几	声	。	[克利斯朵夫]	很	明白	[这	个	意思]	。	[老人]	极想
zufu	*ke*	*le*	*ji*	*sheng*		*Kelisiduofu*	*hen*	*mingbai*	*zhe*	*ge*	*yisi*		*laoren*	*jixiang*
grand-père	tousser	LE	quelque	CL		Christophe *npr-1*	bien	savoir	ce	CL	sens *ER-1*		vieux-homme *ER-2*	désirer

讲	[故事]	，	但	要	[孩子]	向	他	请求	。	[克利斯朵夫]	立刻	凑上去	。	他们俩
jiang	*gushi*		*dan*	*yao*	*haizi*	*xiang*	*ta*	*qingqiu*		*Kelisiduofu*	*like*	*coushangqu*		*tamenlia*
raconter	histoire *ER-3*		mais	vouloir	enfant *ER-4*	vers	3P	demander		Christophe *npr-2*	immédiatement	s'approcher		3Ps

很	投机	。
hen	*touji*	
très	s'entendre	

En troisième lieu, le nom propre peut être une des catégories privilégiées à la place du premier maillon de chaîne pour introduire ou réintroduire un nouveau référent. Malgré leur

absence à cette place dans la chaîne du protagoniste au début du roman, les noms propres chinois, tout comme ceux du français, redémarrent la chaîne suivant le découpage typographique (début de paragraphe) et spatio-temporel (début d'épisode) au milieu du roman : ils occupent la place du premier maillon dans 19 épisodes sur les 47 contenant les maillons de la chaîne du protagoniste[6].

Nous aimerions présenter les noms propres chinois dans la chaîne du protagoniste par le biais d'une comparaison avec ceux du français afin d'en saisir les caractéristiques similaires et les différences. En raison du grand nombre de maillons qui fait que le travail d'analyse de tous les noms propres au cas par cas est gigantesque, nous avons choisi d'effectuer les analyses quantitatives sur un échantillon de 20 occurrences au début de la section II du roman (voir l'*Annexe 30 : Échantillon de noms propres chinois dans la chaîne 'Christophe'* Kelisiduofu *克利斯朵夫 (section II, épisode 11)*). D'autres emplois exceptionnels éventuels des noms propres chinois pourront être exposés plus tard. Nous avons repris les quatre critères des maillons français pour ceux du chinois, avec l'objectif de comparer les points divergents et communs éventuels entre les deux langues :

- la distance inter-maillonnaire entre un nom propre et son antécédent : est-ce que les noms propres chinois favorisent le redémarrage de la chaîne après de grandes distances ou plutôt la maintenance de la chaîne après de petites distances ?
- le nombre de référents dans le contexte : est-ce que les noms propres chinois permettent de faire coïncider plusieurs référents dans le même contexte sans susciter d'ambiguïté référentielle ?
- la place initiale dans les séquences textuelles : est-ce que les noms propres chinois servent à initier une séquence typographique ou spatio-temporelle, comme les paragraphes, les épisodes et les sections du roman-fleuve, ou, dans le cas contraire, à maintenir la continuité référentielle au milieu des séquences textuelles ?
- la localisation dans les différents types de séquence textuelle : est-ce que ce sont les séquences narratives qui favorisent les occurrences des noms propres, ou, dans le cas contraire, les séquences dialogales ou autres types de séquences textuelles qui permettent de les utiliser plus souvent ?

Ces quatre critères contextuels nous ont permis d'obtenir quelques caractéristiques discursives des noms propres dans l'échantillon en question. Les paramètres détaillés de chaque occurrence sont montrés dans le tableau (7.4).

Tableau (7.4) : Caractéristiques discursives des noms propres chinois dans la chaîne du protagoniste (section II, épisodes 12 et 13)

Maillons	Fonction	Distance	Réf	Début	Discours
'petit Christophe' *xiao-Kelisiduofu* 小克利斯朵夫 (« npr-1 »)	S	LG	BP	EP	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-2 »)	CDO	5	9	N	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-3 »)	CDO	1	1	PA	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-4 »)	S	2	3	N	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-5 »)	S	2	1	N	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-6 »)	CDO	1	5	N	N

[6] Pour les 28 épisodes restants, le pronom personnel est le premier maillon de 20 d'entre eux, le SN sans déterminant pour 5 et les autres formes pour 3.

'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-7 »)	CDO	1	3	N	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-8 »)	CDO	1	11	PA	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-9 »)	S	1	0	PA	N
'petit Christophe' *xiao-Kelisiduofu* 小克利斯朵夫 (« npr-10 »)	S	4	12	N	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-11 »)	S	1	1	N	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-12 »)	S	6	22	PA	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-13 »)	S	3	3	N	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-14 »)	S	1	1	N	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-15 »)	S	1	1	N	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-16 »)	S	1	3	N	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-17 »)	CC	1	8	PA	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-18 »)	S	1	2	N	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-19 »)	S	1	0	N	N
'Christophe' *Kelisiduofu* 克利斯朵夫 (« npr-20 »)	S	1	1	N	N
Total / moyenne	70 % S	1,8 (1 LG)	6,4 (1 BP)	5 PA 1 EP (30 %)	0 DD

Selon le recensement d'un échantillon de noms propres, cette catégorie manifeste un très haut degré de stabilité lexicale : seuls 2 occurrences sont modifiées d'un adjectif 'petit' *xiao* 小 et les maillons restants, soit 90 %, sont identiques. Les noms propres occupent majoritairement les sujets dans 70 % des cas et les compléments d'objet à la deuxième place. La distance inter-maillonnaire est relativement faible : la moyenne est de 1,8 phrases. Entre deux maillons il existe en moyenne 6,4 référents. Une seule occurrence apparaît après une longue distance de plus de 10 phrases, correspondant à « beaucoup » de maillons concurrents entre ce nom propre et son antécédent. La place initiale des noms propres semble remarquable : 30 % se situent au début d'un paragraphe ou d'un épisode, cela impliquant que cette forme en chinois peut coïncider avec les changements typographiques et spatio-temporels, conformément à ce qui a été vu pour le français. Pourtant, la proportion des noms propres chinois utilisés dans les séquences dialogales semble peu significative dans cet échantillon qui est extrait d'une séquence narrative : les occurrences éventuelles des noms propres dans les séquences dialogales ne peuvent pas être reflétées par ces exemples.

7.1.1.2. SN sans déterminant

Comme de nombreux auteurs le montrent, il existe, pour la même forme d'expression, des corrélations entre la mode de référence et la fonction syntaxique (*cf.* notamment Zhang, 张伯江, 1997 ; Cornish, 2000 ; Dong, 董秀芳, 2010). Cette spécificité est particulièrement importante pour les SN sans déterminant chinois. Ces derniers sont régulièrement comparés avec la même forme dans les langues qui disposent d'un système d'articles définis. Des questions relatives aux ambiguïtés référentielles sont souvent évoquées : est-ce que le déterminant zéro peut engendrer des ambiguïtés ? Quel est le système complémentaire qui peut permettre d'éviter ces ambiguïtés ? Face à ce type de questions, les SN sans déterminant chinois semblent disposer de leurs propres caractéristiques qui permettent d'éviter ce genre

de problèmes, par le truchement de conditions discursives spécifiques qui éliminent les ambiguïtés réelles. Dans ce sens, de nombreux auteurs signalent que les SN sans déterminant chinois sont équivalents aux SN définis dans les langues qui disposent d'articles définis, y compris notamment l'anglais et le français (*cf*. Liu, 刘丹青, 2002 ; Dong, 董秀芳, 2010 ; Xiong, 熊岭, 2012 ; Wan, 完权, 2016), sans préciser les spécificités relatives à ce dernier. En fait, comme nous l'avons vu lors de la présentation des fonctions référentielles des SN définis français dans la section *6.1.2.2. SN définis* du *Chapitre 6*, les SN définis français peuvent référer, de manière globale, à des entités génériques ou spécifiques au singulier. Prenant en compte ces remarques, s'il existe des interprétations référentielles des SN sans déterminant chinois standardisées dans la grammaire, elles sont censées être très proches de celles des SN définis français. À partir de cette conclusion grammaticale, ce qui paraît plus intéressant, pour beaucoup d'auteurs sinophones, comme notamment Chen (陈平, 1987), Liu & Xu (刘丹青 & 徐烈炯, 1998), Xu (徐烈炯, 1995), Abbott (2006), Xu (许余龙, 2010) et Li (李虹, 2015), ce sont les conditions phrastiques et discursives spécifiques des SN sans déterminant chinois qui peuvent être interprétés par telle ou telle forme dans les langues disposant d'articles définis. Par exemple, Abbott (2006) propose trois traductions possibles des SN sans déterminant chinois vers l'anglais comme le montre le groupe d'exemples [4] et [5], avec notre traduction vers le français : selon le contexte phrastique, le SN sans déterminant 'livre' *shu* 书 peut être interprété par les manières suivantes : (i) le référent est générique ou indéfini au singulier, exprimé par « un livre » dans l'exemple [4], (ii) le référent est indéfini pluriel, exprimé par « des livres » dans l'exemple [4] et (iii) le référent est spécifique au singulier, exprimé par « le livre » dans l'exemple [5].

[4] 桌子 上 有 [书]。
zhuozi *shang* *you* *shu*
table on have book
SN-sans-dét

'There is [**a book**] on the table. / Il y a [**un livre**] sur la table.'
'There is [**books**] on the table. / Il y a [**des livres**] sur la table.'

[5] [书] 在 桌子 上。
shu *zai* *zhuozi* *shang*
book is-located table on
SN-sans-dét

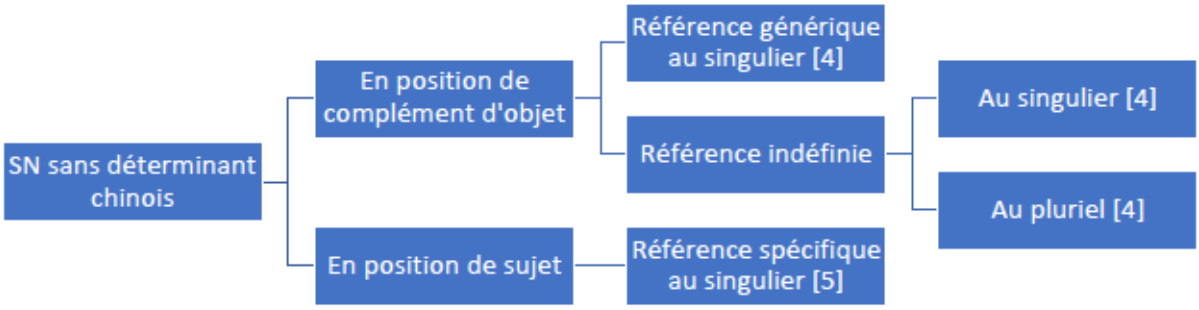
'[**The book**] is on the table. / [**Le livre**] est sur la table.'

(Abbott, 2006 : 108, nous traduisons en français)

La distinction entre la référence définie et la référence indéfinie des exemples [4] et [5] peut être reliée à la différence de la fonction syntaxique : quand le SN sans déterminant occupe la position de sujet, il a tendance à être interprété comme défini ou générique ; quand il occupe la position de complément d'objet, il a la tendance à être interprété comme indéfini. En d'autres termes, c'est la fonction syntaxique du SN sans déterminant qui joue un rôle décisif dans ces contextes phrastiques.

La distinction entre les interprétations au singulier et au pluriel de l'exemple [4] semble plus dépendante de la situation d'énonciation immédiate, parce que cet énoncé peut se produire suite à la question 'qu'est-ce qu'il y a sur la table ?' *zhuozi-shang-you-shenme* 桌子上有什么 ?, dont la réponse met l'accent sur la nature ontologique du référent, au lieu de sa quantité. L'interprétation « un livre » ou « des livres » dépendra de la quantité réelle de référent dans la situation d'énonciation. Cette observation permet d'affirmer que l'interprétation référentielle ne dépend pas uniquement de la fonction syntaxique, mais également du contexte discursif.

Figure (7.2) : Interprétation de SN sans déterminant chinois selon Abbott (2006)



Dans la chaîne du protagoniste, la différence de SN sans déterminant s'avère très importante entre le chinois et le français. Ceux-ci occupent 6 % de la chaîne chinoise, à la deuxième place des catégories nominales après les noms propres, pour 210 maillons, alors que cette proportion, rappelons-le, est de 0,2 % en français avec très peu d'occurrences. Comme nous l'avons observé dans l'extrait [1] au début de ce chapitre, 3 SN sans déterminant chinois 'nouveau-né bébé' *chushengde-yinger* 初生的婴儿, 'enfant' *haizi* 孩子 et 'enfant' *haizi* 孩子 correspondent à 3 SN définis français « le nouveau-né », « l'enfant » et « le petit ». Au vu de ces résultats, il est prévisible qu'une partie importante des SN sans déterminant chinois dans la chaîne du protagoniste équivaudront aux SN définis français. Dans cette partie, nous analyserons, à la suite des remarques d'autres auteurs sur les modes de référence, les caractéristiques référentielles des SN sans déterminant chinois dans la chaîne du protagoniste en fonction de diverses conditions phrastiques et discursives.

En fonction des conditions syntaxiques et discursives, toutes les interprétations référentielles de SN sans déterminant chinois ne correspondent pas à celles de SN définis français. Nous réexaminerons donc la question de la référence des SN sans déterminant chinois en deux temps : (i) selon leur fonction syntaxique dans le cadre phrastique et (ii) selon quelques facteurs discursifs dans le cadre inter-phrastique[7].

7.1.1.2.2. Dans le cadre phrastique : les SN sans déterminant à différentes positions syntaxiques

Cette observation sur la fonction syntaxique des SN sans déterminant est liée au fait que nous nous intéressons ici seulement aux maillons de la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫, à savoir les expressions coréférentielles renvoyant à un personnage spécifique. Ainsi, les discussions dans cette partie ne concernent pas encore les questions plus générales relatives aux SN sans déterminant chinois qui renvoient à d'autres types de référents.

[7] Les caractéristiques syntaxiques et discursives de chaque SN sans déterminant de la chaîne sont exposées dans l'*Annexe 31 : Caractéristiques discursives des SN sans déterminant chinois dans la chaîne 'Christophe'* Kelisiduofu *克利斯朵夫*.

Il semble que, si le cadre d'analyse d'Abbott (2006) s'applique à ce fait, la position syntaxique la plus probable des maillons est le sujet. Cette remarque peut-elle être pertinente pour expliquer l'emploi des SN sans déterminant dans cette chaîne du roman ? Pour répondre à cette question, nous avons recensé les fonctions syntaxiques de tous les SN sans déterminant de la chaîne. Une synthèse des fonctions syntaxiques des maillons est montrée dans le tableau (7.5) ci-dessous, où l'on voit que les sujets constituent la plus grande proportion de maillons, équivalente à 44 % de SN sans déterminant. Toutes les autres fonctions syntaxiques occupent des proportions moindres : les compléments d'objet 26 %, les compléments du nom 19 %, les circonstants 2 % et les autres fonctions 9 %.

Tableau (7.5) : Fonctions syntaxiques des SN sans déterminant chinois dans la chaîne du protagoniste

	Nbre	Pourcentage
Sujet	92	44 %
Complément d'objet	55	26 %
Complément du nom	40	19 %
Circonstant	4	2 %
Autres	19	9 %

Les exemples de sujet sont donnés dans l'extrait [1] au début de ce chapitre, comme 'nouveau-né bébé' *chushengde-yinger* 初生的婴儿 et 'enfant' *haizi* 孩子 ; ceux qui sont en position de complément d'objet, de complément du nom et de circonstant sont présentés dans les extraits [6] – [8] suivants :

[6] 他 表面 上 装做 独往独来， 说话 毫无 忌讳， 瞧不起 名衔 地位， 骨子 里
ta biaomian shang zhuangzuo dulaiduwang shuohua haowu jihui qiaobuqi mingxian diwei guzi li
3P surface sur prétendre indépendant parler aucun tabou mépriser titre classe nature dedans

却 是 挺 天真的 仰慕 金钱， 权势， 荣誉， 声望； 看见 [孙儿] 能
que shi ting tianzhende yangmu jinqian quanshi rongyu shengwang kanjian suner neng
mais être assez naïf admirer argent pouvoir honneur réputation voir petit-fils pouvoir
SN-sans- dét-CDO

接近 那些 有财有势的 人， 他 真 得意 极了。
jiejin naxie youcaiyoushide ren ta zhen deyi jile
approcher ces-là riche-puissant personne 3P vraiment réjouissant trop

'Il affectait bien l'indépendance, l'humeur frondeuse, le mépris des grandeurs ; mais il avait une admiration naïve pour l'argent, le pouvoir, les honneurs, les distinctions sociales ; sa fierté était sans pareille de voir son petit-fils approcher ceux qui y participaient : il en jouissait.'

[7] 老人 也 跟 [孩子] 一样 紧张， 他 没 法 坐下， 老 在
lauren ye gen haizi yiyang jinzhang ta mei fa zuoxia lao zai
vieux-homme aussi comme enfant aussi stressé 3P NEG pouvoir s'asseoir toujours PREP
SN-sans-dét-circonstant

屋子 里 踱来踱去， 按着 文字 的 内容 做出 各种 表情。
wuzi li duolaiduoqu anzhe wenzi de neirong zuochu gezhong biaoqing
chambre intérieur marcher selon texte DE contenu mimer tout visage

'Le vieux n'était pas moins ému. Il n'avait pu rester assis ; il se promenait à travers la chambre, en mimant les expressions de son texte.'

[8] [他] 从 有心讨好的 邻人 那边， 知道 父亲 说 [他] 坏话。 曼希沃 从前 为了
ta cong youxintaohaode lingren nabian zhidao fuqin shuo ta huaihua Manxiwo congqian weile
3P par officieux voisin là-bas apprendre père dire 3P mal Melchior auparavant pour

[儿子] 的 光荣 大为得意， 此 刻 却 不 知 羞耻 的 忌妒 起来。
erzi de guangrong daweideyi ci ke que bu zhi xiuchi de jidu qilai
fils DE gloire très-satisfait ce moment mais NEG savoir honte DE jaloux AUX
SN-sans-dét-CDN

'Il apprit par d'officieux voisins que son père disait du mal de lui. Après avoir été glorieux des succès de son fils, Melchior avait la honteuse faiblesse d'en devenir jaloux.'

Les sujets comptent seulement moins d'une moitié des maillons de la chaîne (44 %). En d'autres termes, une grande moitié de maillons occupent toutes les autres fonctions syntaxiques. Cette proportion de sujets est également beaucoup moins élevée que celle du français (57 %).

En raison de cette proportion moins importante que prévue, les SN sans déterminant occupant toutes les fonctions syntaxiques peuvent en réalité également renvoyer à un référent spécifique et défini avec des occurrences moins significatives que celles des sujets. Cette répartition entre les fonctions syntaxiques implique en même temps que ce facteur dans le cadre phrastique n'est pas le seul à influencer le mode de référence des SN sans déterminant chinois dans le contexte de ce long texte narratif. L'étape suivante consiste donc à se demander : i) quels sont les autres facteurs discursifs potentiels qui peuvent avoir une influence plus ou moins grande sur le mode de référence ; en d'autres termes, quelles sont les conditions contextuelles spécifiques qui favorisent l'usage de SN sans déterminant pour référer à un personnage spécifique ? ; et ii) quelles sont les caractéristiques discursives principales des sujets ?

7.1.1.2.3. Dans le cadre inter-phrastique : influence des facteurs discursifs

À partir de l'examen de l'influence des fonctions syntaxiques sur la référence des SN sans déterminant, il semble prévisible que quelques facteurs discursifs dans le cadre inter-phrastique puissent également influencer la référence des SN sans déterminant dans la chaîne. L'ensemble des détails de chaque expression est exposé dans l'*Annexe 30*. Nous essayons de commenter ces chiffres par la comparaison des maillons français afin de savoir de quelle forme de maillons français les SN sans déterminant se rapprochent.

Premièrement, la distance moyenne entre un SN sans déterminant et son antécédent est de 2,3 phrases. D'après l'échelle de Givón (1983), la distance de « 2 – 3 phrases » correspond au « pronom tonique », catégorie permettant la continuité référentielle en anglais ; la distance de « 10 phrases » correspondant aux « descriptions définies ». La distance obtenue pour les SN sans déterminant chinois est donc plus proche de celle de « pronom tonique » et un peu plus éloignée de celle des « descriptions définies » dans cette échelle. Cette distance paraît impliquer que les lieux d'occurrences de SN sans déterminant chinois peuvent être équivalents à certains pronoms toniques anglais, ce qui peut constituer des équivalences imprévues du fait que ces deux catégories partagent peu de points communs lorsqu'il est question de leur fonction référentielle. Par ailleurs, 5 SN sans déterminant chinois sont utilisés après une longue distance de plus de 10 phrases, dont 2 après une très longue distance de plus de 20 phrases. Il semble que le SN sans déterminant est un des moyens qui peuvent servir à redémarrer une chaîne de référence.

Deuxièmement, entre un SN sans déterminant et son antécédent existent en moyenne 5,5 référents : son contexte tend à être multi-référentiel plutôt que mono-référentiel. Ce chiffre indique en quelque sorte que le contexte peut contenir d'autres référents vis-à-vis desquels les SN sans déterminant ne sont pas censés engendrer d'ambiguïté sémantique. Selon l'approche comparative, ce groupe de maillons chinois présente le chiffre le plus proche

de deux groupes français : les noms propres, avec 6,4 référents, et les SN possessifs, avec 5,7 référents (voir le tableau (6.12) du *Chapitre 6*). Celui des SN définis français s'élève à 7,1 référents, chiffre relativement éloigné de celui des SN sans déterminant chinois. En considérant ce critère contextuel, les SN sans déterminant chinois s'approchent plutôt de l'usage des noms propres et des SN possessifs que des SN définis français.

Troisièmement, en ce qui concerne la place initiale des séquences textuelles, 12 % de SN sans déterminant (26 occurrences) se situent dans la première phrase d'une séquence textuelle, dont 25 sont des paragraphes et un seul est un épisode. Cette proportion montre que les SN sans déterminant chinois peuvent initier les séquences textuelles, plus souvent aux frontières typographiques qu'aux changements spatio-temporels.

Quatrièmement, 23 % de SN sans déterminant (48 occurrences) sont utilisés dans les séquences autres que les narratives à proprement parler : 45 dans les séquences dialogales et 3 autres dans la seule séquence épistolaire dans la partie du roman étudiée ici. Par rapport à la totalité du corpus, les séquences dialogales et épistolaire (voir l'*Annexe 32 : Séquences dialogales et épistolaire en chinois*) totalisent 9110 caractères, ce qui équivaut à 10 % (9110 / 89858) du corpus chinois. La comparaison de ces deux proportions semble indiquer que les SN sans déterminant présentent un plus grand nombre d'occurrences par rapport aux séquences narratives.

7.1.1.2.4. Dans le cadre inter-phrastique : référence des SN sans déterminant à la place de sujet

Selon les interprétations référentielles des SN sans déterminant présentées au début de cette partie, le sujet est la fonction syntaxique privilégiée pour réaliser la référence spécifique. En raison de cette distinction entre le sujet et les autres fonctions, après l'observation globale de tous les SN sans déterminant, nous avons étudié en particulier le contexte discursif des SN sans déterminant en position de sujet selon les mêmes quatre facteurs discursifs susmentionnés. Une synthèse quantitative issue de l'*Annexe 31* est présentée dans le tableau (7.6) ci-dessous, qui offre une comparaison des chiffres des deux groupes de SN sans déterminant : tous les SN sans déterminant de la chaîne (colonne à gauche) et tous les SN sans déterminant en position de sujet (colonne à droite).

Tableau (7.6) : Comparaison entre tous les SN sans déterminant chinois et les SN sans déterminant sujets dans la chaîne du protagoniste

SN sans déterminant	Tous	Sujets
Distance	2,3 phrases	2,8 phrases
Réf	5,5 référents	5,6 référents
Début	12 % (26 / 210)	13 % (12 / 92)
Discours	23 % (48 / 210)	20 % (18 / 92)

La distance inter-maillonnaire est légèrement plus large pour les sujets, avec une moyenne de 2,8 phrases, contre celle de tous les SN sans déterminant étant de 2,3 phrases. Cette différence peut être due à la position du maillon dans la phrase : étant donné que le

sujet concerne les premiers mots d'une phrase, la distance ne peut pas être 0 puisque son antécédent ne peut pas se trouver dans la même phrase, comme le montre l'extrait [9] où la distance entre le SN sans déterminant 'enfant' *haizi* 孩子 et son antécédent 'il' *ta* 他 est au minimum 1 (phrase). Pourtant, les maillons d'autres fonctions syntaxiques peuvent avoir une distance de 0 puisque l'antécédent peut être le sujet de la même phrase, comme le montre l'extrait [10] : la distance entre le SN sans déterminant 'enfant' *haizi* 孩子 et son antécédent 'Christophe' *Kelisiduofu* 克利斯朵夫 est 0 car les deux maillons se situent dans la même phrase. Cet écart peut donc toujours être relié à la fonction syntaxique des maillons.

[9] 祖父 在 黄昏 散步 的 时候 常常 带着 [他] 一块儿 去。**[孩子]**
zufu *zai* *huanghun* *sanbu* *de* *shihou* *changchang* *daizhe* *ta* *yikuaier* *qu* *haizi*
grand-père PREP soir promenade DE temps souvent prenant 3P ensemble aller enfant
antécédent *SN-sans-dét-sujet*

拉着 老人 的 手 在 旁边 急急忙忙的 搬着 小 步。
lazhe *laoren* *de* *shou* *zai* *pangbian* *jijimangmangde* *banzhe* *xiao* *bu*
donnant vieux-homme DE main PREP côté se-dépêchant trottinant petit pas

'Grand-père le prenait souvent avec lui, dans ses promenades du soir. Le petit trottinait à ses côtés, en lui donnant la main.'

[10]可是 有 一 天 她 看见 曼希沃 又 殴打 [克利斯朵夫]， 抢去 了 **[孩子]** 的
keshi *you* *yi* *tian* *ta* *kanjian* *Manxiwo* *you* *ouda* *Kelisiduofu* *qiangqu* *le* *haizi* *de*
mais avoir un jour 3P voir Melchior encore battre Christophe dépouiller LE enfant DE
antécédent *SN-sans-dét-non-sujet*

钱， 便 再也 忍不住 了。
qian *bian* *zaiye* *renbuzhu* *le*
argent donc ne-plus tenir LE

'Mais un jour qu'elle vit, une fois de plus, Melchior battre Christophe et le dépouiller de son argent, elle n'y tint plus.'

Excepté cette différence de distance inter-maillonnaire, les trois autres critères présentent des résultats assez proches. En ce qui concerne le deuxième critère, le nombre de référents situés dans le proche contexte des sujets est de 5,6 en moyenne, ce qui rejoint celui des SN sans déterminant de toutes les fonctions (5,5 référents). Donc, peu importe la fonction syntaxique : les SN sans déterminant ne créent pas d'ambiguïté référentielle dans un contexte multi-référentiel. Au sujet du troisième critère, 13 % de SN sans déterminant sujet sont au début d'un paragraphe : cette proportion est très proche de celle de tous les SN sans déterminant (12 %). En raison de ces deux groupes de chiffres assez similaires, nous pouvons déduire que ni le nombre de référents concurrents, ni la place initiale de séquence textuelle, ne sont des facteurs primordiaux relativement au mode de référence des SN sans déterminant chinois. Quant au dernier critère, 20 % de SN sans déterminant sujet sont utilisés dans les séquences dialogales, donc légèrement moins que celle de tous les SN sans déterminant (23 %). Il existe une faible tendance à moins utiliser cette forme dans les dialogues que dans la narration proprement dite.

7.1.1.3. SN démonstratifs : 'ce (classificateur)' *zhe(ge)* 这 (个) / 'ce-là (classificateur)' *na(ge)* 那 (个) + nom

Le SN démonstratif est le troisième groupe de maillons chinois le plus important dans la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫, malgré une faible quantité de 13 occurrences. Cette quantité est légèrement plus élevée que celle du français dont elle reste proche avec 10 occurrences. Ce groupe de maillons chinois sont exposés dans leur contexte dans l'*Annexe 33 :*

SN démonstratifs chinois dans la chaîne 'Christophe' Kelisiduofu *克利斯朵夫*, avec les extraits parallèles français. Les maillons chinois, accompagnés des gloses françaises, sont présentés avec les indices contextuels dans le tableau (7.7) suivant, avec lesquels nous avons synthétisés trois points significatifs.

Tableau (7.7) : Caractéristiques discursives des SN démonstratifs chinois dans la chaîne du protagoniste

S	E	Maillons	Fonction	Distance	Réf	Début	Discours
I	5	'ce CL enfant' *zhe-ge-haizi* 这个孩子	CDO	TL	BP	N	N
		'toi cet âne' *ni-zhe-lüzi* 你这驴子[8]	S	0	1	N	DD
	10	'cette petite vie'*zhe-xiao-shengming* 这小生命	S	2	0	PA	N
II	13	'ce CL garçon' *zhe-ge-nanhaizi* 这个男孩子	CDN	2	4	N	N
		'ce nouveau-venu enfant' *zhe-xinlaide-haizi* 这新来的孩子	CDN	1	1	N	N
	20	'cet enfant' *zhe-haizi* 这孩子	S	0	2	N	DD
	22	'lui ce CL enfant' *ta-zhe-ge-haizi* 他这个孩子	CDO	1	3	N	N
		'ce-là CL enfant' *na-ge-haizi* 那个孩子	CDO	2	5	N	N
		'cet enfant' *zhe-haizi* 这孩子	S	3	4	N	DD
III	34	'ce CL petit homme' *zhe-ge-xiao-renwu* 这个小人物	CDN	2	8	N	N
	35	'cet air comme bohême DE petit enfant' *zhe-qise-xiang-boximiyaren-ban-de-xiaohaier* 这气色象波希米人般的小孩儿	CDO	1	3	N	N
IV	38	'ce gars' *zhe-jiahuo* 这家伙	S	0	0	N	DD
	47	'ce CL quatorze ans DE enfant' *zhe-ge-shisi-sui-de-haizi* 这个十四岁的孩子	S	1	1	N	N
Total / moyenne			46 % S	1,2	2,5	1 PA (8 %)	4 DD (31 %)

7.1.1.3.1. Trois fonctions syntaxiques des SN démonstratifs

Dans le cadre phrastique, les SN démonstratifs se présentent dans trois positions syntaxiques avec un nombre d'occurrences relativement équilibrées : 5 occurrences (46 %) à la position de sujet ('ce gars' *zhe-jiahuo* 这家伙 de l'extrait [11] ci-dessous, par exemple), 5 à celle de complément d'objet ('ce nouveau-venu enfant' *zhe-xinlaide-haizi* 这新来的孩子 de l'extrait [12], par exemple), et 3 occurrences à celle de complément du nom ('ce petit homme' *zhe-ge-xiao-renwu* 这个小人物 de l'extrait [13], par exemple).

[11]要是 可能， [他] 竟要 恨 音乐 了。曼希沃 说 [他]："**[这 家伙]** 没有 心肝， 没有
yaoshi keneng ta jingyao hen yinyue le Manxiwo shuo ta zhe jiahuo meiyou xingan meiyou
si possible 3P même haïr musique LE Melchior dire 3P ce individu sans cœur sans
SN-dém-sujet

[8] Cette expression concerne la *structure co-topicale* (双向名词句) en chinois (*cf.* Chen, 陈平, 2004) : *ni* 你 et *zhe-haizi* 这孩子 sont les deux topiques juxtaposés au début de la phrase. La même explication s'applique aussi à 'lui ce CL enfant' *ta-zhe-ge-haizi* 他这个孩子. Cette structure est utilisée très rarement dans notre corpus.

感觉。 不 知 [他] 这 种 性格 象 谁。"
ganjue bu zhi ta zhe zhong xingge xiang shui
sentiment NEG savoir son ce type caractère comme qui
'Il l'eût haïe, si c'eût été possible. Melchior disait de lui :
– Cet individu n'a pas de cœur. Il ne sent rien. Je ne sais pas de qui il tient.'

[12][克利斯朵夫] 一 来， 倒是 给 他们 解了围。 两 人 走拢来 打量
Kelisiduofu yi lai daoshi gei tamen jielewei liang ren zoulonglai dalinag
Christophe dès arrivée par-contre AUX 3Ps faire-diversion-LE deux personne se-rapprocher examiner

[这 新来的 孩子]。
zhe xinlaide haizi
ce nouveau-venu enfant
SN-dém

'L'arrivée de Christophe fit diversion. Ils se rapprochèrent pour examiner **[le nouveau venu]**.'

[13]鲁意莎 说 他们 要 叫 [可怜的 孩子] 闹笑话 了， 但 她的 反对 毫无 用处。
Luyisha shuo tamen yao jiao keliande haizi naoxiaohua le dan tade fandui haowu yongchu
Louisa dire 3Ps aller rendre pauvre enfant faire-rire LE mais son protestation sans-aucun utilité
(…) 事情 就 这样 决定 了， 裁缝 给 叫来 量 **[这 个 小 人物]** 的 尺寸。
shiqing jiu zheyang jueding le caifeng gei jiaolai liang zhe ge xiao renwu de chicun
chose donc ainsi décider LE tailleur AUX appeler mesurer ce CL petit homme DE taille
SN-dém

'En vain, la bonne Louisa protestait qu'on voulait rendre ridicule son pauvre garçon. (…) Il en fut fait ainsi, et le tailleur vint prendre mesure pour l'habit **[du petit homme]**.'

À la différence des SN sans déterminant, qui occupent la position de sujet en grande majorité, les fonctions syntaxiques des SN démonstratifs semblent assez bien réparties. Ce facteur phrastique n'exerce donc pas d'influence majeure sur la référence des SN démonstratifs chinois.

7.1.1.3.2. SN démonstratifs

Étant donné que la quantité de SN démonstratifs est faible et similaire dans les deux langues, il se peut que la plupart des maillons de cette catégorie soient traduits à partir de la même forme morpho-syntaxique française. Or, nous avons au contraire observé seulement une petite moitié d'occurrences qui correspondent à cette hypothèse : 5 SN démonstratifs chinois sur 13 sont traduits à partir des SN définis français, alors que la grande moitié d'occurrences restantes ne sont pas utilisées dans les textes parallèles, comme les exemples montrés dans les extraits [12] et [13] : 'ce nouveau-venu enfant' *zhe-xinlaide-haizi* 这新来的孩子 et 'ce CL petit homme' *zhe-ge-xiao-renwu* 这个小人物 traduisent « le nouveau venu » et « le petit homme ». Dans ces deux extraits ainsi que dans ceux contenant un SN démonstratif dans l'*Annexe 33,* le protagoniste n'est souvent pas le seul personnage présent dans le contexte, sauf un maillon en début de paragraphe (à la 3ème ligne du tableau (7.7)) et un autre dans un discours rapporté direct (à la 12ème ligne du tableau (7.7)). Autrement dit, les référents d'humains sont régulièrement nombreux : le contexte à gauche contient en moyenne 2,5 référents concurrents. Lors de la reprise du protagoniste, se pose la question de la distinction entre celui-ci et les autres personnages. Dans un tel contexte, les SN démonstratifs permettent de désigner le protagoniste plutôt que les autres personnages, sans susciter d'ambiguïté référentielle.

Puisque nous avons vu plus haut que les SN sans déterminant peuvent référer de manière définie et spécifique, nous avons également réalisé des tests en supprimant les déterminants démonstratifs chinois pour comparer les différences référentielles éventuelles

entre les SN sans déterminant et les SN démonstratifs. Sachant que le SN sans déterminant peut tout de même référer à une entité spécifique et à une entité générique, la suppression du déterminant démonstratif n'est pas censée changer la nature spécifique du référent. Autrement dit, les lecteurs chinois peuvent toujours comprendre que le référent est spécifique ou défini, comme c'est le cas de certains SN démonstratifs : après la suppression du déterminant démonstratif de l'exemple 'ce nouveau-venu enfant' *zhe-xinlaide-haizi* 这新来的孩子 de l'extrait [12], le SN sans déterminant 'nouveau-venu enfant' *xinlaide-haizi* 新来的孩子 de l'extrait [14]* parvient toujours à référer au même personnage et n'engendre pas non plus de difficultés de compréhension ou d'ambiguïtés pour établir la coréférence avec son antécédent 'Christophe' *Kelisiduofu* 克利斯朵夫 :

[14]*[克利斯朵夫]	一	来，	倒是	给	他们	解了围。		两	人	走拢来		打量	
Kelisiduofu	*yi*	*lai*	*daoshi*	*gei*	*tamen*	*jielewei*		*liang*	*ren*	*zoulonglai*		*daliang*	
Christophe	dès	arrivée	par-contre	AUX	3Ps	faire-diversion-LE		deux	personne	se-rapprocher		examiner	
[新来的	**孩子]**	。											
xinlaide	*haizi*												
nouveau-venu	enfant												
SN-sans-dét													

Pourtant, pour diverses raisons, d'autres SN démonstratifs ne supportent pas ce type de suppression. La phrase [15]* est agrammaticale en chinois dans ce contexte de discours rapporté direct : un SN sans déterminant chinois ne peut pas désigner une personne autre que les participants du dialogue : le déterminant démonstratif est oblitatoire en chinois pour réaliser une référence exophorique. La phrase [16]* engendre des difficultés aux lecteurs pour constituer la coréférence entre 'petit homme' *xiao-renwu* 小人物 et son antécédent 'pauvre enfant' *keliande-haizi* 可怜的孩子 en raison de la suppression du déterminant démonstratif comme marqueur de reprise endophorique, et en raison du changement de la tête lexicale dans l'anaphore nominale infidèle qui ajoute une difficulté supplémentaire à la constitution de la coréférence entre les deux expressions.

[15]*曼希沃	说	[他]：	"**[家伙]**	没有	心肝，	没有	感觉。"								
Manxiwo	*shuo*	*ta*	*jiahuo*	*meiyou*	*xingan*	*meiyou*	*ganjue*								
Melchior	dire	3P	individu	sans	cœur	sans	sentiment								
			SN-sans-dét												
[16]*鲁意莎	说	他们	要	叫	**[可怜的**	**孩子]**	闹笑话	了，	但	她的	反对		毫无	用处。	
Luyisha	*shuo*	*tamen*	*yao*	*jiao*	*keliande*	*haizi*	*naoxiaohua*	*le*	*dan*	*tade*	*fandui*		*haowu*	*yongchu*	
Louisa	dire	3Ps	aller	rendre	pauvre	enfant	faire-rire	LE	mais	son	protestation		sans-aucun	utilité	
(…) 裁缝	给	叫来	量	**[小**	**人物]**	的	尺寸。								
caifeng	*gei*	*jiaolai*	*liang*	*xiao*	*renwu*	*de*	*chicun*								
tailleur	AUX	appeler	mesurer	petit	homme	DE	taille								
				SN-sans-dét											

Les tests de la suppression du déterminant démonstratif dans toutes les occurrences révèlent que la plupart ne supportent pas cette modification. En premier lieu, dans 4 cas localisés dans les séquences dialogales, les phrases deviennent agrammaticales. En deuxième lieu, dans 6 cas localisés dans les anaphores nominales infidèles, les phrases créent des difficultés de compréhension et d'établissement de la coréférence avec l'antécédent, comme la confusion entre un référent spécifique et un référent générique. En troisième lieu, dans les 3 cas restants, les SN démonstratifs sont facultatifs en termes de coréférence et peuvent être remplacés par les SN sans déterminant sans susciter de changements grammaticaux ou coréférentiels majeurs.

7.1.1.3.3. Démonstratif *endophorique* et démonstratif *endo-exophorique*

Comme nous l'avons vu, les démonstratifs peuvent être utilisés de manière endophorique dans l'anaphore et dans la cataphore mais aussi de manière exophorique dans la déixis, ou encore de manière endo-exophorique dans l'ana-déixis. Dans le cadre du roman, il ne semble pas pertinent de parler de la déixis à proprement parler en raison de l'absence d'une vraie situation d'énonciation de l'oral. Toutefois, les séquences dialogales semblent pouvoir établir un contexte différent de celui des séquences narratives. À partir de cette remarque, les SN démonstratifs peuvent être subdivisés en groupe de *démonstratifs endophoriques* dans les séquences narratives, et en groupe de *démonstratifs endo-exophoriques* dans les séquences dialogales. D'après ces critères, le résultat de recensement révèle que 4 SN démonstratifs sur 13 sont endo-exophoriques, comme le montre 'cet individu' *zhe-jiahuo* 这家伙 de l'extrait [11] du plus haut ; les 9 restants sont endophoriques, comme le montrent les extraits [12] et [13]. Les endo-exophoriques sont traduits à partir de la même forme démonstrative française. Les endophoriques, cependant, ne manifestent pas cette équivalence entre les deux langues : les maillons de ce sous-groupe peuvent avoir la même interprétation référentielle des SN définis ou démonstratifs en français.

Cette comparaison nous permet de déduire que les SN démonstratifs chinois ont deux performances distinctives selon les contextes discursifs du roman : les démonstratifs endo-exophoriques peuvent être interprétés de même manière que celle des démonstratifs français, alors que les démonstratifs endophoriques se rapprochent en même temps des endophoriques en deux formes françaises : les définis et les démonstratifs.

7.1.1.4. Deux catégories nominales de plus faible quantité : SN possessifs et SN indéfinis

Comme le montre le tableau (7.2) au début de cette partie, les SN possessif et indéfini sont peu utilisés dans la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫, avec respectivement 3 et 4 occurrences. Or, pour cette chaîne de 3817 maillons, ces nombres sont proportionnellement très peu significatifs : les occurrences des deux catégories peuvent même être considérées comme « exceptionnelles ». Le faible emploi de ces formes peut être expliqué par plusieurs raisons. Premièrement, elles sont peut-être peu utilisées dans le texte-source en français et donc les formes morpho-syntaxiques sont traduites de manière fidèle vers le chinois. Deuxièmement, il se peut qu'elles ne permettent pas de référer à une entité spécifique et définie en chinois ; elles sont donc remplacées par d'autres formes morpho-syntaxiques. Troisièmement, une des fonctions référentielles des SN possessifs français est d'exprimer la relation inaliénable, alors que celle-ci peut être exprimée par les SN sans déterminant en chinois (*cf.* Liu, 刘丹青, 2002 ; Dong, 董秀芳, 2010 ; Xiong, 熊岭, 2012 ; Wan, 完权, 2016). Pour cette raison, une partie de SN possessifs chinois peuvent être remplacés par les SN sans déterminant et les SN définis, malgré le même type d'usage inaliénable de ces catégories en français. Quatrièmement, à part les raisons susmentionnées, d'autres motifs syntaxiques ou discursifs peuvent également contraindre leur emploi comme maillon de la chaîne d'un référent spécifique en chinois. Nous examinerons ces derniers motifs à travers les chiffres et les exemples du corpus.

7.1.1.4.1. SN possessifs : *wode* 我的 / *nide* 你的 / *tade* 他的 + nom

Le groupe des SN possessifs concerne les expressions composées d'un déterminant possessif, d'un nom, et de modifieurs facultatifs antéposés au nom, représentées par la formule suivante :

Déterminant possessif – (modifieur(s)) – nom

En raison de leur très faible quantité (3 maillons seulement), il nous semble peu pertinent de résumer leur usage régulier par telle ou telle tendance principale et tel ou tel contexte privilégié ; nous choisissons donc de les examiner au cas par cas, au lieu de les étudier de manière quantitative ou proportionnelle. Les 3 maillons en question se situent dans les extraits [17] – [19]. Les caractéristiques discursives de chaque maillon sont synthétisées dans le tableau (7.8) qui suit.

[17]她 瞅着 [他]， 又 惭愧 又 欢喜的 笑了笑:
ta chouzhe ta you cankui you huanxide xiaolexiao
3P contemplant 3P et confus et ravi rire-LE

"哦， **[我的 小 乖乖]**， [你]多 难看， 多 难看， 我 多 疼 [你]！"
O wode xiao guaiguai ni duo nankan duo nankan wo duo teng ni
oh mon petit chéri 2P que laid que laid 1P que aimer 2P
SN-poss

'Elle le contempla avec un sourire confus et ravi :
– Oh ! **[mon pauvre petit]**, dit-elle toute honteuse, que tu es laid, que tu es laid, comme je t'aime !'

[18]母亲 紧紧搂着 [他]， 轻轻的 说:
muqin jinjinlouzhe ta qingqingde shuo
mère pressant 3P doux dire
"得啦， 得啦， 别 哭 了， [我的 小 耶稣]， **[我的 小 金鱼]**……"
dela dela bie ku le wode xiao Yesu wode xiao jinyu
c'est-bon c'est-bon NEG pleurer LE mon petit jésus mon petit poisson-d'or
SN-poss

'La mère le presse contre elle, avec de petits mots :
« C'est fini, c'est fini, ne pleurons plus, mon jésus, **[mon petit poisson d'or]**…'

[19]大公爵 没有 忘记 **[他的 钢琴师]**： 这 并不是 说， 以 钢琴师 的 名义 应有的
dagongjue meiyou wangji tade gangqinshi zhe bingbushi shuo yi gangqinshi de mingyi yingyoude
grand-duc NEG oublier son pianiste ce NEG dire à pianiste DE titre devant
SN-poss

一点儿 月俸 会 准起 支付， 那 是 永远 要去 催讨 的。
yidianer yuefeng hui zhunqi zhifu na shi yongyuqn yaoqu cuitao de
modique pension pouvoir exactement payer cela être toujours devoir réclamer DE

'Le grand-duc n'oubliait pas **[son pianiste ordinaire]** : non que la modique pension attribuée pour ce titre fût exactement payée, – il fallait toujours la réclamer.'

Tableau (7.8) : Caractéristiques discursives des SN possessifs chinois dans la chaîne du protagoniste

S	E	Maillons	Fonction	Distance	Réf	Début	Discours
I	1	'mon petit chéri' *wode-xiao-guaiguai* 我的小乖乖	A	0	1	N	DD
	3	'mon petit poisson d'or' *wode-xiao-jinyu* 我的小金鱼	A	0	1	N	DD
IV	37	'son pianiste' *tade-gangqinshi* 他的钢琴师	CDO	1	2	PA	N

Malgré la grande variation qui caractérise la catégorie grammaticale des déterminants possessifs, seules 2 formes figurent dans la chaîne : *wode* 我的 pour toute la variation de genres et de nombres de la première personne ('mon / ma / mes' en français) et *tade* 他的 pour celle du genres et du nombre de la troisième personne masculine ('son / sa / ses' en français). En raison du genre masculin et du nombre au singulier du protagoniste, les seuls équivalents français sont « mon » et « son ». Les 2 SN possessifs de première personne avec 'mon' *wode* 我的 partagent quelques points communs et le seul possessif de troisième personne avec 'son' *wode* 他的 présente d'autres caractéristiques complètement différentes.

Les 2 possessifs de première personne se situent dans les séquences dialogales en position syntaxique d'appellatif, comme le montrent 'mon petit chéri' *wode-xiao-guaiguai* 我的小乖乖 et 'mon petit poisson d'or' *wode-xiao-jinyu* 我的小金鱼 dans les extraits [17] et [18]. Ils concernent un changement de tour de parole du narrateur à un personnage en relation directe, familiale et affective, avec le protagoniste, *i.e.* sa mère 'Louisa' *Luyisha* 鲁意莎. Le troisième SN possessif 'son pianiste' *tade-gangqinshi* 他的钢琴师 dans l'extrait [19] se manifeste dans une séquence narrative au début d'un paragraphe, occupant la place de complément d'objet de la phrase. Il a également trait à un changement de tour de parole, mais de manière indirecte, sans discours rapporté direct, entre le narrateur et le personnage 'le grand-duc' *dagongjue* 大公爵. La relation entre ces deux personnages est moins étroite et affective que celle des possessifs de première personne.

En ce qui concerne les têtes lexicales contenues dans les 3 SN possessifs, aucune n'est identique et chacune est suscitée par une raison spécifique. La première tête lexicale 'chéri' *guaiguai* 乖乖 semble d'être une adresse dans une relation intime, familiale et affective. La troisième 'pianiste' *gangqinshi* 钢琴师 est une désignation par la catégorie socio-professionnelle du protagoniste. Le deuxième SN possessif 'mon petit poisson d'or'[9] *wode-xiao-jinyu* 我的小金鱼 dans l'extrait [18] est encore plus particulier : c'est une traduction littérale d'une expression figée et métaphorique française qui n'existe pas en chinois. Cette expression appellative peut être considérée comme un *énoncé situationnel* en français selon la définition de Marque-Pucheu (2011 : 5), à savoir que ce type d'expressions possèdent leurs interprétations sémantico-référentielles très particulières par rapport aux autres SN possessifs :

> *Leur [les énoncés situationnels] pouvoir référentiel est généralement affecté dans la mesure où ces expressions sont convenues (Vous m'en direz des nouvelles (E / * fraîches)), et souvent leur interprétation met en jeu une atténuation du sens et / ou un sens décalé : N'hésitez pas est proche de l'invitation, Faut pas chercher à comprendre note une « réaction […] devant l'attitude d'un tiers que l'on ne comprend pas ou que l'on n'approuve pas » (Bidaud, 2002, n° 157) et Vous n'auriez pas dû fait office de remerciement. (Marque-Pucheu, 2011 : 5)*

Selon cette définition, l'interprétation référentielle des énoncés situationnels n'est pas forcément reliée au sens littéral d'expression. L'expression française « mon petit poisson d'or » ne réfère en aucun cas à un type de poisson possédé par quelqu'un, mais elle a pour

[9] L'expression 'mon petit jésus' *wode-xiao-Yesu* 我的小耶稣 appartient également aux *énoncés situationnels*. Voir son interprétation référentielle dans la section des noms propres modifiés.

seul but d'appeler quelqu'un en montrant une relation chère et intime avec lui. La traduction littérale vers le chinois peut ainsi être interprétée dans ce sens.

En effet, les SN possessifs chinois sont très peu utilisés par rapport aux 37 occurrences du français : 3 SN possessifs français sont traduits par la forme possessive ; les autres ont d'autres équivalents formels en chinois. Aussi bien dans les séquences narratives que dans les dialogales, cette différence d'emploi du déterminant possessif nous permet de déduire le fait que les référents autres que le protagoniste sont présentés en chinois comme des référents isolés sans relation explicite avec le protagoniste, tandis que ceux du français montrent une relation au protagoniste beaucoup plus forte et visuelle étant donné la forme possessive exprimée. Par exemple, le SN sans déterminant 'petit-fils' *suner* 孙儿, équivalent au SN possessif « son petit-fils » dans la séquence narrative de l'extrait [20], établit la relation familiale par la tête lexicale seulement au lieu du déterminant possessif ; la même explication s'applique au SN sans déterminant 'enfant' *haizi* 孩子, équivalent au SN possessif « mon enfant » dans la séquence dialogale de l'extrait [21].

[20]这 时候， [他] 瞥见 祖父 又 高兴 又 不好意思的， 站 在 走廊 里 包厢
zhe shihou ta piejian zufu you gaoxing you buhaoyiside zhan zai zoulang li baoxiang
ce moment 3P apercevoir grand-père et ravi et honteux debout PREP corridor dedans loge

进口 的 地方； 他 很 想 进来 说 几 句 话， 可是 不 敢， 因为 人家 没
jinko de difang ta hen xiang jinlai shuo ji ju hua keshi bu gan yinwei renjia mei
entrée DE endroit 3P bien vouloir entrer dire quelque CL mot mais NEG oser car on NEG

招呼 他， 只 能 远远的 看着 [孙儿] 的 光荣， 暗中 得意。
zhaohu ta zhi neng yuanyuande kanzhe suner de guangrong anzhong deyi
adresser 3P seulement pouvoir loin regardant petit-fils DE gloire discrètement se-contenter
SN-sans-dét

'À ce moment, il aperçut dans le corridor, à l'entrée de la loge, grand-père, rayonnant et honteux, qui aurait bien voulu se montrer et dire aussi son mot, mais qui n'osait, parce qu'on ne lui avait pas adressé la parole : il jouissait de loin de la gloire de [**son petit-fils**].'
SN-poss

[21][克利斯朵夫] 犹豫 了 一会， 说："哎， **舅舅**， 他 现在 在 哪儿 呢？"
Kelisiduofu youyu le yihui shuo ai jiujiu ta xianzai zai naer ne
Christophe hésiter LE un-instant dire ah oncle 3P maintenant PREP où INTERR

"[**孩子**]， 他 和 上帝 在 一起。"
haizi ta he shangdi zai yiqi
enfant 3P avec seigneur PREP ensemble
SN-sans-dét

'Christophe hésita :
– Ah ! oncle, demanda-t-il, où est-il maintenant ?
Gottfried répondit :
– Il est avec le Seigneur, [**mon enfant**].'
SN-poss

La disparition du possessif en chinois est susceptible d'affaiblir la relation entre le protagoniste et les autres référents. Sans marqueur possessif explicite, ces derniers ont tendance à être présentés comme entités indépendantes du protagoniste ou des autres entités.

7.1.1.4.2. SN indéfinis : '(un) classificateur' *yi(ge)* (一)个 + nom

Les SN indéfinis sont présents en très faible quantité dans les deux langues : 4 occurrences en chinois et 8 en français. Tout comme les SN possessifs, il n'est pas intéressant d'analyser cette faible quantité de maillons de manière quantitative et proportionnelle ; nous les étudierons donc au cas par cas pour voir les conditions spécifiques de l'emploi de ces formes exceptionnelles. Les 4 SN indéfinis chinois sont exposés dans les extraits [22] – [25].

[22]他们俩 很 投机。 老人 非常 喜欢 [孙子]；有 [个 **愿意** 听 他 说话 的
tamenlia hen touji laoren feichang xihuan sunzi you ge yuanyi ting ta shuohua de
3P-deux très s'entendre vieux-homme très aimer petit-fils avoir CL *SN-ind* vouloir écouter 3P parler DE

人] 更 使 他 快乐。
ren geng shi ta kuaile
personne plus rendre 3P content
'Ils s'entendaient ensemble. Le vieux avait une immense affection pour son petit-fils ; et ce lui était une joie de trouver en lui [**un public complaisant**].'

[23][克利斯朵夫] 的 脸 涨得通红。 小姑娘 撅着 嘴 轻轻的 和 她的 兄弟 说：
Kelisiduofu de lian zhangdetonghong xiaoguniang juezhe zui qingqingde he tade xiongdi shuo
Christophe DE visage cramoisi fillette faisant bouche doucement avec son frère dire

"[他] 是 [个 **穷** **小子**]。"
ta shi ge qiong xiaozi
3P être CL *SN-ind* pauvre garçon
'Christophe devint cramoisi. La fillette fit la moue et souffla à son frère – Christophe l'entendit, – que c'était [**un petit pauvre**].'

[24][孩子] 这样的 躲 在 钢琴 后边 物我两忘，— 直到 [他] 忽然 觉得 蚂蚁
haizi zheyangde duo zai gangqin houbian wuwoliangwang zhidao ta huran juede mayi
enfant tel se-cacher PREP piano derrière tout-oublier jusque 3P brusquement sentir fourmi

爬上 [他] 大腿 的 时候，才 记起 [自己] 是 [个 **小** **孩子**]，[指甲 乌黑，把 **鼻子**
pashang ta datui de shihou cai jiqi ziji shi ge xiao haizi zhijia wuhei ba bizi
monter son jambe DE temps alors se-souvenir soi être CL *SN-ind* petit enfant ongle noir BA nez

望 **墙** 上 **轻轻** **挨着**，双 手 **攀着** 脚 的 小 孩子]。
wang qiang shang qingqing aizhe shuang shou panzhe jiao de xiao haizi
vers mur dessus doucement contre deux main tenant pieds DE petit enfant
'Ainsi, l'enfant restait, oublié, oubliant, dans le coin du piano, – jusqu'à ce que brusquement il sentît des fourmis lui monter dans les jambes. Et il se souvenait alors qu'il était [**un petit garçon, avec des ongles noirs, et qu'il frottait son nez contre le mur, en tenant ses pieds entre ses mains**].'

[25]从 那 天 起，因为 [他] 是 [个 **作曲家**]，[他] 就 开始 作曲 了。连 字
cong na tian qi yinwei ta shi ge zuowujia ta jiu kaishi zuoqu le lian zi
à-partir-de ce-là jour depuis car 3P être CL *SN-ind* compositeur 3P donc commencer composer LE même lettre

还不怎么 写 得起来，[他] 已经 在 家用 账簿 上 撕下 纸片，
haibuzenme xie deiqilai ta yijing zai jiayong zhangbu shang sixia zhipian
à-peine écrire AUX 3P déjà PREP ménage cahier-de-compte dessus arracher lambeau-de-papier

涂着 蝌蚪 似的 音符 了。
tuzhe kedou side yinfu le
griffonnant têtard CMP croche LE
'À partir de ce jour, puisqu'il était [**un compositeur**], il se mit à composer. Avant de savoir à peine ses lettres, il s'évertua à griffonner des noires et des croches sur des lambeaux de papier, qu'il arrachait aux cahiers de comptes du ménage.'

En ce qui concerne les composants de ce groupe de maillons, aucun SN indéfini ne contient le déterminant indéfini chinois *yi* — 'un / une', alors que ses équivalents français 'un'

et 'une' sont considérés comme un des marqueurs prototypiques des SN indéfinis français dans la grammaire. Dans toutes les occurrences obtenues, c'est le classificateur chinois *ge* 个 qui occupe seul le rôle de marqueur de SN indéfinis chinois : contrairement aux langues absentes du système des classificateurs, le chinois possède un dispositif des classificateurs obligatoires relatifs aux noms, occupant le même rôle sémantique et grammatical que celui des articles indéfinis singuliers.

Tous les SN indéfinis sont traduits à partir de ceux du français ; cette catégorie semble en quelque sorte équivalente dans les deux langues. Toutefois, cette équivalence s'établit sur la base de l'interprétation référentielle du déterminant indéfini *yi* — 'un / une' en chinois. Or, comme le constate Paris (1981 : 83), le chinois expose une interprétation référentielle relativement plus limitée que celle des articles indéfinis français « un / une », car l'article indéfini *yi* — ne marque pas le générique. C'est une des raisons pour lesquelles les SN indéfinis chinois sont légèrement moins nombreux que ceux du français. D'autres raisons et conditions spécifiques qui influencent l'usage des SN indéfinis sont susceptibles de se trouver dans le cadre intra-phrastique et inter-phrastique. Ces caractéristiques sont recensées dans le tableau (7.9) qui suit.

Tableau (7.9) : Caractéristiques discursives des SN indéfinis chinois dans la chaîne du protagoniste

S	E	Maillons	Fonction	Réf	Début	Discours
I	8	'CL vouloir écouter 3P parler DE personne' *ge-yuanyi-ting-ta-shuohua-de-ren* 个愿意听他说话的人	CDO	0	N	N
II	13	'CL pauvre gars' *ge-qiong-xiaozi* 个穷小子	A	0	N	DD
	22	'CL petit enfant' *ge-xiao-haizi* 个小孩子	A	0	PA	N
III	31	'CL compositeur' *ge-zuoqujia* 个作曲家	A	0	EP	N

Dans le cadre phrastique, les fonctions syntaxiques occupées par les SN indéfinis semblent assez particulières : contrairement à tous les groupes de maillons nominaux présentés, aucun maillon n'est sujet de phrase. 1 SN indéfini 'CL vouloir écouter 3P parler DE personne' *ge-yuanyi-ting-ta-shuohua-de-ren* 个愿意听他说话的人 se situe dans la structure existentielle initiée par 'il y a' *you* 有 (extrait [22]). Dans la grammaire traditionnelle du chinois, la structure existentielle *有 'il y a' + SN indéfini* permet d'introduire un nouveau référent non dominant comme topique ou référent saillant (*cf*. Huang, 1987, cité par Xu, 许余龙, 2005b : 123 ; Xu, 徐赳赳, 2005 ; Yue, 乐耀, 2010 ; Zhang, 张旺熹, 2013 ; Fang, 方梅, 2018). Ainsi le SN indéfini de l'extrait [22] est utilisé dans le but de réintroduire le protagoniste au milieu de la narration selon le point de vue d'un autre personnage.

Les 3 autres SN indéfinis sont utilisés à la suite du verbe copule *shi* 是 'être' dans la structure attributive pour re-catégoriser le protagoniste pour ses caractéristiques 'pauvre gars', 'petit enfant' et 'compositeur', comme le montrent les extraits [23] – [25].

Contrairement à la fonction référentielle des SN indéfinis dans la structure existentielle, celle des attributs nominaux semble plus discutable. Avant de voir les exemples du corpus en chinois, rappelons que les attributs nominaux français peuvent exprimer l'état ou la qualité de leur sujet, en tant qu'expressions non référentielles. Les linguistes chinois, notamment Chen (陈平, 2015) et Li (李先银, 2016), constatent également que les attributs nominaux

chinois sont susceptibles d'être qualitatifs plutôt que référentiels. Rappelons également que nous avons appliqué une méthode d'annotation de tous les attributs nominaux comme expressions référentielles : il existe donc un risque d'annoter les attributs nominaux qualitatifs comme référentiels. Par conséquent, nous pouvons douter du fait que les 3 attributs nominaux que nous avons annotés soient les vraies expressions référentielles : ne sont-ils pas plutôt qualitatifs ? Par exemple, le SN indéfini 'CL pauvre garçon' *ge-qiong-xiaozi* 个穷小子 de [23] se situe dans le discours rapporté direct d'un personnage. D'après ce contexte, c'est un commentaire subjectif d'un autre personnage qui cherche à qualifier et à redéfinir la classe du protagoniste de son point de vue. Le SN indéfini 'CL compositeur' *ge-zuoqujia* 个作曲家 de [25] est un auto-commentaire subjectif du protagoniste lui-même qui s'inscrit dans la catégorie socio-professionnelle « compositeur » pour démontrer son talent et ses efforts consacrés à la composition. Ces deux expressions peuvent avoir deux interprétations : elles sont non référentielles lorsque l'objectif est de qualifier les sujets de phrase ; elles sont référentielles lorsque le but est de renvoyer à la classe générique des individus qui partagent ces caractéristiques. Dans aucun cas ces attributs nominaux ne sont coréférentiels de leur sujet de phrase, qui renvoie à un personnage spécifique. Il nous semble donc que ces SN indéfinis chinois ne font pas partie de la chaîne du protagoniste.

Les maillons précédant et suivant des SN indéfinis sont des pronoms personnels et des SN de différentes têtes lexicales et des SN modifiés, puisque le changement de la tête lexicale et les modifieurs permettent de mettre l'accent sur d'autres caractéristiques du protagoniste.

7.1.1.5. Bilan

Après avoir effectué l'examen détaillé de chaque catégorie de maillons nominaux, nous récapitulons leurs caractéristiques par le biais des approches quantitative, qualitative et comparative franco-chinoises. Le nombre et la proportion de chaque catégorie en chinois sont extrêmement différents les uns des autres. Les deux grands groupes de maillons nominaux sont les noms propres et les SN sans déterminant, respectivement avec 410 et 210 maillons. Les trois petits groupes n'ont que 13, 3 et 4 maillons, correspondant aux SN démonstratifs, SN possessifs et SN indéfinis. Les paramètres de chaque groupe de maillons sont présentés de la même manière avec les mêmes définitions de critères que celles du français (voir aussi la section *6.1.2. Divers maillons nominaux* du *Chapitre 6*) dans le tableau (7.10).

Tableau (7.10) : Bilan des maillons nominaux chinois dans la chaîne du protagoniste

Catégorie	Nbre	Fonction	Distance	Réf	Début	Discours
Nom propre[10]	410	70 % S	1,8 (1 LG)	6,4 (1 BP)	30 % (5 PA, 1 EP)	0
SN sans déterminant	210	44 % S	2,3 (3 LG, 2 TL)	5,5 (5 BP)	12 % (25 PA, 1 EP)	23 % (45 DD, 3 DL)
SN démonstratif	13	46 % S	1,2	2,5	8 % (1 PA)	31 % (4 DD)
SN possessif	3	0 S	0,3	1,3	33 % (1 PA)	67 % (2 DD)
SN indéfini	4	0 S	0,3	0	50 % (1 PA, 1 EP)	25 % (1 DD)

[10] Ce sont les paramètres obtenus dans l'échantillon de 20 occurrences du tableau (7.4).

En raison de ces nombres considérablement éloignés, tous les maillons ne présentent pas la même importance pour la chaîne : les noms propres et les SN sans déterminant sont beaucoup plus réguliers par rapport aux autres catégories qui sont exceptionnelles, comme les SN démonstratifs, les SN possessifs et les SN indéfinis. Malgré ces chiffres, il semble que certains paramètres liés aux trois petits groupes soient aléatoires et peu représentatifs en raison de leurs occurrences trop peu nombreuses. Les tendances reflétées par ces proportions ne peuvent pas être considérées comme des régularités reproductibles.

- *Fonction*

Les fonctions syntaxiques occupées par chaque groupe de maillons sont diverses. Le sujet est la fonction privilégiée pour la totalité des maillons, concernant la plus grande partie de ceux-ci, alors que la petite moitié de maillons restants engagent toutes les autres fonctions syntaxiques. Mais, comme on peut le voir dans le tableau (7.10), les proportions s'éloignent beaucoup entre les différentes catégories. Le nom propre est le seul groupe dont les sujets occupent une grande partie (70 %). Pour les SN sans déterminant et les SN démonstratifs, les sujets concernent une petite moitié de maillons, respectivement 44 % et 46 %. Les SN possessifs et les SN indéfinis présentent des proportions radicalement différentes, puisqu'aucun maillon de ces deux formes n'est sujet de phrase. La fonction syntaxique de sujet n'est dominante que pour les SN sans déterminant et les SN démonstratifs et non pour les autres formes.

- *Distance*

La distance inter-maillonnaire moyenne des maillons nominaux chinois est assez courte pour tous les groupes. Celle des SN sans déterminant est de 2,3 phrases, étant la plus longue de toutes les catégories ; celle des noms propres n'est que de 1,8 phrases, à la deuxième place. Ce sont aussi les deux catégories qui peuvent être utilisées après de longues distances : les noms propres (1 fois) et les SN sans déterminant (5 fois). Ce sont donc les deux catégories employées pour réactiver le référent dominant dans le roman-fleuve. Trois autres catégories ne sont jamais utilisées après une longue distance : les SN démonstratifs, les SN possessifs et les SN indéfinis. Ce sont donc les trois catégories nominales employées en chinois pour maintenir la continuité de la chaîne dominante dans le roman-fleuve. La distance entre un SN possessif et son antécédent est extrêmement faible, à 0,3 phrase, c'est-à-dire que cette catégorie est régulièrement utilisée dans la même phrase que leur antécédent.

- *Référents* (« réf »)

Le nombre moyen de référents concurrents va en ordre décroissant, du plus élevé, celui des noms propres (6,4 référents), en passant par celui des SN sans déterminant (5,5 référents), celui des SN démonstratif (2,5 référents) en finissant par celui des SN possessifs (1,3 référents). Ces formes présentent une tendance à être utilisées dans le contexte multi-référentiel et à pouvoir éviter des risques d'ambiguïtés référentielles.

- *Début*

Le *début* coïncide avec des changements typographiques et spatio-temporels dans le roman-fleuve. La comparaison des proportions de maillons à ces endroits dans les différentes chaînes ne donne pas de résultat très distinct, puisque toutes les formes présentent quelques occurrences qui coïncident avec le début des séquences textuelles. La proportion la plus élevée est celle des SN indéfinis, avec 50 %, mais il ne s'agit que de 2 occurrences. Si l'on tient compte de la quantité d'occurrences et non de la proportion, il nous semble que ce sont les SN sans déterminant qui sont le plus couramment utilisés pour initier une séquence textuelle : leur proportion n'atteint que 12 % mais 26 occurrences sont concernées.

- *Discours*

Comme le montre l'*Annexe 32*, les séquences dialogales concernent environ 10 % du corpus en chinois. Cette proposition peut servir de seuil pour dire si le contexte dialogal ou narratif est favorable ou défavorable pour l'usage de telle ou telle forme de maillons. De manière générale, les noms propres exceptés, tous les maillons nominaux sont plus souvent employés dans les dialogues que dans les narrations : il s'agit de 23 % de SN sans déterminant, 31 % de SN démonstratifs, 67 % de SN possessifs et 25 % de SN indéfinis. Cela signifie que ce sont les autres maillons non nominaux qui sont probablement plus souvent utilisés dans les narratives.

7.1.2. Modifieurs des maillons nominaux

Après avoir présenté les maillons nominaux selon la catégorie de nom et de déterminant, notre observation s'oriente vers les composants facultatifs des maillons nominaux, à savoir les *modifieurs*. Ce groupe totalise 75 maillons, quantité plus importante que celle des groupes déjà mentionnés. Tous les maillons modifiés sont montrés avec les extraits parallèles français dans l'*Annexe 34 : Noms communs modifiés chinois dans la chaîne 'Christophe'* Kelisiduofu *克利斯朵* et l'*Annexe 35 : Noms propres modifiés chinois dans la chaîne 'Christophe'* Kelisiduofu *克利斯朵夫*. Nous présenterons d'abord les composants et les sous-types de ce groupe de maillons dits *modifiés*, et ensuite leurs spécificités référentielles en chinois dans la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫 par comparaison avec celles du français.

7.1.2.1. Trois composants des maillons nominaux modifiés

En chinois, les maillons nominaux modifiés dans la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫 peuvent être formulés par la composition suivante avec trois éléments, dont les facultatifs sont mis entre parenthèses :

(Déterminant) – modifieur(s) – nom

Le premier composant est *déterminant*. Il est mis entre parenthèse puisqu'il n'est pas obligatoire vu la grande quantité de noms propres (par exemple, 'cher Christophe' *qinaide-Kelisiduofu* 亲爱的克利斯朵夫) et de SN sans déterminant (par exemple, 'cher petit-fils' *qinaide-sunzi* 亲爱的孙子).

Le deuxième composant est *modifieur(s)* : il concerne tous les composants grammaticalement facultatifs dont la suppression n'affecte pas la structure syntaxique principale. Un maillon nominal chinois peut porter un seul ou plusieurs modifieurs, dont la seule position grammaticalement acceptable est antéposée au nom, contrairement à ceux du français avec deux positions dont la plupart sont postposés au nom.

Selon le critère formel des marqueurs grammaticaux, la typologie des sous-types de modifieurs chinois est moins évidente que celle du français, où les sous-types de modifieurs sont marqués différemment. Par exemple, les adjectifs sont marqués par les accords de genre et de nombre ; les subordonnées relatives le sont par les pronoms relatifs et les groupes prépositionnels par les prépositions. En revanche, tous les modifieurs chinois sont, soit directement antéposés au nom, soit reliés par l'auxiliaire général et polyvalent *de* 的 'DE'[11] . Le premier cas concerne les adjectifs monosyllabiques (par exemple, 'pauvre gars' *qiong-xiaozi* 穷小子 de l'extrait [26])[12], et le second les adjectifs dyssyllabiques (par exemple, 'et honteux et heureux DE enfant' *you-xiu-you-xi-de-haizi* 又羞又喜的孩子 de l'extrait [27]) ainsi que les propositions subordonnées relatives (par exemple, 'cramponnant 3P DE Christophe' *pinming-zhuazie-ta-de-Kelisiduofu* 拼命抓着他的克利斯朵夫 de l'extrait [28])[13].

[26]两 个 有钱的 孩子， 突然 对 **[穷 小子]** 起 了 一 种 儿童的 、
liang ge youqiande haizi turan dui qiong xiaozi qi le yi zhong ertongde
deux CL riche enfant brusquement pour pauvre petit prendre LE un CL enfantin
adj-monosyllabique-nom

残忍的 、莫名其妙的 反感， 看 [他] 默不作声 更 大胆 了， 想 用 什么 好玩的
canrende momingqimiaode fangan kan ta mobuzuosheng geng dadan le xiang yong shenme haowande
cruel inconnu antipathie voir 3P silencieux plus enhardi LE penser utiliser quelque amusant

方法 折磨 [他] 。
fangfa zhemo ta
moyen tourmenter 3P

'Enhardis par son silence, les deux petits riches, qui avaient pris brusquement pour le petit pauvre une de ces antipathies d'enfant, cruelles et sans raison, cherchèrent quelque moyen amusant de le tourmenter.'

[27]哈斯莱 一边 笑 一边 亲自 替 **[又 羞 又 喜 的 孩子]** 擤鼻子 。 他 把 [他]
Hasilai yibian xiao yibian qinzi ti you xiu you xi de haizi xingbizi ta ba ta
Hassler et rire et personnel pour et honteux et heureux DE enfant moucher 3P BA 3P
adj-dyssyllabique-DE-nom

放 在 地 下， 拉 [他] 到 桌子 旁边， 把 糕饼 塞满 了 [他的] 口袋 。
fang zai di xia la ta dao zhuozi pangbian ba gaobing saiman le tade koudai
déposer PREP terre dessous mener 3P vers table côté BA gâteau bourrer LE son poche

'Hassler riait, et il moucha lui-même l'enfant honteux et heureux. Il le déposa à terre, puis le prit par la main, le mena à une table, bourra ses poches de gâteaux.'

[11] Pour distinguer le 'DE' du chinois de celui du français, le premier est toujours mis en majuscule dans notre thèse.

[12] Les noms épithèses sont aussi inclus dans ce sous-type d'après la grammaire, mais aucune occurrence n'a été repérée dans cette chaîne.

[13] Les groupes prépositionnels sont aussi inclus dans ce sous-type d'après la grammaire, mais aucune occurrence n'a été repérée dans cette chaîne.

[28]曼希沃 马上 抖擞 一 下， 把 [拼命抓着 他 的 克利斯朵夫] 摔 在
Manxiwo mashang dousou yi xia ba pinmingzhuazhe ta de Kelisiduofu shuai zai
Melchior soudainement se-réveiller un coup BA cramponnant 3P DE Christophe rouler PREP
sub-rel-DE-nom

地砖 上。
dizhuan shang
carreau dessus
'Ce ne fut qu'un éclair. Melchior se secoua et envoya rouler contre le carreau Christophe, qui se cramponnait avec fureur.'

Certains auteurs, comme Liu (刘丹青, 2005), traitent tous les modifieurs, avec ou sans *de* 的 'DE', comme un seul type sans subdivision ; d'autres, comme Teng (1981), abordent la typologie des modifieurs par les approches comparatives entre plusieurs langues : les propositions subordonnées sont considérées comme un sous-type distinct des autres modifieurs en raison de leur complexité grammaticale. De ce point de vue, seules les traductions des subordonnées relatives d'autres langues vers le chinois sont considérées comme des subordonnées relatives chinoises, malgré l'absence de marqueurs grammaticaux explicites par les pronoms relatifs. Citons deux exemples de Teng (1980) de subordonnées relatives anglaises en position de modifieurs avec leur traduction vers des SN modifiés de relatives en chinois :

[29]I want [the watch [which you are holding in your hand]].
sub-rel

我 要 [[你 手里 拿着的] 那个 表]。
wo yao ni shouli nazhiede nage biao
I want you hand-in holding that watch
sub-rel

[30]I've heard [the story [which you're going to tell]].
sub-rel

[[你 现在 要 说的] 这个 故事]， 我 听过 。
ni xianzai yao shuode zhege gushi wo tingguo
you now want say this story I have-heard
sub-rel

(Teng, 1981 : 11)

Vu que les typologies des sous-types formels de modifieurs chinois sont loin d'être consensuelles entre les auteurs, nous chercherons à les classer et à les décrire selon d'autres critères. Comme le montrent les formules au début de cette section, la position de modifieur est unique pour le chinois, à savoir l'antéposition ; celle pour le français est double, à savoir l'antéposition et la postposition. Cette différence formelle nous semble intéressante pour deux raisons. Premièrement, la présence ou l'absence de modifieur postposé chaque langue est reliée au système linguistique. Les contraintes grammaticales de chaque langue peuvent influencer la position des modifieurs sans qu'il y ait de modification sémantique. Cette spécificité est reflétée par les formules des composants de chaque langue. Deuxièmement, la différence de modifieur antéposé peut susciter des interprétations spécifiques selon les langues. Pour le français, la position de modifieur peut être dépendante de sa longueur (*cf.* Adam, 2015 : 69) : la postposition permet des modifieurs plus longs. En d'autres termes, seuls les éléments courts peuvent être antéposés au nom, alors que tous les modifieurs chinois sont antéposés au nom, quelles que soient la catégorie et la longueur, comme le montrent les différents ordres *modifieur-nom* en chinois et *modifieur-nom-modifieur* en français dans les extraits [31] et [32] : tous les modifieurs chinois sont antéposés au nom, tandis que les modifieurs français sont partiellement antéposés et partiellement postposés.

[31][他] 老是 断断续续的 悲啼。仿佛 [这 一 堆 无意识的 尚未成形的 肉]，对 [他]
ta laoshi duanduanxuxude beiti fangfu zhe yi dui wuyishide shangwuchengxingde rou dui ta
3P toujours entrecoupé plaindre comme ce un CL inconscient informe viande à son
mod-nom

命中注定的 痛苦的 生涯 已经 有 了 预感。
mingzhongzhudingde tongkude shengya yijing you le yugan
réservé misérable vie déjà avoir LE pressentiment

[32][Il] continue toujours [sa] plainte entrecoupée. On dirait que [**cette misérable masse inconsciente**
mod-nom-mod

et informe] a le pressentiment de la vie de peines qui [lui] est réservée.

Adam (2015 : 69) remarque également que cette différence de position peut engendrer d'autres effets en français, car la position représente le degré d'inhérence des caractéristiques du référent. Autrement dit, les modifieurs antéposés au nom tendent à exprimer des caractéristiques plus importantes et intrinsèques au référent, alors que ceux postposés ont plus souvent trait à des caractéristiques accessoires et moins importantes. Ainsi, les caractéristiques exprimées par les modifieurs longs postposés en français sont *a priori* censés être plus lâches et moins fondamentales pour qualifier les référents. La postposition du nom permet une émergence de ses valeurs plus longue que celle de l'antéposition. Pour cette raison, la position peut permettre de savoir si les caractéristiques du référent exprimées par les modifieurs sont intrinsèques ou secondaires.

Pour le chinois, toutes ces remarques relatives aux modifieurs français ne sont plus pertinentes, étant donné l'absence de la postposition du modifieur. Nous en concluons que tous les modifieurs chinois permettent d'exprimer les caractéristiques intrinsèques et accessoires du référent sans distinction de position relative au nom, tandis que les caractéristiques accessoires peuvent être transformées en d'autres formes morpho-syntaxiques ou en d'autres structures phrastiques.

Le troisième composant des maillons modifiés est *nom* : il s'agit du noyau des maillons nominaux. Tout comme ceux du français, les maillons nominaux modifiés chinois peuvent être classés en *noms communs modifiés* et *noms propres modifiés*, avec les mêmes définitions que celles du français présentées dans la section *6.1.3. Modifieurs des maillons nominaux : longueur et restriction* du *Chapitre 6*. Pour rappel, les compositions des deux groupes de maillons modifiés peuvent être formulées comme suit :

Nom commun modifié = (déterminant) – modifieur(s) – nom commun

Nom propre modifié = (déterminant) – modifieur(s) – nom propre

Chaque groupe de maillons chinois est présent respectivement dans les *Annexes 34* et *35*, avec 47 noms communs modifiés et 28 noms propres modifiés.

7.1.2.2. Différences entre les maillons modifiés dans les deux langues

Il est plausible que le nombre de noms communs modifiés et celui de noms propres modifiés soient proches puisque la traduction de ce groupe d'expressions du français vers le chinois est théoriquement fidèle en termes de forme morpho-syntaxique. En effet, même si les modifieurs ne jouent pas un rôle grammatical indispensable dans chaque langue comme le font les déterminants, ils ne sont pas facultatifs puisqu'ils contiennent des informations sémantiques du référent qui ne peuvent pas être omises. Les maillons modifiés en chinois et en français sont inventoriés et mis en comparaison dans le tableau (7.11) : en chinois, 47 noms

communs modifiés et 28 noms propres modifiés, contre 42 noms communs modifiés et 34 noms propres modifiés en français.

Tableau (7.11) : Synthèse comparative des maillons modifiés en chinois et en français dans la chaîne du protagoniste

Langue	Chinois	Français
Noms communs modifiés	47	42
Noms propres modifiés	28	34

Selon les chiffres et les exemples obtenus, tous les maillons modifiés chinois et français ne sont pas parallèles à strictement parler : de légères différences quantitatives sont dues aux modifications de la structure syntaxique. Par exemple, le nom commun modifié 'une seule personne soi-même' *ziji-yigeren* 自己一个人 de l'extrait [33] correspond au pronom personnel français « il » de l'extrait parallèle [34] ; le nom propre modifié français [14] « Christophe, qui avait maintenant six ans » de l'extrait [35] est traduit par un nom propre sans modifieur 'Christophe' *Kelisiduofu* 克利斯朵夫 et une reprise pronominale immédiate '3P' *ta* 他 de l'extrait [36].

[33][他的] 做气 不 许 [他] 向 别人 求助。[他] 发誓 要 凭 [自己 一 个 人] 的
tade aoqi bu xu ta xiang bieren qiuzhu ta fashi yao ping ziji yi ge ren de
son orgueil NEG permettre 3P vers autrui recourir 3P jurer aller par soi-même un CL personne DE
nco-mod

力量 去 解决 困难。
liliang qu jiejue kunnan
force aller résoudre difficulté

[34]Son orgueil lui défendait de recourir à la charité des autres. **[Il]** se jura de se tirer d'affaire seul.
pro

[35]Melchior ne s'en occupait jamais. Louisa, forcée de sortir, les confiait à **[Christophe, qui avait maintenant six ans]**.
npr-mod

[36]曼希沃 从来不 照顾 他们。鲁意莎 要 出门，就 得 把 两 个 小的 交给
Manxiwo conglaibu zhaogu tamen Luyisha yao chumen jiu dei ba liang ge xiaode jiaogei
Melchior ne-jamais s'occuper 3Ps Louisa devoir sortir donc devoir BA deux CL petit donner

[克利斯朵夫]，**[他]** 现在 已经 有 六 岁 了。
Kelisiduofu ta xianzai yijing you you sui le
Christophe 3P maintenant déjà avoir six an LE
npr-non-mod pro

L'observation de ces deux groupes d'extraits parallèles révèle que les maillons modifiés peuvent être raccourcis, et *vice versa*, d'une langue vers l'autre et réciproquement. Tous les maillons modifiés chinois ne coïncident pas avec ceux du français, malgré leur quantité assez proche.

[14] Voir notre définition dans la section *6.1.3.2. Noms propres modifiés* du *Chapitre 6.*

7.1.3. Quatre formes de maillons non nominaux

Nous présentons dans cette partie les quatre groupes de maillons non nominaux chinois qui totalisent 83 % des composants de la chaîne : les *pronoms personnels*, les *déictiques*, les *déterminants possessifs* et les *zéros*. Sachant que ces quatre groupes présentent des quantités d'occurrences considérables, il semble difficile de les analyser quantitativement dans leur intégralité comme nous l'avons fait pour les groupes de maillons nominaux. Par conséquent, les analyses peuvent se fonder sur des échantillons d'occurrences aléatoires, afin d'obtenir quelques caractéristiques régulières de cette chaîne du protagoniste en chinois.

7.1.3.1. *Pronoms personnels de troisième personne* : composants principaux de la chaîne

Selon l'échelle d'accessibilité d'Ariel, plusieurs formes en anglais, comme les suivantes, renvoient aux référents du plus haut degré d'accessibilité. Ces formes se distinguent par ailleurs de leur degré d'accessibilité entre elles : les pronoms pleins (*full pronouns*) renvoient aux référents les moins accessibles parmi ces formes, alors que les accords zéro (*zero agreement*) expriment un degré d'accessibilité extrêmement haut parmi toutes les formes référentielles :

> *Full pronouns → Cliticized pronouns → Rich / First- and second-person agremment → Poor / third-person agreement → Zero agreement*
>
> *(Ariel, 1990 : 99)*

Malgré l'échelle pour l'anglais ci-dessus, toutes les variations formelles et les nuances d'accessibilité ne sont pas explicites en chinois : les pronoms chinois constituent un seul groupe ; les accords des personnes n'existent pas. Par conséquent, sous l'étiquette de *pronoms personnels*, nous parlons des pronoms personnels chinois de la troisième personne, y compris les formes accentuées et inaccentuées (*full pronouns* et *cliticized pronouns*) ; les pronoms personnels de première et deuxième personnes seront présentés plus tard sous l'étiquette de *déictique*.

Tout comme en français, les pronoms personnels chinois peuvent refléter le genre et le nombre de l'antécédent en emploi anaphorique. En raison du nombre au singulier et du genre masculin du protagoniste, seules deux formes sont obtenues dans ce groupe important. La forme pronominale la plus privilégiée est *ta* 他, correspondant à toutes les variations du pronom masculin au singulier en français, comme « il », « lui » et « le ». La seconde forme obtenue est *ziji* 自己, équivalant à toutes les formes variables du pronom réfléchi et réflexif en français, comme « se », « soi » et « lui-même ». Les deux pronoms chinois sont montrés dans l'extrait [37].

[37][克利斯朵夫]	骑着	马	跳过	危崖绝壁	。	(…)	玩	这些	游戏	的	时候，	[他]	常常
	Kelisiduofu	*qizhe*	*ma*	*tiaoguo*	*weiyajuebo*			*wan*	*zhexie*	*youxi*	*de*	*shihou*	*ta*	*changchang*
	Christophe	montant	cheval	sauter	précipice			jouer	ces	jeux	DE	temps	3P *pro-per*	souvent

会 懵懵懂懂的 出神。周围的 一切 都 隐灭 了，[他] 不 知道 [自己]在
hui mengmengdongdongde chushen zhouweide yiqie dou yinmie le ta bu zhidao zjiji zai
pouvoir étrange rêver entourant tout tous s'effacer LE 3P NEG savoir 3P PREP
pro-per *pro-per*

那里 做 些 什么，甚至 把[自己] 都 忘 了。
nali zuo xie shenme shenzhi ba zjiji dou wang le
là-bas faire quelque quoi même BA 3P même oublier LE
pro-per

'Christophe montait à cheval, sautait des précipices. (…) Au milieu de ces jeux, il avait des instants de rêvasserie étrange et de complet oubli. Tout ce qui l'entourait s'effaçait, il ne savait plus ce qu'il faisait, il ne se souvenait même plus de lui-même.'

Vu leur immense densité dans le texte, nous avons choisi aléatoirement les trois premiers paragraphes au début de l'épisode 37 de la section IV du roman (voir l'*Annexe 36 : Échantillon de pronoms personnels chinois dans la chaîne 'Christophe'* Kelisiduofu *克利斯朵夫 (section IV, épisode 37)*). Cet extrait contient au total 43 maillons, dont 22 sont pronoms personnels. Cette proportion (51 %) est assez proche de celle de ce groupe dans l'intégralité du corpus (44 %). Il est ainsi permis de considérer que les emplois réguliers des pronoms personnels dans cet extrait pourront refléter en quelque sorte ceux de tous les pronoms personnels dans le roman. Nous formulerons quelques remarques à partir de l'observation des maillons dans cet extrait[15].

Tableau (7.12) : Caractéristiques discursives des pronoms personnels chinois dans la chaîne du protagoniste (section IV, épisode 37)

Maillons	Fonction	Distance	Réf	Début
Ta 他 (« M-2-pro-1 »)	S	1	0	N
Ta 他 (« M-4-pro-2 »)	S	1	1	N
Ta 他 (« M-5-pro-3 »)	CDO	1	4	N
Ta 他 (« M-6-pro-4 »)	S	0	0	N
Ta 他 (« M-7-pro-5 »)	S	0	1	N
Ta 他 (« M-11-pro-6 »)	S	1	1	N
Ta 他 (« M-13-pro-7 »)	S	1	3	N
Ta 他 (« M-15-pro-8 »)	CC	0	2	N
Ta 他 (« M-16-pro-9 »)	S	1	3	N
Ta 他 (« M-19-pro-10 »)	S	1	2	N
Ta 他 (« M-22-pro-11 »)	S	1	1	N
Ta 他 (« M-28-pro-12 »)	S	1	4	N
Ta 他 (« M-29-pro-13 »)	S	0	2	N
Ta 他 (« M-31-pro-14 »)	S	0	0	N
Ta 他 (« M-33-pro-15 »)	S	0	1	N
Ta 他 (« M-34-pro-16 »)	S	1	2	N
Ta 他 (« M-35-pro-17 »)	S	1	8	N
Ta 他 (« M-37-pro-18 »)	S	1	2	N
Ta 他 (« M-39-pro-19 »)	CDO	0	2	N
Ta 他 (« M-40-pro-20 »)	S	1	0	N
Ta 他 (« M-41-pro-21 »)	CDO	0	6	N
Ta 他 (« M-42-pro-22 »)	S	1	1	N
Total / moyenne	82 % S	0,6	2,1	0

[15] Cet extrait ne contient pas de séquence dialogale, ce qui fait que nous ne discutons pas de l'usage des pronoms personnels chinois dans les différents types de séquences textuelles.

Au niveau formel, lorsque l'on voit les maillons pronominaux français, leur forme représente déjà leur fonction syntaxique, comme la différence entre le sujet en pronom inaccentué « il », le complément en pronom accentué « lui » et le déterminant possessif « son » ; mais les maillons chinois seuls ne permettent pas ce genre d'observation directe, car leur usage s'ancre profondément dans leur contexte phrastique et textuel. Dépourvu de la variation formelle liée à la fonction syntaxique, les pronoms personnels et les déterminants possessifs peuvent partager la même forme, comme *ta* 他 (« dét-poss-M-1 ») de « sa nature indisciplinée » qui est identique au pronom personnel *ta* 他 (« pro-per-M-2 ») :

[38]但	因	[他]	天性	倔强，	[Ø] 倒反	更喜欢	那些	和声。	[他]	最	高兴
dan	*yin*	*ta*	*tianxing*	*juejiang*	*daofan*	*gengxihuan*	*naxie*	*hesheng*	*ta*	*zui*	*gaoxing*
mais	car	son	nature	indiscipliné	contrairement	préférer	ces-là	accord	3P	le-plus	content
		dét-poss-M-1							*pro-per-M-2*		

在	人人佩服的	大	音乐家	的	作品	中	找出	这	一	类	例子，	拿去给	祖父
zai	*renrenpeifude*	*de*	*yinyuejia*	*de*	*zuopin*	*zhong*	*zhaochu*	*zhe*	*yi*	*lei*	*lizi*	*naqugei*	*zufu*
PREP	admirable	grand	musicien	DE	œuvre	dedans	trouver	ce	un	type	exemple	ramener	grand-père

或	老师	看。
huo	*laoshi*	*kan*
ou	maître	voir

Une très forte proportion de pronoms personnels occupent la place de sujet de phrase : 82 %, ce qui est plus élevé que toutes les autres formes. Cette proportion dans le roman-fleuve est également plus élevée que celle des textes narratifs journalistiques en chinois, qui atteint seulement un pourcentage de 61,82 % (Xu, 1990 : 325).

La distance entre un pronom personnel et son antécédent est toujours très proche, de 0 ou 1 phrase, avec seulement 0,6 phrases en moyenne.

Le nombre de référents concurrents est relativement moins important par rapport aux autres catégories de maillons, avec seulement 2,1 référents en moyenne.

La proportion des pronoms personnels chinois est relativement moins élevée que celle du français : 44 % contre 60 % dans l'intégralité de la chaîne du protagoniste. Cet écart peut être dû à diverses raisons morpho-syntaxiques et discursives. Étant donné les proportions de maillons nominaux et non nominaux chinois qui se rapprochent beaucoup de celles du français (voir le tableau (6.2) du *Chapitre 6* pour le français et le tableau (7.2) au début de ce chapitre pour le chinois), nous formulons l'hypothèse selon laquelle il existe pour les pronoms personnels chinois une possibilité de s'adapter aux autres catégories non nominales, comme notamment les zéros. Nous vérifierons cette hypothèse dans la partie suivante de manière comparative avec les autres groupes non nominaux.

7.1.3.2. *Pronoms déictiques* : emploi dans les séquences dialogales et narratives

Sous l'étiquette de *déictiques*, nous parlons des pronoms personnels de première et deuxième personnes, pour les distinguer de ceux de troisième personne. Ce groupe de pronoms sont aussi appelés *pronoms de dialogue* par Charolles (2002 : 195), étant donné leur usage fondamental dans les dialogues ; dans notre corpus, il pourrait en effet s'agir principalement de séquences dialogales. C'est pour cette raison que nous focalisons notre

observation sur le contexte de leur occurrence : si les séquences dialogales sont considérées comme un contexte typique des déictiques, existe-t-il des usages exceptionnels dans les séquences narratives ou autres types de contexte éventuels ?

La répartition des déictiques ainsi que le type de leur contexte sont présentés dans l'*Annexe 37 : Pronoms déictiques chinois dans la chaîne 'Christophe'* Kelisiduofu *克利斯朵夫*. Ce groupe concerne 4 % de la chaîne, avec 161 maillons ; cette proportion est identique à celle du français, en raison de l'usage équivalent de ce groupe de maillons dans les deux langues. Les séquences dialogales ne concernent que 10 % du corpus en chinois (voir l'*Annexe 32*), où 93 % de déictiques sont repérés ; seuls 12 occurrences (7 %) apparaissent dans les séquences narratives. Lorsque les pronoms déictiques sont utilisés dans les séquences narratives, il peut s'agir de différents *plans d'énonciation*, dans le cas où le narrateur prend directement la parole et intervient dans la narration qui porte sur les personnages du roman :

> *On retrouve également, bien que de façon beaucoup plus marginale, des changements de plan d'énonciation. Il s'agit principalement de parenthèses « méta-narratives » (comme le souligne Schnedecker, 1997 : 131), c'est-à-dire de commentaires du narrateur sur son propre récit. Ce sont en général des énoncés pris en charge par l'énonciateur-narrateur, qui viennent s'insérer à l'intérieur du récit afin d'introduire une qualification supplémentaire du personnage, une attribution de propriété, un jugement appréciatif, un ajustement* a posteriori*, etc. (Baumer, 2015 : 89)*

Un des exemples est le suivant :

[39]慈悲的 温柔的 笑容！ 把 [他的] 心 都 融化 了！ [他] 觉得 多 舒畅，多 爱 它！
cibeide wenroude xiaorong ba tade xin dou ronghua le ta juede duo shuchang duo ai ta
indulgent aimant sourire BA son cœur même fondre LE 3P sentir que agréable que aimer 3P

啊，再来 一 次 罢！ 再 对 **[我]** 笑 一 下 罢！ 你 别 走 呀！
a zailia yi ci ba zai dui wo xiao yi xia ba ni bie zou ya
ah encore un fois INTERJ encore à 1P sourire un fois INTERJ 2P NEG partir INTERJ
déictique

'Le beau sourire indulgent et aimant ! il fond le cœur de tendresse ! qu'il fait du bien, qu'on l'aime ! Encore ! Souris-moi encore ! Ne t'en va point !…'

Contrairement aux pronoms déictiques utilisés dans les séquences dialogales qui renvoient aux participants, le référent de ceux des séquences narratives est moins évident, ce qui fait qu'ils peuvent susciter, dans certaines situations, des ambiguïtés éventuelles. Prenons l'extrait [40] comme exemple : les deux déictiques '2P' *ni* 你 peuvent être interprétés comme coréférentiels à 'enfant' *haizi* 孩子 du début de l'extrait ; ils sont donc maillons de la chaîne du protagoniste. Pourtant, il semble aussi plausible que ces deux déictiques désignent les gens en général sans référer à aucune personne spécifique. La troisième interprétation possible est que le narrateur cherche à créer un quasi-dialogue entre le personnage et les lecteurs du roman : les pronoms déictiques peuvent donc désigner des lecteurs quelconques. Selon ces deux dernières interprétations, ils ne sont plus composants de la chaîne du protagoniste mais potentiellement maillons de la chaîne de « gens en général » ou de la chaîne de « lecteurs en général ». Lorsque l'on regarde l'extrait parallèle [41] en français, l'interprétation référentielle de « vous » est aussi équivoque que celle de '2P' *ni* 你 en chinois : en français, deux antécédents d'humains se trouvent dans le contexte, permettant également trois interprétations référentielles de celles du chinois, à savoir « l'enfant », renvoyant au protagoniste, et « on », renvoyant aux gens en général ou aux lecteurs en général.

[40][孩子] 一 个 又 一 个 的 听上 老半天， 听 它们 低下去， 没有 了； 它们
haizi *yi* *ge* *you* *yi* *ge* *de* *tingshang* *laobantian* *ting* *tamen* *dixiaqu* *meiyou* *le* *tamen*
enfant un CL encore un CL DE écoutant longuement écouter 3Ps diminuer s'éteindre LE 3Ps
protagoniste

有如 田野 里 的 钟声， 飘飘荡荡， 随着 风 吹过来 又 吹远去； 细
youru *tianye* *li* *de* *zhongsheng* *piaopiaodangdang* *suizhe* *feng* *chuiguolai* *you* *chuiyuanqu* *xi*
comme champs dedans DE cloche balançant avec vent ramener et s'éloigner soigneusement

听 之下， 远远的 还 有 别的 不同的 声音 交错 回旋， 仿佛 羽虫
ting *zhixia* *yuanyuande* *hai* *you* *biede* *butongde* *shengyin* *jiaocuo* *huixuan* *fangfu* *yuchong*
écouter dessous lointain encore avoir autre différent voix se-mêler se-tourner comme insecte

飞舞； 它们 好象 在 那儿 叫 **[你]**， 引 **[你]** 到 远的 地方……
feiwu *tamen* *haoxiang* *zai* *naer* *jiao* *ni* *yin* *ni* *dao* *yuande* *difang*
voler 3Ps comme PREP là-bas appeler 2P *déictique* attirer 2P *déictique* à loin endroit

[41]**[L'enfant]** les écoute longuement, un à un, diminuer et s'éteindre ; ils se balancent comme les cloches, lorsqu'**[on]** est dans les champs, et que le vent les apporte et les éloigne tour à tour ; puis, quand **[on]** prête l'oreille, **[on]** entend dans le lointain d'autres voix différentes qui se mêlent et tournent, ainsi que des vols d'insectes ; elles ont l'air de **[vous]** appeler, de **[vous]** attirer loin… loin…

Dans tous les cas, l'interprétation référentielle des pronoms personnels déictiques peut avoir un emploi prototypique équivalent dans les dialogues des deux langues. En dehors de ce contexte, l'interprétation référentielle de ce groupe devient plus vague et multiple : ils peuvent être interprétés soit par un personnage spécifique, soit par les gens en général, soit par un groupe d'individus. Ces interprétations se rapprochent beaucoup de celles du pronom « on » en français.

7.1.3.3. Déterminants possessifs

Les déterminants possessifs occupent la troisième place en chinois (8 %, 305 occurrences) et la deuxième place en français (16 %, 720 occurrences). Ce groupe de maillons permet de renvoyer au protagoniste, en établissant en même temps une relation avec d'autres référents, secondaires. Différents de ceux du français, les déterminants possessifs chinois ne reflètent pas le genre et le nombre du référent renvoyé par le SN possessif. Rappelons que chaque déterminant possessif chinois peut avoir une ou deux formes selon le nom qui se trouve à sa suite (voir le tableau (5.4) du *Chapitre 5*) pour exprimer la relation de parenté et la relation inaliénable (*cf*. Rabut *et al*., 2012). Par exemple, deux formes de déterminants possessifs sont acceptables pour dire « mon père » (relation de parenté) : 'mon (père)' *wo(baba)* 我 (爸爸) et 'mon (père)' *wode(baba)* 我的 (爸爸) ; pour exprimer la relation de possession temporaire d'objet, la seule forme acceptable est *wode* 我的, comme 'mon (sac)' *wode(bao)* 我的 (包) et 'mon (train)' *wode(huoche)* 我的 (火车). Selon le seul critère morpho-syntaxique, il n'est pas possible de catégoriser *wo* 我 comme pronom personnel ou comme déterminant possessif[16]. Ces deux types de déterminants possessifs sont marqués dans l'extrait [42] :

[16] C'est pour cette raison que nous ne pouvons pas distinguer les déterminants possessifs des pronoms personnels dans l'*Annexe 29*.

[42]祖父 临终的 景象 老 是 在 **[他的]** 记忆 中， [他]还 听到 那 可怕的
zufu *linzhongde* *jingxiang* *lao* *shi* *zai* *tade* *jiyi* *zhong* *ta* *hai* *tingdao* *na* *kepade*
grand-père mourant scène toujours être PREP son souvenir dedans 3P encore entendre ce-là terrible
dét-poss-1

呼吸。 整个的 天地 都 改变 了， 仿佛 布满着 一 片 冰雾。 在 **[他]**
huxi *zhenggede* *tiandi* *dou* *gaibian* *le* *fangfu* *bumanzhe* *yi* *pian* *bingwu* *zai* *ta*
respiration tout nature tout changer LE semblant étendu un CL glace-brume PREP son
dét-poss-2

周围， 不论 转 向 哪 一 边， 总 觉得 那 盲目的 野兽 有 股 血腥 气
zhouwei *bulun* *zhuan* *xiang* *na* *yi* *bian* *zong* *juede* *na* *mangmude* *yeshou* *you* *gu* *xuexing* *qi*
alentour peu-importe tourner vers quel un côté toujours sentir ce-là aveugle bête avoir CL meurtrier air

吹 在 **[他]** 脸 上。
chui *zai* *ta* *lian* *shang*
souffler PREP son visage dessus
dét-poss-3

'Les souvenirs de l'agonie le poursuivaient ; il entendait l'horrible respiration. La nature entière avait changé ; il semblait que se fût étendue sur elle une brume de glace. Autour de lui, partout, de quelque côté qu'il se tournât, il sentait sur sa face le souffle meurtrier de la Bête aveugle.'

Ce groupe est beaucoup moins utilisé comme maillons : seulement 8 % en chinois, contre 16 % en français.

Les analyses quantitatives se fondent sur un échantillon aléatoire des 19 premières occurrences consécutives situées au début de la section III (épisodes 25 – 28), comme le montre l'*Annexe 39 : Échantillon de déterminants possessifs chinois dans la chaîne 'Christophe' Kelisiduofu 克利斯朵夫 (section III, épisodes 25 – 28)* et le tableau (7.13)[17] ci-dessous.

Tableau (7.13) : Caractéristiques discursives des déterminants possessifs chinois dans la chaîne du protagoniste (section III, épisodes 25 – 28)

Maillons	Distance	Réf	Début	Discours
Ta 他 (« dét-poss-1 »)	0	1	N	N
Tade 他的 (« dét-poss-2 »)	2	3	N	N
Tade 他的 (« dét-poss-3 »)	1	1	N	N
Tade 他的 (« dét-poss-4 »)	1	3	N	N
Ta 他 (« dét-poss-5 »)	1	0	N	N
Ta 他 (« dét-poss-6 »)	0	2	N	N
Ta 他 (« dét-poss-7 »)	1	1	PA	N
Ta 他 (« dét-poss-8 »)	1	3	N	N
Ta 他 (« dét-poss-9 »)	0	4	N	N
Tade 他的 (« dét-poss-10 »)	0	2	N	N
Tade 他的 (« dét-poss-11 »)	0	0	PA	N
Tade 他的 (« dét-poss-12 »)	0	1	N	N
Ta 他 (« dét-poss-13 »)	0	1	N	N
Tade 他的 (« dét-poss-14 »)	0	1	N	N
Ta 他 (« dét-poss-15 »)	0	0	N	N
Ta 他 (« dét-poss-16 »)	0	2	N	N
Tade 他的 (« dét-poss-17 »)	1	1	N	N
Nide 你的 'ton / ta / tes' (« dét-poss-18 »)	0	2	N	DD
Tade 他的 (« dét-poss-19 »)	0	4	N	DD

[17] *Ta* 他 et *tade* 他的 sont les deux déterminants possessifs de la troisième personne au singulier ; la traduction des autres formes est mise à côté de chaque maillon.

Total / moyenne	0,4	1,7	2 PA (11 %)	2 DD (11 %)

Malgré une partie de formes identiques à celle des pronoms personnels, les caractéristiques référentielles des déterminants possessifs sont malgré tout différentes de ces derniers. Les paramètres observés dans l'échantillon de ces maillons représentent quelques tendances de leurs emplois.

La distance entre un déterminant possessif et son antécédent est très faible : 0,4 phrase en moyenne ; les deux se situent très souvent dans la même phrase. Ainsi, les déterminants possessifs représentent un plus haut degré de continuité référentielle que celui des pronoms personnels. Ils ne sont jamais utilisés après une distance relativement longue pour redémarrer la chaîne du protagoniste. En raison de cette petite distance le nombre de référents est également peu significatif, avec une moyenne de 1,7 : les déterminant possessifs sont utilisés relativement plus souvent dans le contexte mono-référentiel, où d'autres référents n'interviennent pas. Malgré la faible distance, les déterminants possessifs peuvent être utilisés au début des séquences textuelles : 2 maillons (11 %) apparaissent dans la première phrase de paragraphe. Le texte est découpé en paragraphes, entre lesquels la chaîne établit une continuité étroite. Ce groupe de maillons est aussi constaté dans les discours rapportés directs : 2 maillons (11 %) correspondent à l'interlocuteur en deuxième personne et à un autre personnage en troisième personne.

7.1.3.4. *Zéro* : typologie et différences sino-françaises

Dans l'échelle d'accessibilité d'Ariel, les expressions dites de haut degré et d'extrême haut degré d'accessibilité référentielle ne concernent pas seulement des formes explicites mais également des zéros. Selon les langues, il peut s'agir, d'une part, des pronoms atones dans celles où ces pronoms sont obligatoirement exprimés, et, d'autre part, des zéros dans celles où les pronoms atones ne sont pas obligatoirement exprimés. Le français appartient au premier groupe et le chinois au second. Huang (1994 : 21) propose de distinguer les langues par leur *température* (*hot-cool division*), à savoir *langue chaude*, *langue moyenne* et *langue froide*, en fonction de l'usage des zéros dans la relation anaphorique de chaque langue. Selon cette typologie, le chinois est classé comme *langue très froide* :

> *Chinese is characterized as a very "cool" language in that it allows the massive use of zero anaphora and the interpretation of anaphoric expressions in Chinese relies heavily on inference, context and world knowledge. (Huang, 1994 : 21)*

Dépourvus de conjugaison, les verbes prédicatifs chinois se présentent toujours à la forme indicative sans variation morpho-syntaxique. Cela pouvant susciter davantage de risques d'ambiguïté référentielle par rapport aux formes explicites, l'emploi des sujets zéro rencontre ainsi des limites. Par exemple, il existe peu de zéros dans les genres officiels, comme les lois et les documents gouvernementaux (*cf*. Hou & Sun, 侯敏 & 孙建军, 2004). Au contraire, les textes littéraires semblent mieux tolérer ce type d'ambiguïté éventuelle : le fait que des lecteurs comprennent un texte de différentes manières n'a aucune conséquence grave, et cela peut même rajouter de l'intérêt aux textes littéraires.

Comme nous l'avons vu avec la présentation des résultats de Li & Thompson (1979) sur les zéros chinois dans la section *3.2.2.2. Chaînes pronominales et expressions anaphoriques zéro en chinois* du *Chapitre 3*, les auteurs donnent une proportion de zéros beaucoup plus élevée que celle que nous avons nous-même obtenue ; cela peut s'expliquer de deux manières. Premièrement, le texte chinois, dans notre corpus, est traduit à partir d'un texte français : les pronoms explicites français sont traduits par la forme parallèle vers le chinois. Deuxièmement, les textes dans le corpus de Li & Thompson (1979) ont été écrits aux 16ème et 17ème siècle, alors que notre corpus date du début du 20ème siècle : les pronoms explicites chinois aient réalisé une évolution durant ce grand laps de temps.

Lorsque les zéros engagent la fonction de sujet de phrase, ils peuvent constituer une relation anaphorique et coréférentielle avec les éléments en position de plusieurs fonctions syntaxiques de la phrase précédente (*cf*. Fang, 方梅, 1985, 2018), comme le sujet ('Paul' *Baoluo* 保罗 de la phrase [43]), le complément de sujet ('Paul' *Baoluo* 保罗 de la phrase [44]) et le complément d'objet ('une tasse de café' *bei-kafei* 杯咖啡 de la phrase [45]) dans nos exemples :

[43][保罗] 喝 了 杯 咖啡， 然后 [**Ø**] 走 了。
Baoluo he le bei kafei ranhou zou le
Paul boire LE tasse café ensuite partir LE
sujet

'Paul a pris un café et (il) est parti.'

[44][保罗] 的 咖啡 很烫， [**Ø**] 喝 不 了。
baoluo de kafei hentang he bu liao
Paul DE café trop-chaud boire NEG LE
complément-sujet

'Le café de Paul est trop chaud et (il) ne peut pas le boire.'

[45]保罗 喝 了 [杯 咖啡]， [**Ø**] 有点 烫。
baoluo he le bei kafei youdian tang
Paul boire LE tasse café un-peu trop-chaud
CDO

'Paul a pris un café, (le café) est un peu trop chaud.'

D'après l'observation de Fang (方梅, 1985, 2018) à partir de son corpus, plusieurs types de relations anaphoriques peuvent être réalisées par les zéros, mais environ 80 % des cas concernent la reprise de sujet du type de l'exemple [43], dit « sujet zéro coréférentiel au sujet de la proposition principale » (1985 : 44, 2018 : 15). Pour cette raison, le sujet zéro est aussi un zéro susceptible d'être présent en grand nombre, ce qui nous intéresse grandement par rapport aux zéros aux autres positions syntaxiques. En raison de très nombreux zéros dans notre corpus, l'observation est réduite à un échantillon aléatoire, celui des premières occurrences au début de la section III (épisode 25), comme le montrent l'*Annexe 40 : Échantillon de zéros chinois dans la chaîne 'Christophe'* Kelisiduofu *克利斯朵夫 (section III, épisode 25)* et le tableau (7.14) ci-dessous.

Tableau (7.14) : Caractéristiques discursives des sujets zéro chinois dans la chaîne du protagoniste (section III, épisode 25)

Maillons	Type de coréférence	Catégorie d'antécédent / subséquent	Distance avec l'antécédent
« zéro-1 »	Cataphorique au sujet	Pronom personnel	1
« zéro-2 »	Cataphorique au sujet	Pronom personnel	0
« zéro-3 »	Cataphorique au sujet	Pronom personnel	0
« zéro-4 »	Anaphorique au sujet	Pronom personnel	0

« zéro-5 »	Anaphorique au sujet	Pronom personnel	0
« zéro-6 »	Anaphorique au sujet	Nom propre	0
« zéro-7 »	Anaphorique au sujet	Pronom personnel	0
« zéro-8 »	Anaphorique au sujet	Pronom personnel	0
« zéro-9 »	Anaphorique au déterminant de sujet	Déterminant possessif	0
« zéro-10 »	Anaphorique au déterminant de sujet	Déterminant possessif	0
« zéro-11 »	Anaphorique au sujet	SN sans déterminant	0
« zéro-12 »	Anaphorique au sujet	Nom propre	1
« zéro-13 »	Anaphorique au sujet	Pronom personnel	0
« zéro-14 »	Anaphorique au sujet	Pronom personnel	0
« zéro-15 »	Anaphorique au sujet	SN sans déterminant	0
« zéro-16 »	Anaphorique au sujet	SN sans déterminant	0
« zéro-17 »	Anaphorique au sujet	Nom propre	0
« zéro-18 »	Anaphorique au sujet	Nom propre	0
« zéro-19 »	Anaphorique au sujet	Nom propre	0
« zéro-20 »	Anaphorique au déterminant de sujet	Déterminant possessif	0
« zéro-21 »	Anaphorique au sujet	Pronom personnel	0
« zéro-22 »	Anaphorique au sujet	Pronom personnel	0
« zéro-23 »	Anaphorique au sujet	Pronom personnel	0
« zéro-24 »	Anaphorique au sujet	Pronom personnel	0
« zéro-25 »	Anaphorique au sujet	Pronom personnel	0
« zéro-26 »	Anaphorique au sujet	Pronom personnel	0
« zéro-27 »	Anaphorique au sujet	Nom propre	0
« zéro-28 »	Anaphorique au sujet	Pronom personnel	0
« zéro-29 »	Anaphorique au sujet	Pronom personnel	0
Total	23 anaphoriques au sujet (79 %)	17 pronoms personnels (63 %)	27 intra-phrastiques (93 %)

L'observation des 29 sujets zéro dans l'échantillon de la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫 nous a permis de constater un emploi privilégié et deux autres relativement rares des zéros en chinois[18] : nous les présentons selon l'ordre décroissant de leur proportion.

i) L'usage privilégié est le zéro anaphorique au sujet de la proposition précédente, comme le montrent les zéros des extraits [46] et [47] :

[46]好几 天 以前，他 先 把 情节 详细 解释 了 一番。[克利斯朵夫] 一点也不
haoji tian yiqian ta xian ba qingjie xiangxi jieshi le yifan Kelisiduofu yidianyebu
plusieurs jour avant 3P d'abord BA sujet détail expliquer LE un-coup Christophe rien
sujet-antécédent

明白， **[Ø]** 只 记得 有 些 可怕的 事； [他] 一边 迫不及待的 想 看，
mingbai zhi jide you xie kepade shi ta yibian pobujidaide xiang kan
comprendre seulement retenir avoir des terrible chose 3P et brûlant désirer voir
zéro-1-anaphorique

一边 **[Ø]** 也 十分 怕 看。
yibian ye shifen pa kan
et aussi assez avoir-peur voir
zéro-2-anaphorique

'Plusieurs jours à l'avance, il lui avait raconté longuement le sujet de la pièce. Christophe n'y avait rien compris ; mais il avait retenu qu'il se passerait des choses terribles ; et, tout en brûlant du désir de les voir, il en avait grand peur.'

[47]逢 到 惊心动起的 情节，他 装做 若无其事的 轻轻咳嗽，遮掩 心中的 激动；
feng dao jingxindongqide qingjie ta zhuangzuo ruowuqishide qingqingkesou zheyan xinzhongde jidong
passer à dramatique intrigue 3P prétendre indifférent toussoter cacher intérieur troublé

[18] Outre ces trois types repérés dans l'échantillon des zéros, d'autres possibilités existent encore (*cf.* Fang, 方梅, 1985, 2018), mais aucune d'entre elles n'a été trouvée dans notre corpus.

可是 [克利斯朵夫] 看得 很 清楚， [Ø] 觉得 很 高兴。 [他] 热 极 了，
keshi *Kelisiduofu* *kande* *hen* *qingchu* *juede* *hen* *gaoxing* *ta* *re* *ji* *le*
mais Christophe voyant très clair sentir très content 3P avoir-chaud horrible LE
sujet-antécédent *zéro-3-anaphorique*

[Ø] 昏昏欲睡， [Ø] 坐 在 那儿 又 非常 不舒服。
hunhunyushui *zuo* *zai* *naer* *you* *feichang* *bushufu*
endormi s'asseoir PREP là-bas aussi très mal
zéro-4-anaphorique *zéro-5-anaphorique*

'Aux passages dramatiques, il toussotait d'un air indifférent, pour cacher son trouble ; mais Christophe le voyait ; et cela lui faisait plaisir. Il avait horriblement chaud, il tombait de sommeil, et il avait très mal où il était assis.'

Cet usage des zéros est majoritaire, concernant 22 cas sur 29 ; les 7 zéros restants sont utilisés de deux manières.

ii) Les zéros des extraits [48] et [49] montrent leur deuxième usage : ceux-ci ne coréfèrent pas au sujet de phrase, mais au déterminant possessif du sujet : « zéro-1 » et « zéro-2 » coréfèrent à 'son' *tade* 他的 de 'son cœur' *tade-xin* 他的心 ; « zéro-3 » et « zéro-4 » coréfèrent à 'son' *ta* 他 de 'sa gorge' *ta-houlong* 他喉咙.

[48]终于 老人 出现 了， 他们俩 动身 了。 [他的] 心 在 胸 中
zhongyu *laoren* *chuxian* *le* *tamenlia* *dongshen* *le* *tade* *xin* *zai* *xiong* *zhong*
enfin vieux-homme paraître LE 3Ps-deux partir LE son cœur PREP poitrine dedans
dét-antécédent

乱跳， [Ø] 口干舌燥， [Ø] 连 一 个 字 都 说不上来。
luantiao *kouganshezao* *lian* *yi* *ge* *zi* *dou* *shuobushanglai*
sauter langue-sèche même un CL mot tout articuler-NEG
zéro-1-anaphorique *zéro-2-anaphorique*

'Enfin le vieux parut et ils partirent ensemble. Le cœur lui sautait dans la poitrine. Il avait la langue sèche, il ne pouvait articuler une syllabe.'

[49]但 那 时 [孩子] 觉得 要 闭过气去 了； [他] 喉咙 象 着了凉 一样的
dan *na* *shi* *haizi* *juede* *yao* *biguoqiqu* *le* *ta* *houlong* *xiang* *zhaoleliang* *yiyangde*
mais ce-là temps enfant sentir aller étouffer LE son gorge comme prendre-froid-LE CMP
dét-antécédent

难过， [Ø] 双手 掐着 脖子， [Ø] 连 口水 都 咽不下 了。
nanguo *shuangshou* *qiazhe* *bozi* *lian* *koushui* *dou* *yanbuxia* *le*
douloureux deux-main serrant cou même salive tout avaler-NEG LE
zéro-3-anaphorique *zéro-4-anaphorique*

'L'enfant crut qu'il allait étouffer ; il avait la gorge douloureuse, comme quand il avait pris froid ; il se serrait le cou avec ses mains, il ne pouvait plus avaler sa salive.'

iii) Le dernier usage, encore moins fréquent, a trait aux cas de cataphore : le zéro est coréférentiel au sujet explicite de la proposition suivante. Seuls 3 cas sont concernés, comme « zéro-1 » et « zéro-2 » et « zéro-3 » de l'extrait [50].

[50][Ø] 又 要 用心， [Ø] 又 是 厌烦， [Ø] 大 颗 大 颗
you *yao* *yongxin* *you* *shi* *yanfan* *da* *ke* *da* *ke*
aussi devoir faire-attention aussi être ennuyé gros CL gros CL
zéro-1-cataphorique *zéro-2-cataphorique* *zéro-3-cataphorique*

的 眼泪 沿着 鼻子 眼 腮帮 淌着： [他] 把 常常 冻得 红肿的 小
de *yanlei* *yanzhe* *bizi* *yan* *saibang* *tangzhe* *ta* *ba* *changchang* *dongde* *hongzhongde* *xiao*
DE larme suivant nez yeux joue coulant 3P BA souvent refroidissant rouge-gourde petit
sujet-subséquent

手 在 黑 白 的 键子 上 搬动。
shou *zai* *hei* *bai* *de* *jianzi* *shang* *bandong*
main PREP noir blanc DE touche dessus remuer

'Crispé d'attention et d'ennui, de grosses larmes coulant le long de ses joues et de son nez, il remuait sur les touches blanches et noires ses petites mains rouges, souvent gourdes de froid.'

Ces diverses possibilités de zéros permettent de constituer plus facilement une solide continuité référentielle phrastique et inter-phrastique en chinois par rapport au français.

Dans notre corpus, 27 % des maillons de la chaîne du protagoniste sont des zéros, qui se situent ainsi à la troisième place de toutes les formes de maillons après les pronoms personnels et les déterminants possessifs. Par contraste, la chaîne française ne comporte que 4 % de zéros. Comparons parallèlement les extraits [46] – [50]. Ce sont les pronoms personnels explicites français qui occupent les places parallèles des zéro chinois. Ce type de différence peut expliquer les grands écarts quantitatifs entre les nombres de zéros ainsi que les nombres de pronoms explicites dans les deux langues.

7.1.4. Bilan

Dans cette partie, nous avons présenté la chaîne du protagoniste en chinois selon la forme de ses composants : pour les maillons nominaux, il s'agit de leur tête lexicale (nom propre et nom commun), de leur déterminant (démonstratif, possessif, indéfini et sans déterminant) et de leur modifieur ; pour les maillons non nominaux, il s'agit de trois catégories explicites (pronom personnel, déictique et déterminant possessif) et une forme implicite (zéro). L'emploi des noms propres et des déictiques sont presque identiques en français et en chinois : ces expressions se localisent dans les endroits parallèles dans les deux langues. Les différences sont observées sur deux groupes de maillons. Le premier concerne les SN sans déterminant chinois, qui ne réfèrent pas du tout de la même manière que ceux du français. En fonction de leur contexte, les SN sans déterminant chinois réfèrent de la même manière que celle des SN de divers déterminants en français. Le second groupe concerne les zéros, qui sont énormément récurrents et occupent souvent les mêmes places de pronoms explicites dans les textes parallèles en français. Nous poursuivons notre observation dans les parties suivantes sur les autres chaînes de référence en chinois dans le roman-fleuve *Jean-Christophe*.

7.2. Chaînes des personnages secondaires

Après la présentation de la chaîne du protagoniste, nous poursuivons dans cette partie par l'observation des chaînes des personnages secondaires qui constituent trois autres chaînes globales dans le corpus en chinois, avec pour objectif de vérifier si ce groupe présente les mêmes spécificités que celles de la chaîne du protagoniste en chinois et celles des chaînes des mêmes personnages en français.

Selon le tableau (7.2) au début du chapitre, trois personnages sont représentés par une chaîne globale. Il s'agit des trois personnages secondaires : 'Louisa' *Luyisha* 鲁意莎, 'Melchior' *Manxiwo* 曼希沃 et 'Jean-Michel' *Yuehan-Mixier* 约翰·米希尔. Ces trois chaînes sont distribuées à environ une moitié d'épisodes des quatre sections du corpus. Leur répartition est presque identique à celle du français dans le corpus, comme le synthétise le tableau (7.15). Pourtant, les trois chaînes chinoises présentent des longueurs plus courtes et des densités moins élevées que celles du français : leur longueur moyenne est de 701,7 maillons, contre 738,7 maillons pour le français ; leur densité est de 0,0078, contre 0,0119.

Tableau (7.15) : Répartition, longueur et densité des chaînes des personnages secondaires en chinois et en français

Critère	Répartition		Longueur		Densité	
Langue	Ch	Fr	Ch	Fr	Ch	Fr
'Louisa' *Luyisha* 鲁意莎	58 %	56 %	472	511	0,0053	0,0083
'Melchior' *Manxiwo* 曼希沃	64 %	64 %	774	788	0,0086	0,0127
'Jean-Michel' *Yuehan-Mixier* 约翰·米希尔	54 %	54 %	859	917	0,0096	0,0148
Moyenne	59 %	58 %	701,7	738,7	0,0078	0,0119

7.2.1. Démarrage et redémarrage

Manifestement, en raison de la relation de traduction du français vers le chinois, les premiers maillons et ceux de redémarrage des trois personnages secondaires en chinois se localisent dans des endroits parallèles à ceux du français. Les extraits [51] et [52] (parallèles aux extraits [77] et [78] du *Chapitre 6*) comprennent le premier maillon de ces chaînes de personnages secondaires en chinois : 'vieux-homme' *lao-ren* 老人, celui de la chaîne 'Jean-Michel' *Yuehan-Mixier* 约翰·米希尔 ; 'mère' *muqin* 母亲, celui de la chaîne 'Louisa' *Luyisha* 鲁意莎 et 'ton mari' *ni-zhangfu* 你丈夫, celui de la chaîne 'Melchior' *Manxiwo* 曼希沃.

[51]初生的 婴儿 在 摇篮 里 扭动。**[老人]** 进来 虽然 把 木靴 脱 在 门
chushengde yinger zai yaolan li niudong laoren jinlai suiran ba muxue tuo zai men
nouveau-né bébé PREP berceau dedans s'agiter vieux-homme entrer malgré BA sabot laisser PREP porte
premier-maillon-JM

外， 走路 的 时候 地板 还是 格格的响， 孩子 哼啊嗐啊 的 哭 了。**[母亲]**
wai zoulu de shihou diban haishi gegedexiang haizi hengahaia de ku le muqin
extérieur marcher DE temps plancher encore craquer enfant ONOMA DE pleurer LE mère
premier-maillon-L

从 床 上 探出 身子 抚慰 他； 祖父 摸索着 点起灯来， 免得 孩子 在
cong chuang shang tanchu shenzi fuwei ta zufu mosuoshe dianqidenglai miande haizi zai
par lit dessus sortir corps assurer 3P grand-père tâtonnant allumer-lampe évitant enfant PREP
黑夜 里 害怕。
heiye li haipa
nuit durant avoir-peur

[52]他 停 了 一会， 想着 要不要 把 这 意思 再 申说 一番； 但 他 再也
ta ting le yihui xiangzhe yaobuyao ba zhe yisi zai shenshuo yifan dan ta zaiye
3P se-taire LE un-instant méditant si BA ce pensée encore parler une-fois mais 3P ne-plus

找不到 话， 于是 静默 了 半晌， 又 很 生气的 问："怎么 **[你 丈夫]**
zhaobudao hua yushi jingmo le banshang you hen shengqide wen zenme ni zhangfu
trouver-NEG mot donc se-taire LE un-moment aussi très irrité demander comment ton mari
premier-maillon-M

还 不 回来？"
hai bu huilai
encore NEG rentrer

Les deux SN sans déterminant chinois 'vieux-homme' *lao-ren* 老人 et 'mère' *muqin* 母亲 sont à la place des SN définis français « le vieux » et « la mère » ; le SN possessif chinois 'ton mari' *ni-zhangfu* 你丈夫 correspond à la même forme du français. Selon ces trois chaînes, le SN sans déterminant et le SN possessif semblent être les catégories privilégiées pour introduire les nouveaux référents d'importance secondaire dans le roman-fleuve en chinois. Ces trois personnages ne sont pas dénommés lors de leur introduction, mais ils sont évoqués par leur identité familiale et leur relation avec le protagoniste.

À la différence de la répartition omniprésente de la chaîne du protagoniste, ces trois chaînes présentent des ruptures dans des épisodes. Lorsqu'elles sont discontinues et que leur référent a besoin d'être réactivé, quelles sont les catégories chinoises privilégiées après une longue distance ? Le tableau (7.16) synthétise le nombre de ruptures et les catégories variables des maillons au moment du redémarrage des chaînes. La chaîne 'Louisa' *Luyisha* 鲁意莎 en présente 10 ; la chaîne 'Melchior' *Manxiwo* 曼希沃 et la chaîne 'Jean-Michel' *Yuehan-Mixier* 约翰·米希尔 11. Toutes les formes ne redémarrent pas les chaînes : aucun SN démonstratif ou SN indéfini n'y est repéré. Les noms propres et les SN sans déterminant sont les deux catégories dominantes qui favorisent potentiellement la réactivation des personnages secondaires après la rupture de chaînes.

Tableau (7.16) : Catégories des maillons chinois au démarrage et au redémarrage des chaînes des personnages secondaires

Nom de chaîne	'Louisa' *Luyisha* 鲁意莎	'Melchior' *Manxiwo* 曼希沃	'Jean-Michel' *Yuehan-Mixier* 约翰·米希尔	Total
Noms propres	3	7	2	12
SN sans déterminant	7	2	9	18
SN possessifs	0	1	0	1
Non nominaux	0	1	0	1
Total	10	11	11	32

7.2.2. Catégories grammaticales des composants

Nous nous intéressons dans cette partie aux caractéristiques de la composition des chaînes secondaires en chinois par rapport à celles de la chaîne du protagoniste en chinois et à celles des chaînes secondaires en français. L'ensemble des maillons de ces trois chaînes est montré dans les *Annexes 41 – 43*[19] ; le nombre et la proportion de leurs composants sont synthétisés dans le tableau (7.17). Les points communs aux trois chaînes se manifestent déjà selon le critère formel. Trois catégories sont constituants principaux : les noms propres, les SN sans déterminant et les maillons non nominaux ; les autres catégories sont minoritaires ou absentes. Nous analyserons les raisons et les conditions discursives qui suscitent ces points communs et les différences éventuelles selon la catégorie.

[19] L'*Annexe 41 : Maillons chinois de la chaîne 'Louisa'* Luyisha *鲁意莎*, l'*Annexe 42 : Maillons chinois de la chaîne 'Melchior'* Manxiwo *曼希沃* et l'*Annexe 43 : Maillons chinois de la chaîne 'Jean-Michel'* Yunehan-Mixier *约翰·米希尔*.

Tableau (7.17) : Composants des chaînes des personnages secondaires en chinois et en français

Nom de chaîne	'Louisa' *Luyisha* 鲁意莎	'Melchior' *Manxiwo* 曼希沃	'Jean-Michel' *Yuehan-Mixier* 约翰·米希尔	Moyenne	
				Ch	Fr
Noms propres	14 % (64)	18 % (142)	6 % (49)	12 % (255)	12 % (270)
SN sans déterminant	21 % (100)	17 % (130)	26 % (221)	21 % (451)	6 % (137)
SN possessifs	1 % (7)	0,5 % (4)	0,3 % (3)	0,7 % (14)	6 % (133)
SN démonstratifs	0	0,4 % (3)	0,1 % (1)	0,2 % (4)	0,1 % (2)
SN indéfinis	0,6 % (3)	0,4 % (3)	0,1 % (1)	0,3 % (7)	0,5 % (11)
Pronoms personnels	29 % (140)	28 % (216)	31 % (256)	29 % (204)	42 % (319)
Déictiques	5 % (24)	4 % (27)	4 % (33)	4 % (28)	5 % (103)
Zéros	24 % (111)	26 % (201)	28 % (242)	26 % (151)	6 % (132)
Déterminants possessifs	5 % (23)	6 % (48)	6 % (53)	6 % (43)	13 % (104)
Total	472	774	859	702	739

7.2.2.1. Groupes de maillons nominaux dominants : noms propres et SN sans déterminant

Les noms propres et les SN sans déterminant sont regroupés puisqu'ils partagent trois points communs remarquables. Premièrement, ils occupent les deux premières places parmi tous les maillons nominaux dans les trois chaînes : la chaîne 'Louisa' *Luyisha* 鲁意莎 présente 14 % de noms propres et 21 % de SN sans déterminant ; la chaîne 'Melchior' *Manxiwo* 曼希沃 respectivement 18 % et 17 % ; la chaîne 'Jean-Michel' *Yuehan-Mixier* 约翰·米希尔 6 % et 26 %. Tous les autres maillons nominaux sont beaucoup moins nombreux par rapport à ces deux groupes. Deuxièmement, les noms propres et les SN sans déterminant (en correspondance avec les SN définis) en chinois sont susceptibles d'exprimer les référents les moins accessibles. Troisièmement, sachant que les noms propres et les noms communs en chinois ne se différencient pas de leur lettre initiale, les noms propres peuvent être considérés comme un des sous-types de SN sans déterminant particulier. Tout comme la graphie chinoise du protagoniste, celle du nom propre 鲁意莎 (transcription en pinyin : /lu-yi-sha/) est une traduction phonétique du prénom français « Louisa ». Ce choix vaut dans tout le roman-fleuve, comme 曼希沃 /man-xi-wo/ pour « Melchior » et 约翰·米希尔 /yue-han · mi-xi-er/ pour « Jean-Michel ».

Excepté ces trois points communs, les noms propres et les SN sans déterminant présentent des différences lorsqu'il s'agit de leur emploi comme maillons dans les chaînes secondaires. En raison de leur grand nombre d'occurrences, nous avons choisi de fonder les analyses sur un échantillon d'occurrences mixtes des trois chaînes au début de la section II du roman-fleuve : l'*Annexe 44 : Échantillon de noms propres chinois dans les chaînes des personnages secondaires (section II, épisodes 11 – 12)* présente les noms propres ; l'*Annexe 45 : Échantillon de SN sans déterminant chinois dans les chaînes des personnages secondaires (section II, épisodes 11 – 13)* les SN sans déterminant[20].

En ce qui concerne les conditions contextuelles qui favorisent l'emploi des noms propres chinois, il semble que cette catégorie soit plutôt utilisée comme sujet de phrase (82 %)

[20] L'extrait ne contenant pas de séquence dialogale, il est impossible d'observer de manière comparative les emplois des noms propres dans les différents types de séquences textuelles.

après une distance relativement longue (5,0 phrases en moyenne) dans le contexte multi-référentiel (avec 10,4 référents concurrents en moyenne entre le maillon et son antécédent). Par exemple, l'extrait [53] suivant contient 3 noms propres des 2 chaînes : tous les trois sont sujets de phrase ; 9 autres référents apparaissent dans le même contexte :

[53]那 时 [他] 已经 不 是 独子 了。[**曼希沃**] 给 妻子 [每 年] 生 [一 个 孩子],
na shi ta yijing bu shi duzi le Manxiwo gei qizi mei nian sheng yi ge haizi
ce-là moment 3P déjà NEG être seul-enfant LE Melchior à femme chaque an naître un CL enfant
réf-1 *npr-17* *réf-2* *réf-3*

完全 不 管 [[将来] 的 结局]。[两 个] 在 [很 小 的 时候] 就 死了。
wanquan bu guan jianglai de jieju liang ge zai hen xiao de shihou jiu sile
complètement NEG s'inquiéter futur DE fin deux CL PREP très jeune DE temps donc mort
réf-4 *réf-5* *réf-6* *réf-7*

[其余 两 个] 正好 是 三 岁 和 四 岁。[**曼希沃**] 从来不 照顾 [他们]。[**鲁意莎**]
qiyu liang ge zhenghao shi san sui he si sui Manxiwo conglaibu zhaogu tamen Luyisha
autre deux CL juste être trois an et quatre an Melchior ne-jamais s'occuper 3Ps Louisa
réf-8 *npr-18* *réf-8* *npr-19*

要 出门, 就 得 把 [两 个 小的] 交 给 [克利斯朵夫], [他] [现在] 已经 有
yao chumen jiu dei ba liang ge xiaode jiao gei Kelisiduofu ta xianzai yijing you
devoir sortir donc devoir BA deux CL petit donner à Christophe 3P maintenant déjà avoir
réf-8 *réf-1* *réf-1* *réf-9*

六 岁 了。
liu sui le
six an LE

'Il n'était plus seul enfant. Melchior faisait un enfant à sa femme chaque année, sans s'inquiéter de ce qui en arriverait plus tard. Deux étaient morts en bas âge. Deux autres avaient trois et quatre ans. Melchior ne s'en occupait jamais. Louisa, forcée de sortir, les confiait à Christophe, qui avait maintenant six ans.'

Par contraste, les conditions favorables à l'usage des SN sans déterminant ne sont pas toutes similaires. La fonction syntaxique privilégiée n'est plus le sujet : les SN sans déterminant sont utilisés à toutes les positions de phrase, dans des proportions assez équilibrées. Par exemple, dans l'extrait [54], 'fils' *erzi* 儿子 et 'exécutant' *yanzoujia* 演奏家 sont compléments d'objet de phrase :

[54]老 约翰·米希尔 对 [**儿子**] 的 声名 很 得意, 看到 [**演奏家**] 的 卖弄 技巧
lao Yuehan-Mixier dui erzi de shengming hen deyi kandao yanzoujia de mainong jiqiao
vieux Jean-Michel à fils DE succès très satisfaisant voir exécutant DE jouer virtuosité
SN-sans-dét *SN-sans-dét*

简直 出神 了; [**老人 自己**] 就 从来不 能 好好的 弄 一 种 乐器。
jianzhi chushen le laoren ziji jiu conglaibu neng haohaode nong yi zhong yueqi
presque extase LE vieux-homme soi juste ne-jamais pouvoir proprement jouer un sorte instrument
SN-sans-dét

'Michel savourait les succès de son fils ; il était en extase devant les tours de force du virtuose, lui qui n'avait jamais su jouer proprement d'aucun instrument.'

Excepté cette différence, les autres critères présentent des chiffres très proches de ceux des noms propres : les SN sans déterminant peuvent se présenter dans le contexte multi-référentiel (10,9 référents en moyenne), après une distance relativement longue (5,0 phrases en moyenne), et au début des séquences textuelles (15 % d'occurrences en début de paragraphe et d'épisode). Les paramètres des noms propres et des SN sans déterminant sont synthétisés dans le tableau (7.18) ci-dessous.

Tableau (7.18) : Comparaison des noms propres et des SN sans déterminant en chinois dans les chaînes des personnages secondaires (section II[21])

Total / moyenne	Fonction	Distance	Réf	Début
Noms propres	82 % S	5,0 (1 LG, 2 TL)	10,4 (3 BP)	14 % (3 PA)
SN sans déterminant	35 % S	5,0 (1 LG, 2 TL)	10,9 (3 BP)	15 % (2 PA, 1 EP)

7.2.2.2. Maillons nominaux de faible quantité : SN possessifs, SN démonstratifs et SN indéfinis

De manière globale, trois groupes concernent 99 % de composants des chaînes des personnages secondaires : les noms propres, les SN sans déterminant et les maillons non nominaux. En d'autres termes, tous les autres maillons nominaux concernent moins de 1 % de chaque chaîne ; nous parlons ainsi de trois groupes de quantité peu significative : les SN possessifs, les SN démonstratifs et les SN indéfinis.

Par rapport aux SN possessifs français, qui occupent 7 % des chaînes des personnages secondaires en moyenne, les SN possessifs chinois sont très peu utilisés : moins de 1 % de maillons dans chaque chaîne, malgré la relation familiale étroite entre eux et avec le protagoniste. Les SN possessifs totalisent 14 occurrences dans les trois chaînes (voir l'*Annexe 46 : SN possessifs chinois dans les chaînes des personnages secondaires*) : 7 dans la chaîne 'Louisa' *Luyisha* 鲁意莎, 4 dans celle de 'Melchior' *Manxiwo* 曼希沃 et 3 dans celle de 'Jean-Michel' *Yuehan-Mixier* 约翰·米希尔. Dans le cadre phrastique, les SN possessifs occupent rarement la place de sujet : seulement 2 occurrences sur 14. Dans le cadre textuel, ce groupe de maillons peuvent apparaître après une longue distance (6 occurrences sur 14). 2 SN possessifs sont repérés en début de paragraphe ; 4 sont utilisés dans les discours rapportés directs. Un exemple de SN possessif chinois est montré dans l'extrait suivant :

[55]哦！ 克利斯朵夫 多 恨 **[他的 父亲]**， 恨 [他] 的 不 想到 他们， [Ø] 连
o *Kelisiduofu* *duo* *hen* *tade* *fuqin* *hen* *ta* *de* *bu* *xiangdao* *tamen* *lian*
ah Christophe comment détester son *SN-poss* père détester 3P DE NEG penser 3Ps même

吃掉 了 他们的 份儿 都 没想到！
chidiao *le* *tamende* *fener* *dou* *meixiangdao*
manger LE leur part même se-douter

'Ah ! comme Christophe détestait **[son père]**, comme il lui en voulait de ne pas penser à eux, de ne même pas se douter qu'il leur mangeait leur part !'

Les SN démonstratifs constituent le deuxième groupe de maillons de faible quantité, avec, au total, 4 occurrences, dont 3 dans la chaîne 'Melchior' *Manxiwo* 曼希沃 et 1 dans celle 'Jean-Michel' *Yuehan-Mixier* 约翰·米希尔 (voir l'*Annexe 47 : SN démonstratifs chinois dans les chaînes des personnages secondaires*). Le nombre de SN démonstratifs est assez similaire à celui des chaînes secondaires en français, qui n'en totalise que 2 dans les trois chaînes. Malgré ce faible nombre d'occurrences, la plupart des SN démonstratifs sont utilisés dans les discours rapportés directs, en mode de référence exophorique, comme dans l'extrait [56] :

[56]约翰·米希尔 沉着 嗓子， 气冲冲的 接着 说：“我 犯 了 什么 天条， 生下
Yuehan-Mixier *chenzhe* *sangzi* *qichongchongde* *jiezhe* *shuo* *wo* *fan* *le* *shenme* *tiantiao* *shengxia*
Jean-Michel baissant voix éclatant continuant dire 1P faire LE quel faute naître

[21] Épisodes 11 – 12 pour les noms propres ; épisodes 11 – 13 pour les SN sans déterminant.

[这 个 酒鬼 的 儿子]？ 我 这 一 辈子 省吃俭用的， **真** 是 够受了！
zhe ge jiugui de erzi *wo zhe yi beizi shengchijianyongde zhen shi shougoule*
ce CL ivrogne DE fils 1P ce un vie privé vraiment être suffisant
SN-dém-M-1

'Jean-Michel continua d'une voix plus sourde, avec des éclats de colère :
– Qu'ai-je fait au bon Dieu pour avoir **[cet ivrogne de fils]** ? C'est bien la peine d'avoir vécu comme j'ai vécu, de m'être privé de tout !'

Enfin, 7 SN indéfinis sont repérés dans les trois chaînes : 3 dans la chaîne 'Louisa' *Luyisha* 鲁意莎, 3 dans celle de 'Melchior' *Manxiwo* 曼希沃 et 1 dans celle de 'Jean-Michel' *Yuehan-Mixier* 约翰·米希尔 (voir l'*Annexe 48 : SN indéfinis chinois dans les chaînes des personnages secondaires*). Cette quantité est très proche de celle du français, qui en totalise 11 dans les trois chaînes : c'est une catégorie rare dans les chaînes des personnages secondaires dans les deux langues. Dans le cadre phrastique, aucun SN indéfini n'est sujet de phrase : ils occupent d'autres fonctions, comme celle de complément d'objet ('une CL cuisinière' *yi-ge-chuniang* 一个厨娘 de l'extrait [57]).

[57]无人 管束 的 船 会 向 暗礁 直 撞 过去， 而 足智多谋的 **曼希沃**
wuren guanshi de chuan hui xiang anjiao zhi zhuang guoqu er zuzhiduomoude Manxiwo
sans-personne contrôler DE bateau aller vers écueil tout-droit cogner vers et intrigant Melchior

居然 娶 了 **[一 个 厨娘]**。
juran qu le yi ge chuniang
surprenant épouser LE un CL cuisinière
SN-ind-L

'Le bateau laissé libre va droit contre l'écueil ; et l'intrigant Melchior épousa **[une cuisinière]**.'

7.2.2.3. Pronoms personnels, déterminants possessifs et zéros

Les maillons non nominaux sont les composants principaux des trois chaînes, pour environ deux tiers de chaque chaîne. Selon le résultat de notre recensement des catégories non nominales synthétisé dans le tableau (7.19), on obtient quatre groupes de maillons : les pronoms personnels, les déictiques, les déterminants possessifs et les zéros. Les déictiques présentent des proportions très proches dans les deux langues (4 % en chinois et 5 % en français), alors que les trois autres groupes manifestent des différences : 29 % de pronoms personnels chinois contre 42 % en français ; 6 % de déterminants possessifs chinois contre 13 % en français ; 26 % de zéros chinois contre 6 % en français. Ces trois formes sont susceptibles d'avoir des rôles différents dans les chaînes de référence des deux langues.

En raison de leur grand nombre d'occurrences, nous avons choisi d'effectuer les analyses sur un échantillon d'occurrences prélevé dans l'épisode 2[22] de la section I du livre (voir l'*Annexe 49 : Échantillon de maillons non nominaux chinois dans les chaînes des personnages secondaires (section I, épisode 2)*). L'extrait en question contient 39 pronoms personnels, 13 déterminants possessifs et 19 zéros. La distance entre l'occurrence et son antécédent est très faible pour les trois formes : la moyenne ne dépasse pas 1 phrase. Celle des zéros est seulement de 0,1 phrase, beaucoup plus courte donc que celle des formes explicites, respectivement à 0,49 et 0,46 phrase. Malgré cette petite distance, les pronoms

[22] L'extrait ne contenant pas de séquence dialogale, ce sont les paramètres des maillons spécifiques à la séquence narrative.

personnels peuvent coïncider avec le début des séquences textuelles, contrairement aux deux autres formes. Ces trois formes non nominales tendent à apparaître dans les contextes mono-référentiels où la moyenne du nombre de référents est très faible. Tout comme la distance, celle des zéros n'atteint que 0,8 référent, et est beaucoup moins élevée que celle des pronoms personnels (1,5 référents) et des déterminants possessifs (1,4 référents).

Tableau (7.19) : Caractéristiques discursives des maillons non nominaux chinois dans les chaînes des personnages secondaires (section I, épisode 2)

Catégorie	Nbre	Fonction	Distance	Réf	Début
Pronom personnel	39	56 % S	0,49	1,5	5 % (1 EP, 1 PA)
Déterminant possessif	13	0 S	0,46	1,4	0
Zéro	19	100 %	0,1	0,8	0

7.2.3. Bilan

À la différence de la chaîne du protagoniste qui couvre tout le corpus, les ruptures qui interviennent dans les chaînes des personnages secondaires fournissent une bonne matière pour étudier le redémarrage des chaînes de référence en chinois. Ce sont les SN sans déterminant et les noms propres qui se manifestent principalement à la place de redémarrage dans les chaînes des humains.

La composition des chaînes des personnages secondaires présente des spécificités par rapport à celle du protagoniste en chinois. Les chiffres du tableau (7.20) montrent les points communs et les différences : les SN possessifs, les SN démonstratifs et les SN indéfinis sont rarement utilisées dans les deux groupes de chaînes en chinois ; les déictiques, les zéros et les déterminants possessifs ne manifestent pas non plus de différence proportionnelle significative entre les chaînes secondaires et celle du protagoniste. Leurs distinctions sont pourtant remarquables dans trois catégories : les noms propres, les SN sans déterminant et les pronoms personnels. Les noms propres concernent 13 % des chaînes secondaires, proportion beaucoup plus élevée que celle de la chaîne du protagoniste (7 %) ; de même, les SN sans déterminant sont présents dans 21 % des chaînes secondaires, mais seulement à hauteur de 11 % dans la chaîne du protagoniste ; les pronoms personnels (29 %) sont, par contre, beaucoup moins utilisés dans les chaînes secondaires que dans celle du protagoniste (44 %).

Tableau (7.20) : Comparaison entre les chaînes des personnages secondaires et celle du protagoniste en chinois

Catégorie	Moyenne des chaînes secondaires	Protagoniste
Nom propre	13 %	7 %
SN sans déterminant	21 %	11 %
SN possessif	0,6 %	1 %
SN démonstratif	0,3 %	0,5 %
SN indéfini	0,4 %	0,4 %
Pronoms personnel	29 %	44 %
Déictique	4 %	4 %
Zéro	26 %	27 %
Déterminant possessif	6 %	8 %
Total	702	3817

7.3. Chaînes d'autres humains : épisodiques, éphémères et indéfinis

Les chaînes d'autres humains qui apparaissent dans le roman-fleuve sont en principe les mêmes que celles du français : les humains épisodiques, les humains éphémères et les humains indéfinis. Nous aborderons dans un premier temps les points communs des deux premiers groupes d'humains, et, dans un second temps, les particularités des successions des pronoms personnels indéfinis en chinois.

7.3.1. Chaînes des référents humains épisodiques et éphémères parallèles

La répartition, la longueur ainsi que les maillons des chaînes des personnages épisodiques sont listés dans l'*Annexe 50 : Maillons chinois dans les chaînes des personnages épisodiques* et l'*Annexe 51 : Caractéristiques des chaînes des personnages épisodiques en chinois*. Ce groupe concerne 37 chaînes, dont la répartition s'étale entre 1 – 12 épisodes sur 50 ; leur longueur se situe entre 10 – 275 maillons. Diverses catégories occupent la place du premier maillon : 7 noms propres, 2 SN possessifs, 9 SN indéfinis, 9 SN sans déterminant et 10 non nominaux. Les maillons non nominaux sont composants dominants pour la plupart des chaînes épisodiques, se rapportant à 29 chaînes sur 37, avec une moyenne de 61 % de maillons dans chaque chaîne épisodique.

Les maillons des chaînes des personnages éphémères sont listés dans l'*Annexe 52 : Maillons chinois dans les chaînes des personnages éphémères* et l'*Annexe 53 : Caractéristiques des chaînes des personnages éphémères en chinois*. Ces personnages sont appelés ainsi puisque leur chaîne n'est repérée que dans 1 – 2 épisodes, avec un nombre de maillons inférieur à 10. Diverses catégories occupent la place du premier maillon : 13 formes non nominales, 24 SN sans déterminant, 5 SN démonstratifs, 13 SN indéfinis, 7 noms propres, 5 SN possessifs. Ce groupe concerne 67 chaînes, dont 46 sont constituées par des maillons non nominaux comme composants privilégiés. La moyenne des maillons non nominaux est de 69 % par chaîne.

7.3.2. Succession des expressions personnelles équivoques en chinois

Il est difficile de dire s'il existe un ou plusieurs équivalents du pronom français « on » en chinois étant donné les nombreuses interprétations référentielles de ce pronom personnel indéfini français, en fonction desquelles les équivalents chinois s'avèrent, eux aussi, multiples. À notre connaissance, aucun auteur n'a étudié les chaînes constituées des pronoms polyvalents et équivoques en chinois. C'est pour cette raison que nous nous cherchons à établir les équivalents chinois éventuels du pronom français « on ».

Les pronoms personnels indéfinis chinois sont nombreux : qu'est-ce qui nous permet d'établir des équivalences entre ces pronoms dans les deux langues ? Pour répondre à cette question, nous avons tout d'abord, passé en revue les extraits français caractérisés par une succession de « on », pour voir si les pronoms personnels indéfinis, dans les extraits parallèles en chinois, se succèdent eux aussi. Dans ce cas-là, des zéros peuvent apparaître en substitution des pronoms personnels indéfinis explicites. Ensuite, il est aussi possible que les suites de pronoms personnels indéfinis n'apparaissent pas parallèlement dans les mêmes endroits en français. Nous avons donc annoté les pronoms indéfinis qui se succèdent dans chaque épisode en tant qu'expressions coréférentielles, pour examiner enfin si celles-ci constituent une vraie chaîne de référence. Après avoir effectué ces deux étapes, nous avons obtenu une liste des équivalents chinois plausibles du pronom « on » en français, comme le montre le tableau (7.21), totalisant 152 maillons, légèrement moins nombreux que ceux du français.

Tableau (7.21) : Succession des expressions personnelles équivoques en chinois

	Suite des expressions équivoques chinoises
1	*Ø – Ø – Ø – Ø – Ø – zijide* 自己的
2	*Dajia* 大家 – *renjia* 人家 – *Ø – Ø – women* 我们 – *womende* 我们的
4	*Ni* 你 – *ni* 你 – *ni* 你
6	*Ø – Ø – Ø – Ø – ren* 人
9	*Renjia* 人家 – *Ø – yigeren* 一个人 – *ren* 人
11	*Ren* 人 – *ren* 人 – *dajia* 大家 – *Ø – renjia* 人家 – *dajia* 大家 – *dajia* 大家 – *dazhong* 大众
12	*Renjia* 人家 – *yigeren* 一个人 – *bieren* 别人
13	*Renjia* 人家 – *bieren* 别人 – *quanshijie-de-ren* 全世界的人
14	*Renjia* 人家 – *renjia* 人家 – *renjia* 人家
18	*Yigeren* 一个人 – *ren* 人 – *Ø – Ø – ren* 人 – *Ø – ren* 人 – *ren* 人 – *ren* 人
20	*Ren* 人 – *ren* 人 – *ren* 人 – *Ø – ren* 人家 – *ren* 人 – *ren* 人 – *renjia* 人家 – *dajia* 大家 – *ni* 你 – *ni* 你 – *Ø – Ø – yigeren* 一个人 – *ren* 人
21	*Renjia* 人家 – *ni* 你 – *ni* 你 – *renjia* 人家 – *ni* 你 – *ni* 你 – *ni* 你 – *ren* 人 – *ren* 人
26	*Ren* 人 – *renjia* 人家 – *renjia* 人家 – *Ø – ren* 人 – *Ø – Ø*
27	*Dajia* 大家 – *renjia* 人家 – *renjia* 人家 – *Ø – renjia* 人家
31	*Ren* 人 – *ren* 人家 – *ren* 人 – *Ø – dajia* 大家 – *Ø – renjia* 人家 – *ren* 人
32	*Ren* 人 – *renjia* 人家 – *renjia* 人家 – *renjia* 人家 – *renjia* 人家 – *Ø*
33	*Renjia* 人家 – *shui* 谁 – *Ø – Ø – women* 我们
35	*Ren* 人 – *ren* 人 – *renjia* 人家 – *renjia* 人家 – *ren* 人 – *ren* 人 – *renjia* 人家
37	*Renjia* 人家 – *bieren* 别人 – *renjia* 人家 – *ren* 人 – *renjia* 人家 – *ren* 人 – *Ø – Ø – Ø – renjia* 人家 – *Ø – Ø – ren* 人
39	*Renjia* 人家 – *renjia* 人们 – *renjia* 人家 – *renjia* 人家

40	*Renjia* 人家 – *Ø* – *renjia* 人家 – *renjia* 人家 – *ren* 人
41	*Ren* 人 – *renjia* 人家 – *renjia* 人家
42	*Renjia* 人家 – *ren* 人 – *renjia* 人家 – *ren* 人 – *ren* 人
43	*Renjia* 人家 – *ren* 人 – *yigeren* 一个人 – *Ø* – *ziji* 自己 – *Ø* – *Ø* – *ni* 你 – *Ø* – *ziji* 自己 – *Ø* – *ni* 你 – *nide* 你的 – *yigeren* 一个人
48	*Ren* 人 – *ni* 你 – *Ø* – *ni* 你 – *renjia* 人家 – *ren* 人
Total	152 maillons

Les expressions obtenues sont variables :

- des zéros ;
- des pronoms génériques ('tout le monde' *dajia* 大家, 'gens' *ren* 人, 'gens' *dazhong* 大众 et 'tout le monde' *quanshijie-de-ren* 全世界的人) ;
- des pronoms déictiques ('nous' *women* 我们 et 'tu' *ni* 你) ;
- des pronoms indéfinis ('des gens / quelqu'un / autre(s) personne(s)' *renjia* 人家, 'une personne' *yi-ge-ren* 一个人, 'autre personne' *bieren* 别人 et 'n'importe qui' *shui* 谁 ;
- des pronoms réciproques ('soi-même' *ziji* 自己) ;
- des déterminants possessifs qui s'accordent avec des pronoms susmentionnés ('son / sa / ses' *zijide* 自己的, 'notre / nos' *womende* 我们的 et 'ton / ta / tes' *nide* 你的).

Toutes ces expressions susmentionnées peuvent être équivalentes du pronom français « on », mais l'emploi de chaque forme chinoise dépend plus étroitement de son contexte.

7.4. Chaînes des référents non humains

Nous avons repéré au total 236 chaînes répartissant dans un seul ou plusieurs épisodes du corpus. Leur longueur est variable entre 3 et 35 maillons (la plus longue chaîne étant celle de 'Rhin'). Ces référents appartiennent aux sept catégories sémantiques, comme le montre le tableau (7.22), parallèles à celles du français. Leurs maillons et leur répartition sont présentés dans l'*Annexe 54 : Chaînes des référents non humains en chinois*. Les chaînes de certains référents sont longues, avec un nombre de maillons supérieur à 10 ; celles de certains autres sont récurrentes, à savoir que leurs maillons sont répartis dans plusieurs épisodes. Nous examinons tout d'abord les composants divers dans les chaînes parallèles du fleuve « Rhin » et de « la musique », et ensuite les chaînes des deux groupes de référents distinctes entre deux langues, à savoir celles des notions temporelles et celles des parties du corps.

Tableau (7.22) : Chaînes des référents non humains en chinois

Catégorie de référent	Nbre de CR	Longueur	Répartition
« musique » et relatifs	36	3 – 49 maillons	1 – 8 épisodes
« fleuve » et relatifs	7	3 – 40 maillons	1 – 2 épisodes
Objets concrets	67	3 – 34 maillons	1 – 4 épisodes
Lieux	40	3 – 30 maillons	1 – 8 épisodes
Temps	8	3 – 10 maillons	1 – 2 épisodes
Corps et relatifs	34	3 – 27 maillons	1 – 7 épisodes
Référents abstraits et discursifs	46	3 – 19 maillons	1 – 3 épisodes
Total	236	-	-

7.4.1. Composants des chaînes parallèles du « fleuve » et de « la musique »

Tout comme dans les textes en français, dans les textes en chinois, les deux référents non humains cruciaux du roman-fleuve sont « le Rhin » et « la musique ».

La chaîne 'Rhin' *Laiyinhe* 莱茵河 est constituée de 40 maillons, répartis dans 2 épisodes. Elle est plus courte et moins distribuée que celle du français, composée de 50 maillons répartis dans 5 épisodes. Pourtant, cette chaîne chinoise présente des composants plus variables que ceux du français : 14 pronoms personnels (y compris probablement des déterminants démonstratifs), 11 SN sans déterminant et 11 zéros sont composants majoritaires de la chaîne, alors que les composants français sont beaucoup moins riches, avec 21 pronoms personnels et 25 SN définis pour la plupart de la composition. Une séquence de maillons chinois et français est montrée dans les extraits parallèles suivants :

[58][莱茵河] 在 屋 下 奔流。人 在 楼梯 的 窗口 临 **[河]** 眺望， 好似
Laiyinhe *zai* *wu* *xia* *benliu* *ren* *zai* *louti* *de* *chuangkou* *lin* *he* *tiaowang* *haosi*
Rhin PREP maison bas couler on PREP escalier DE fenêtre vers fleuve voir CMP
Rhin-1-npr *Rhin-2-SN-sans-dét*

悬 在 动荡的 天空。克利斯朵夫 平常 一拐一拐 下楼 的 时候 总是 对
xuan *zai* *dongdangde* *tiankong* *Kelisiduofu* *pingchang* *yiguaiyiguai* *xialou* *de* *shihou* *zongshi* *dui*
suspendu PREP mouvant ciel Christophe habituel clopinantl descendre DE temps toujours à

[河] 瞧上 一 眼 的， 但 从来没 见到 今天 这样的 景色。悲伤 使
he *qiaoshang* *yi* *yan* *de* *dan* *conglaimei* *jiandao* *jintian* *zheyangde* *jingse* *beishang* *shi*
fleuve voir un œil DE mais ne-jamais voir aujourd'hui ce paysage chagrin rendre
Rhin-3-sans-dét

感觉 格外 锐敏； 眼睛 经过 泪水 的 洗涤， 往事 的 遗迹 给 一扫而空， 一切
ganjue *gewai* *ruimin* *yanjing* *jingguo* *leishui* *de* *xidi* *wangshi* *de* *yiji* *gei* *yisaoerkong* *yiqie*
sens particulièrement aigu yeux après pleur DE laver souvenir DE trace AUX tout-lavé tout

在 眼膜 上 刻划得 更 清楚 了。在 孩子 心目 中， **[河]** 仿佛 是 个
zai *yanmo* *shang* *kehuade* *geng* *qingchu* *le* *zai* *haizi* *xinmu* *zhong* *he* *fangfu* *shi* *ge*
PREP yeux dessus tracé plus clair LE PREP enfant cœur dedans fleuve CMP être CL
Rhin-4-sans-dét

有生命的 东西， **[Ø]** 是 个 不可思议的 生物， 但 **[Ø]** 比 他 所见到的 一切
youshengmingde *dongxi* *shi* *ge* *bukesiyide* *shengwu* *dan* *bi* *ta* *suojiandaode* *yiqie*
vivant chose être CL inexplicable être mais CMP 3P vu tout
Rhin-5-zéro *Rhin-6-zéro*

都 强得 多！
dou *qiangde* *duo*
tout puissant plus

[59][Le Rhin] coulait en bas, au pied de la maison. De la fenêtre de l'escalier, on était suspendu
Rhin-1-npr

au-dessus [**du fleuve**] comme dans un ciel mouvant. Christophe ne manquait jamais de [**le**]
Rhin-2-SN-déf *Rhin-3-pro*
regarder quand il descendait les marches en clopinant ; mais jamais il ne [**l'**]avait vu encore, comme
Rhin-4-pro
aujourd'hui. Le chagrin aiguise les sens ; il semble que tout se grave mieux dans les regards, après que les pleurs ont lavé les traces fanées des souvenirs. [**Le fleuve**] apparut à l'enfant comme un
Rhin-5-SN-déf
être, – inexplicable, mais combien plus puissant que tous ceux qu'il connaissait !

La chaîne du référent générique 'musique' *yinyue* 音乐, l'autre chaîne de référents non humains très importante, comporte 49 maillons répartis dans 8 épisodes : cette longueur et cette répartition sont même plus importantes que certaines chaînes d'humains épisodiques et éphémères. Dans les deux langues les chaînes présentent une longueur et une répartition très proches, mais leurs composants sont toujours très différents : 10 SN sans déterminant constituent la partie principale de la chaîne chinoise, ainsi que 5 zéros et un seul pronom personnel ; en français, ce sont 23 SN définis, 9 SN sans déterminant et 11 pronoms personnels qui font partie de manière privilégiée de cette chaîne.

Tableau (7.23) : Chaînes du « fleuve » et de la « musique » en chinois et en français

Nom de chaîne	Langue	Ch	Fr
« Fleuve »	Répartition	2 épisodes	5 épisodes
	Longueur	40 maillons	50 maillons
	Composants	1 nom propre 14 pronoms personnels **11 SN sans déterminant** **11 zéros** **1 pronom démonstratif**	4 noms propres 21 pronoms personnels **25 SN définis**
« Musique »	Répartition	8 épisodes	7 épisodes
	Longueur	49 maillons	50 maillons
	Composants	1 SN indéfini 2 SN démonstratifs 1 pronom personnel **40 SN sans déterminant** **5 zéros**	6 SN indéfinis 1 SN démonstratif 11 pronoms personnels **23 SN définis** **9 SN sans déterminant**

La comparaison des chaînes des deux référents non humains nous a permis de constater que les zéros peuvent référer à des entités non humaines avec des proportions significatives ; cependant, les pronoms personnels sont beaucoup moins utilisés en chinois qu'en français pour référer aux entités non humaines.

7.4.2. Chaînes des notions spatio-temporelles

Les référents spatiaux ainsi que les référents temporels fixent ensemble les cadres spatio-temporels de chaque intrigue dans la narration, dans lesquels s'inscrivent tous les autres types de référents humains et non humains. Ce groupe d'expressions référentielles ne constituent pas des chaînes longues et régulières, puisqu'elles apparaissent *a priori* au début de chaque changement spatio-temporel, sans la nécessité d'être reprises dans la suite du récit.

Selon notre recensement, 16 épisodes sur 50 contiennent une ou quelques chaînes spatiales ; elles totalisent 85 maillons. Quelques lieux spécifiques sont récurrents dans plusieurs épisodes, tels que « la ville » et « la maison », comme le montre l'extrait [60] :

[60]Un soir que Melchior dînait en [**ville**], Gottfried, resté seul dans la salle du bas, tandis que Louisa couchait les deux petits, sortit, et alla s'asseoir à quelques pas de [**la maison**], [**au bord du fleuve**].

Rappelons que selon la définition du *roman passif* de Leblond (2010) présentée dans la première section du *Chapitre 6*, les histoires racontées dans le roman-fleuve suivent l'ordre temporel de la réalité, sans mettre en relief des intrigues tendus et dramatiques. Les expressions spatio-temporelles se situent régulièrement au début des unités typographiques. D'une part, elles peuvent servir de marqueurs des changements spatio-temporels et de la progression des intrigues ; d'autre part, les expressions spatiales sont plus récurrentes que celles du temps parce que les premières peuvent aussi accompagner l'introduction des référents qui se situent dans d'autres espaces.

En tant que langue analytique, le chinois ne possède pas de système de conjugaison des verbes, qui permettrait d'exprimer le temps. Une hypothèse porte ainsi sur les expressions référentielles renvoyant au temps : le chinois contiendrait plus d'expressions temporelles que le français afin de pallier l'absence d'indication temporelle des verbes. Pourtant, le résultat de l'annotation ne confirme pas cette hypothèse : 5 % des expressions référentielles chinoises expriment des notions temporelles, ce qui est relativement proche de la proportion du français, qui s'élève à 4 % ; 0,3 % de ces expressions sont des maillons de chaînes, proportion également assez proche et légèrement moins significative que celle du français, à 0,4 %. Toutes les chaînes sont assez courtes en se répartissant dans 1 ou 2 épisodes : aucune ne dépasse 10 maillons, ni 2 épisodes.

7.4.3. Chaînes des référents relatifs au corps

Les référents relatifs au corps ne concernent pas seulement des parties du corps comme « sa tête » et « ses pieds », mais également des entités relatives comme « ses larmes » et « son esprit ». Nous utiliserons « parties du corps » pour désigner toutes ces entités relatives au corps. Les chaînes des parties du corps sont très nombreuses et assez courtes. Leurs maillons sont très épars dans le texte ; les formes nominales sont dominantes.

Les parties du corps sont des référents inaliénables. Il s'agit d'un groupe de chaînes de toutes les parties du corps de tous les personnages, totalisant 143 maillons, comme « le cou (de Christophe) », « les doigts (de Christophe) » et « la mâchoire (de Christophe) » de l'extrait [61].

[61]Il se balance, il se tord [**le cou**] pour regarder au plafond, il fait des grimaces, il tire grand-père par son habit, il étudie les pailles de sa chaise, il tâche d'y faire un trou avec [**ses doigts**], il écoute les cris d'oiseaux, il bâille à se décrocher [**la mâchoire**].

Selon de nombreuses études de la grammaire du français, ce sont les déterminants définis et les déterminants possessifs qui se chargent d'exprimer cette relation inaliénable entre un personnage et une partie de son corps[23].

Ces chaînes et leurs maillons en chinois attirent notre attention pour deux raisons. D'un côté, selon l'approche sémantique, les parties du corps font partie des objets concrets ; mais ces référents présentent une similitude particulièrement étroite avec les référents humains. Cette relation inaliénable est plus souvent exprimée par les SN sans déterminant (57 %) que par les SN possessifs (12 %) en chinois.

D'un autre côté, selon l'approche formelle, les expressions renvoyant aux parties du corps présentent des spécificités que d'autres objets ne possèdent pas. En effet, un nom de localisation peut être étroitement attaché à un nom de partie du corps en constituant une locution corporelle locative[24], comme le montre la formulation suivante, la préposition devant la locution est facultative ; c'est le nom de localisation qui marque le locatif :

(PREP) Locution corporelle = (PREP) nom de partie du corps – nom de localisation

Le nom de localisation n'est pas obligatoire dans toutes les locutions corporelles, mais il est très récurrent et obligatoire lorsque les parties du corps sont à la place de circonstants locatifs de phrases :

[62]他	听到	[自己的	心]	在	[胸	中]		忐忑乱跳，	[血液]		在	那里	沸腾，
ta	*tingdao*	*zijide*	*xin*	*zai*	*xiong*	*zhong*		*tanteluantiao*	*xueye*		*zai*	*nali*	*feiteng*
3P	entendre	son	cœur	PREP	poitrine	dedans		agiter	sang		PREP	là-bas	bouillonner
		partie-corps-CDO			*partie-corps-loc-circonstant*				*partie-corps-sujet*				

[脸	上]	给	一	阵阵	的	狂	风	吹着。
lian	*shang*	*gei*	*yi*	*zhenzhen*	*de*	*kuang*	*feng*	*chuizhe*
visage	dessus	par	un	CL	DE	frénétique	vent	soufflant
partie-corps-loc-circonstant								

'Il entendait les battements de ce cœur forcené qui saute dans la poitrine, ce sang tumultueux, il sentait sur sa face ces coups de vent frénétiques.'

Nous avons constaté dans notre corpus des expressions comme 'cœur-dedans' *xin-li* 心里, 'corps-dessus' *shen-shang* 身上 et 'pieds-sous' *jiao-xia* 脚下. Tout comme les cas de contraction en français entre une préposition et un article défini (comme « du », « des », « au » et « aux », *cf. Manuel DEMOCRAT*, 2018 : 12), ces noms de localisation sont considérés comme composants obligatoires des locutions corporelles en chinois, puisque les locutions seront agrammaticales après leur suppression. Comme le montre le test [63]*, les noms seuls 'poitrine' *xiong* 胸 et 'visage' *lian* 脸 ne sont pas acceptables selon la grammaire chinoise :

[63]*他	听到	[自己的	心]	在	[胸]	忐忑乱跳，	[血液]	在	那里	沸腾，	[脸]	给	一
ta	*tingdao*	*zijide*	*xin*	*zai*	*xiong*	*tanteluantiao*	*xueye*	*zai*	*nali*	*feiteng*	*lian*	*gei*	*yi*
3P	entendre	son	cœur	PREP	poitrine	agiter	sang	PREP	là-bas	bouillonner	visage	par	un
		partie-corps			*partie-corps*		*partie-corps*				*partie-corps*		

阵阵	的	狂	风	吹着。
zhenzhen	*de*	*kuang*	*feng*	*chuizhe*
CL	DE	frénétique	vent	soufflant

[23] *Cf.* Dubois & Lagane (1973 : 65), Tellier (1992 : 166), Corblin (1987), De Mulder (1994), Guéron (2003 : 189), Vinet & Zhou (2003 : 169), entre autres.

[24] Ce type d'attachement de nom de localisation concerne également des locutions spatiales et des locutions temporelles (*cf.* Zhang, 张旺熹, 2013 : 195).

7.5. Bilan

Dans ce chapitre, nous avons examiné les chaînes de référence dans le corpus *Jean-Christophe* afin de dégager les régularités et spécificités des chaînes de référence chinoises dans le genre roman-fleuve. Ces chaînes sont subdivisées en quatre groupes selon la catégorie sémantique du référent et selon leur importance dans le roman-fleuve : la chaîne du protagoniste constitue le premier ; les trois chaînes des personnages secondaires le deuxième ; les chaînes des autres personnages, dits épisodiques et éphémères, comptent 104 chaînes, et constituent le troisième groupe ; les chaînes des autres référents non humains, totalisant 236 chaînes, forment le quatrième et dernier.

L'échelle des chaînes de référence est évidente non seulement selon leur nombre de maillons, mais également selon leur longueur et leur répartition dans le roman-fleuve : la longueur est radicalement différente entre chaque groupe de chaînes : celle du protagoniste compte 3817 maillons, dépassant largement celle des autres groupes; celle des chaînes secondaires totalise entre 472 – 859 maillons, beaucoup moins élevée que celle du protagoniste mais largement supérieure à celle des autres personnages qui ne dépasse pas 275 maillons ; celle des chaînes des référents non humains est inférieure à 49 maillons. Sans parler de différents degrés d'importance des référents non humains pour le roman-fleuve, toutes les chaînes de ce groupe ne contiennent pas plus de 50 maillons. En comparaison de la longueur des chaînes des personnages, ces longueurs sont peu significatives rapportées à celles des référents humains.

La répartition des maillons reflète la même tendance que la longueur : la chaîne du protagoniste est la seule dont les maillons répartissent dans presque tout le corpus ; celle des chaînes secondaires est présente dans environ une moitié du corpus ; les chaînes des autres personnages et des autres référents non humains ne sont pas continues : elles ne couvrent que quelques épisodes dans le corpus.

En ce qui concerne l'introduction de nouveaux référents, diverses formes sont observées à la place de premier maillon. Le protagoniste et les personnages secondaires sont introduits par les SN sans déterminant et les SN possessifs. Pour les chaînes des autres personnages et des référents non humains, ce sont toujours les SN sans déterminant qui occupent le plus souvent la place de premier maillon. Certaines autres catégories présentent aussi des occurrences. Par exemple, les formes non nominales introduisent régulièrement les chaînes des personnages fusionnés, équivalent à 22 % des chaînes épisodiques et éphémères, puisque leur premier maillon sert seulement à reprendre les anciens référents comme une seule entité : aucun nouvel individu n'est concerné dans cette fusion ; les SN indéfinis sont utilisés pour introduire les référents moins importants, équivalant à 21 % des chaînes épisodiques et éphémères et 22 % des chaînes des référents non humains.

Tableau (7.24) : Paramètres des chaînes de référence chinoises selon le référent

Référent	Nombre	Longueur	Répartition	Premier maillon privilégié
Protagoniste	1 chaîne	3817 maillons	47 épisodes	SN sans déterminant (100 %)
Personnages secondaires	3 chaînes	472 – 859 maillons	27 – 32 épisodes	SN sans déterminant (67 %), SN possessif (33 %)
Personnages épisodiques et éphémères	104 chaînes	3 – 275 maillons	1 – 12 épisodes	SN sans déterminant (32 %), non nominaux (22 %), SN indéfinis (21 %)
Non humains	236 chaînes	3 – 49 maillons	1 – 8 épisodes	SN sans déterminant (56 %), SN indéfinis (22 %)

Quant à la composition des chaînes de référence chinoises, elles présentent des caractéristiques très diversifiées.

La chaîne du protagoniste est assez stable au vu de la variété des déterminants des maillons nominaux : les noms propres (sans déterminant *a priori*), les SN sans déterminant et les maillons non nominaux occupent plus de 99 % de composants ; le reste de moins de 1 % est partagé entre les maillons nominaux avec trois déterminants : les démonstratifs, les possessifs et les indéfinis.

Les chaînes secondaires présentent une tendance similaire, avec des SN dont les déterminants sont stables, mais la proportion des SN sans déterminant (21 %) est beaucoup plus élevée que dans la chaîne du protagoniste (6 %) : les personnages secondaires sont beaucoup plus souvent mentionnés par leur statut social, familial, professionnel et des caractéristiques qualitatives.

La composition des chaînes des personnages épisodiques et éphémères est par contre assez différente et plus variable au vu de la proportion relativement haute des divers SN : les SN possessifs, démonstratifs et indéfinis partagent 8 % de composants, proportion beaucoup plus significative que celle des autres chaînes.

Pour les chaînes des référents non humains, la composition est radicalement différente: les noms propres y sont presque absents ; les SN sans déterminants (56 %) et avec divers déterminants (21 %) sont les deux types de composants privilégiés ; les formes non nominales (23 %), y compris les zéros (7 %), ne sont plus les composants principaux de ce groupe de chaînes.

Tableau (7.25) : Composants des chaînes de référence en chinois selon le référent

Référents	Noms propres	SN sans déterminant	Autres SN	Zéros	Autres non nominaux
Protagoniste	11 %	6 %	0,5 %	27 %	56 %
Personnages secondaires	12 %	21 %	1 %	28 %	39 %
Personnages épisodiques et éphémères	9 %	22 %	8 %	24 %	37 %
Non humains	0,1 %	56 %	21 %	7 %	16 %

Chapitre 8 : Synthèse comparative

En raison du parallélisme des textes franco-chinois qui caractérise notre corpus, les chaînes de référence dans les deux langues s'avèrent avoir beaucoup de points communs et similaires. La comparaison des paramètres dans les deux chapitres précédents a permis de mettre en lumière des différences qui nous paraissent non négligeables car elles sont en rapport direct avec le sujet de notre thèse. Dans ce chapitre, nous proposons un bilan comparatif pour synthétiser les spécificités contrastives et partagées des chaînes de référence dans chaque langue, accompagné d'explications relatives à différentes approches qui nous semblent pertinentes.

8.1. Référents du monde réel reflétés dans un roman-fleuve en français et en chinois

Les référents mentionnés dans une œuvre littéraire, dans sa version originelle et dans sa version traduite, sont censés être les mêmes puisque la définition de *référent* est commune pour toutes les langues. Pourtant, l'observation des faits dans le corpus nous permet de remarquer qu'il ne s'agit là que d'une supposition théorique : les référents ne sont pas identiques dans les textes parallèles en deux langues ; ils ne sont pas non plus exprimés de même manière en français et en chinois. Dans cette partie, nous comparons les référents contenus dans les deux langues pour clarifier les différences, significatives ou non, en cherchant à les expliquer par diverses approches.

8.1.1. Nombre et densité des expressions référentielles, des référents et des chaînes de référence

L'état global des expressions référentielles et des chaînes de référence dans le roman-fleuve *Jean-Christophe* en français et sa traduction en chinois est présenté dans le tableau (8.1). Trois critères sont utilisés pour les comparer : le nombre, la proportion et la densité. Quel que soit le critère choisi, les textes en français montrent une tendance générale à dépasser celui du chinois : en français, le texte contient plus d'expressions référentielles, plus de référents et plus de chaînes. La densité référentielle reflète le ratio entre le nombre d'expressions et le nombre de mots / caractères du texte. Si ces nombres supérieurs en français sont relatifs au nombre de mots du texte, la densité d'expressions référentielles et de maillons en français est toujours plus élevée qu'en chinois.

Tableau (8.1) : Nombres, pourcentages et densités des expressions référentielles en français et en chinois

Langue	Fr		Ch	
Critères	Pourcentage / nbre	Densité	Pourcentage / nbre	Densité
ER	17367	0,281	16407	0,166
Singletons	48 % (8320)	-	44 % (7176)	-
Maillons	52 % (9687)	0,157	56 % (9231)	0,096
Référents	7875	-	7343	-
CR	9 % (679)	0,071	9 % (656)	0,042

Au vu de cet état général, les expressions référentielles sont plus remarquables dans le texte en français que dans son texte parallèle en chinois. La raison principale de ces chiffres relativement bas en chinois semble être liée aux différences des structures syntaxiques ainsi que des catégories référentielles et non référentielles entre les deux langues. Le chinois présente une légère tendance à ne pas mentionner les entités non concrètes par des catégories grammaticales dites référentielles. Les différences ui se situent aux niveaux syntaxique et morphologique entraînent donc, d'une langue à l'autre, la disparition de certaines expressions référentielles ainsi que celle de leurs référents.

D'un côté, certaines structures syntaxiques apportent moins d'expressions référentielles en chinois qu'en français. Comparons par exemple les extraits parallèles [1] et [2] : le pronom « lui » (« ER-6-réf-1-M-3 ») est utilisé pour s'accorder au verbe « déplaire » dans la structure SVO en français, alors que ce fait est traduit par l'adjectif 'déplaisant' *saoxing* 扫兴 dans la structure attributive en chinois.

[1] Bien qu'**[il]** ne se permît pas de discuter **[les actes de [[son] héros]]**, **[il]** **[lui]**
ER-1-réf-1-M-1 *ER-2-réf-2* *ER-3- réf-1-M-2* *ER-4-réf-3-M-1* *ER-5-réf-3-M-2* *ER-6- réf-1-M-3*
déplaisait que **[celui-ci]** eût pensé à **[des choses laides]**, quand il ne devait y avoir,
ER-7-réf-3-M-3 *ER-8-réf-4*
[ce soir-là], que **[[des figures] et [des pensées lumineuses]]**.
ER-9-réf-5 *ER-10-réf-6* *ER-11-réf-7* *ER-12- réf-8*

[2]

虽然	[他]	崇拜	[哈斯莱]	,	[Ø]	不	敢	议论	[[他的]	行为]	,	可是	[他]
suiran	*ta*	*chongbai*	*Hasilai*			*bu*	*gan*	*yilun*	*tade*	*xingwei*		*keshi*	*ta*
bien-que	3P	apprécier	Hassler			NEG	oser	discuter	son	acte		mais	3P
	ER-1-réf-1-M-1		*ER-2-réf-2-M-1*		*ER-3-réf-1-M-2*				*ER-4-réf-2-M-2*	*ER-5-réf-3*			*ER-6-réf-1-M-3*

觉得	[今天	晚上]	[Ø]	应当	和颜悦色	,	只有	[些	快乐的	念头]	才对	,	[哈斯莱]
juede	*jintian*	*wanshang*		*yingdang*	*heyanyuese*		*zhiyou*	*xie*	*kuailede*	*niantou*	*caidui*		*Hasilai*
penser	aujourd'hui	soir		devoir	plaisant		seulement	des	lumineux	pensée	juste		Hassler
	ER-7-réf-4		*ER-8-réf-5*					*ER-9-réf-6*					*ER-10-réf-2-M-3*

想到	[那些	丑恶的	事]	未免太	扫兴	了	。
xiangdao	*naxie*	*chouede*	*shi*	*weimiantai*	*saoxing*	*le*	
penser	ces-là	laid	chose	trop	déplaisant	LE	
	ER-11-réf-7						

D'un autre côté, alors que certaines notions sont présentes sous forme d'expressions référentielles en français, elles le sont sous forme d'expressions non référentielles en chinois. Par exemple, dans l'extrait [3], le SN possessif « leurs interminables bavardages » en français est équivalent au verbe 'bavarder' *tan* 谈 dans l'extrait parallèle [4] en chinois.

[3] Grand-père et le voiturier continuaient sans se lasser **[leurs interminables bavardages]**.
référentielle

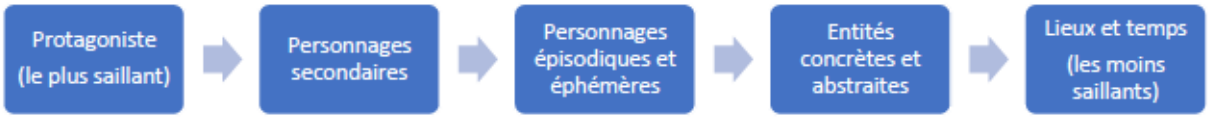
[4]

祖父	跟	赶车的	拉拉扯扯的	老是	**谈**不完。
zufu	*gen*	*ganchede*	*lalachechede*	*laoshi*	*tanbuwan*
grand-père	et	voiturier	sans-se-lassant	toujours	bavarder-NEG
					non-référentielle

8.1.2. Référents communs et distincts

Les chaînes des référents humains et non humains sont en principe identiques dans les deux langues. Nous établissons une échelle commune de la saillance des référents constatés dans notre corpus en français et en chinois, comme le montre la figure (8.1) : les humains sont subdivisés par le *protagoniste* (le plus saillant) les *personnages secondaires* et les *personnages épisodiques et éphémères* ; les référents non humains sont subdivisés par les *entités concrètes et abstraites* et les *lieux et temps* (les moins saillants) au vu de leur nombre et de leur répartition dans le corpus.

Figure (8.1) : Échelle de la saillance des référents dans *Jean-Christophe* en français et en chinois



Pour expliquer pourquoi les référents d'humains sont identiques dans différentes langues, Hamon (1972) propose de subdiviser ces humains selon les critères sémiologiques, par lesquels les humains sont *personnages-référentiels*, *personnages-embrayeurs* et *personnages-anaphores* avec les définitions suivantes :

> *« Personnages-référentiels » : personnages historiques (Napoléon III dans* les Rougon-Macquart, *Richelieu chez A. Dumas…), mythologiques (Vénus, Zeus…), allégoriques (l'Amour, la Haine…) ou sociaux (l'ouvrier, le chevalier, le picaro…). Tous renvoient à un sens plein et fixe, immobilisé par une culture, et leur lisibilité dépend directement du degré de participation du lecteur à cette culture (ils doivent être* appris *et* reconnus*).*
>
> *« Personnages-embrayeurs » : ils sont les marques de la présence en texte de l'auteur, du lecteur, ou de leurs délégués : personnages « porte-parole », chœurs de tragédies antiques, interlocuteurs socratiques, personnages d'*Impromptus, *conteurs et auteurs intervenant, Watson à côté de Sherlock Holmes, etc. Le problème de leur repérage sera parfois difficile. Là aussi, du fait que la communication peut être différée (textes écrits), divers effets de brouillage ou de masquages peuvent venir perturber le décodage immédiat du « sens » de tels personnages (il est nécessaire de connaître les présupposés, le « contexte » : a priori, l'auteur par exemple n'est pas moins présent derrière un « il » que derrière un « je »).*
>
> *« Personnages-anaphores » : ici une référence au système propre de l'œuvre est seule indispensable. Ces personnages tissent dans l'énoncé un réseau d'*appels *et de* rappels *à des segments d'énoncés disjoints et de longueur variable (un syntagme, un mot, un paragraphe…) ; éléments à fonction essentiellement organisatrice et cohésive, ils sont en quelques sorte les signes mnémotechniques du lecteur : personnages de prédicateurs, personnages doués de mémoire, personnages qui sèment ou interprètent des indices, etc.*
>
> *(Hamon, 1972 : 95-96)*

Selon cette typologie, le protagoniste et les personnages secondaires semblent s'accorder à la description de *personnages-embrayeurs* dans le roman-fleuve, puisque les lecteurs n'ont pas besoin de connaissances encyclopédiques ou d'un contexte extra-linguistique spécifique pour comprendre ce que signifient les noms propres comme « Jean-Christophe » et « Jean-Michel ». Les autres humains correspondent aux définitions de *personnages-référentiels* et de *personnages-anaphores*. Pour ceux qui sont personnages-référentiels, ces humains sont directement repérables dans le monde extra-linguistique grâce aux connaissances encyclopédiques les lecteurs : tel est le cas des humains génériques, comme « les enfants » et « les musiciens », et des humains historiques, comme « Napoléon » et « Mozart ». Pour ceux qui sont personnages-anaphores, à savoir que les rappels contextuels sont obligatoires pour pouvoir les identifier, il s'agit de la plupart de personnages spécifiques du roman-fleuve, comme « grand-duc » et « Ernst », et la plupart de groupes de personnages faisant l'objet de chaînes fusionnées, comme « Jean-Michel & Melchior » et « Louisa & Jean-Christophe », dont tous les personnages ne réfèrent pas à quelqu'un dans le monde réel et nécessitent un antécédent.

Malgré les proportions assez proches de chaque groupe de chaînes dans les deux langues, comme le montre le tableau (8.2), une grande différence se trouve dans le groupe des expressions des référents abstraits : 24 % en français contre 17 % en chinois. Pour expliquer cet écart, reprenons les exemples parallèles [1] et [2] dans la section précédente : excepté le nombre total de référents, plus élevé en français qu'en chinois (12 contre 11), trois référents abstraits en français (« des figures », « des pensées lumineuses » et « des figures et des pensées lumineuses ») correspondent à un seul référent abstrait en chinois ('des joyeux idée' *xie-kuailede-niantou* 些快乐的念头) : lorsque le français cherche à préciser les entités abstraites par diverses expressions, le chinois présente la tendance à généraliser ces mêmes entités, sans expliciter les détails des notions potentielles comprises.

Tableau (8.2) : Catégories des référents dans le corpus en français et en chinois

Expressions	ER		Maillons	
Langue	Fr	Ch	Fr	Ch
Humains	8736 (50 %)	8335 (51 %)	8098 (84 %)	7680 (83 %)
Référents concrets	3022 (17 %)	3183 (19 %)	843 (9 %)	858 (9 %)
Référents abstraits	**4081 (24 %)**	**2789 (17 %)**	504 (5 %)	425 (5 %)
Lieux	903 (5 %)	1329 (8 %)	184 (2 %)	212 (2 %)
Temps	590 (3 %)	771 (5 %)	68 (0,7 %)	55 (0,6 %)

8.1.3. Référents indépendants, référents associés et référents inaliénables : possessif obligatoire et facultatif

Les référents présents dans une même œuvre littéraire ne peuvent pas être complètement indépendants les uns des autres ; des relations les relient, comme la possession, la parenté et la relation partie-tout. D'après l'observation des faits, ce sont surtout les possessifs qui établissent ces relations entre les référents humains et non humains comme « Christophe … son lit » (possession), « Christophe … sa mère » (parenté) et « Christophe … ses pieds » (relation partie-tout). Ils n'apportent pas d'ambiguïté et n'ajoutent pas de

répétitions de complément du nom comme « Christophe … la mère de Christophe ». Les possessifs peuvent sans aucun doute d'être une garantie pour établir ces relations inter-référentielles, mais, la comparaison de leur fréquence, y compris les SN possessifs et les déterminants possessifs, permet de remarquer que ces relations sont plus explicites en français qu'en chinois, vu que les expressions possessives sont beaucoup moins nombreuses en chinois qu'en français, au niveau de l'intégralité du corpus, au niveau des maillons des chaînes et à celui de la chaîne du protagoniste, comme montre le tableau (8.3). Par conséquent, les référents ont tendance à être mentionnés comme des entités indépendantes les unes des autres en chinois, tandis que, en français, ils sont présentés comme des entités associées ou des entités inaliénables les unes aux autres (spécifiquement pour les expressions renvoyant aux parties du corps) de manière plus limpide et directe.

Tableau (8.3) : Catégories des expressions référentielles dans le corpus en français et en chinois

Pourcentage de	ER		Maillons		Maillons JC	
Langue	Fr	Ch	Fr	Ch	Fr	Ch
Noms propres	5 %	4 %	8 %	6 %	11 %	11 %
SN définis	23 %	-	11 %	-	3 %	-
SN sans déterminant	6 %	39 %	2 %	22 %	0,2 %	6 %
SN démonstratifs	3 %	3 %	1 %	2 %	0,2 %	0,3 %
SN possessifs	7 %	4 %	4 %	2 %	1 %	0,1 %
SN indéfinis	12 %	9 %	2 %	2 %	0,2 %	0,1 %
Déictiques	3 %	3 %	4 %	4 %	4 %	4 %
Déterminants possessifs	8 %	4 %	13 %	6 %	16 %	8 %
Pronoms personnels	27 %	18 %	46 %	31 %	60 %	44 %
Zéro	3 %	15 %	5 %	25 %	4 %	27 %
Autres formes	4 %	2 %	3 %	1 %	0,2 %	0,1 %

8.1.4. Corrélation entre la catégorie d'expression et la saillance du référent

Les catégories d'expressions référentielles sont un des sujets primordiaux pour les études comparatives ainsi que pour notre observation dans ce corpus franco-chinois. Chacune d'elles présente ses particularités dans chaque langue, dans des contextes spécifiques et à différentes fins. Selon le bilan proposé dans le tableau (8.3), les catégories peuvent être regroupées en quatre types en fonction de la saillance de leur référent.

Trois catégories, dans les deux langues, se trouvent dans le premier groupe : les noms propres, les déterminants possessifs et les pronoms personnels. Leur proportion est la plus élevée pour toutes les expressions référentielles (les deux premières colonnes du tableau), celle des maillons à la place intermédiaire (les deux colonnes au milieu du tableau) et celle des maillons du protagoniste la plus basse (les deux dernières colonnes du tableau).

Le deuxième groupe concerne la plupart des autres catégories qui manifestent la tendance inverse : les SN définis (français), les SN sans déterminant, les SN démonstratifs, les SN possessifs et les SN indéfinis. Ces formes sont le plus souvent utilisées pour les expressions référentielles et le moins souvent pour les maillons du protagoniste.

Une seule catégorie constitue le troisième groupe : elle ne présente pas de tendance remarquable relativement aux trois groupes d'expressions de référents de différente saillance :

leur proportion est constante pour les référents des singletons et des chaînes ; c'est le cas des expressions déictiques dans les deux langues.

Une autre catégorie représente le dernier groupe : les zéros. Pour le français, la proportion de cette forme vide est relativement constante dans les trois groupes d'expressions, alors que pour le chinois, leur proportion augmente fortement au gré de la saillance du référent : ils sont peu utilisés pour les singletons mais engagent une proportion très importante dans la chaîne du protagoniste.

Parmi ces catégories, certaines présentent des proportions d'usage très proches ; d'autres affichent de grandes différences entre le français et le chinois ; d'autres encore manifestent des tendances singulières. Trois d'entre elles sont similaires : il s'agit des noms propres, des SN démonstratifs et des déictiques ; six autres sont proportionnellement très différentes : les SN définis, les SN sans déterminant, les SN possessifs, les déterminants possessifs, les pronoms personnels et les zéros. La seule catégorie restante est celle des SN indéfinis, dont les proportions dans les trois groupes d'expressions présentent des tendances différentes : leurs proportions sont proches pour les maillons, mais les singletons français sont beaucoup plus souvent récurrents.

8.2. Continuité des chaînes et progression du roman-fleuve

Par rapport aux référents présents dans le corpus, le parallélisme des chaînes de référence dans les deux langues est plus évident. Comme le montre le tableau (8.4), leurs différences inter-linguistiques ne sont pas significatives : la chaîne du protagoniste couvre l'intégralité du corpus, du premier épisode jusqu'à la fin ; les autres présentent aussi des répartitions et des longueurs très proches.

Tableau (8.4) : Répartition et longueur des chaînes de référence en français et en chinois

Référent	Répartition		Longueur	
Langue	Fr	Ch	Fr	Ch
Protagoniste	49 épisodes	47 épisodes	4366 maillons	3817 maillons
Personnages secondaires	27 – 32 épisodes	27 – 32 épisodes	511 – 917 maillons	472 – 859 maillons
Personnages épisodiques et éphémères	1 – 9 épisodes	1 – 12 épisodes	3 – 287 maillons	3 – 275 maillons
Non humains	1 – 7 épisodes	1 – 8 épisodes	3 – 50 maillons	3 – 49 maillons

La portée de la chaîne du protagoniste est importante pour le roman-fleuve, puisque c'est sur cette base topicale que le discours présente une continuité référentielle. Selon la définition de Li & Thompson (1979) et de Givón (1983), cette chaîne du protagoniste constitue la chaîne topicale de ce roman-fleuve dans les deux langues. Au fondement de ce topique, les autres référents apparaissent sous formes de singletons ou constituants des chaînes, ce qui permet au roman-fleuve de progresser au fur et à mesure avec de nouvelles informations. Dans cette section, nous analyserons la manière dont les chaînes permettent la continuité et la progression dans le roman-fleuve en français et en chinois.

8.2.1. Premier maillon dans la chaîne de référence pour introduire les nouveaux référents

Le premier maillon présente diverses formes, certaines étant privilégiées selon la catégorie sémantique du référent et selon la langue, comme on peut le voir dans le tableau (8.5). Étant donné le petit nombre de chaînes globales des humains, les premiers maillons ne présentent que deux formes : le SN défini et le SN possessif. La variation formelle est plus significative pour les chaînes des autres groupes étant donné leur nombre de chaînes plus élevé.

Dans les chaînes locales des humains, les différences les plus significatives sont constatées au niveau des SN définis et des SN sans déterminant. Les 16 % de SN définis français au premier maillon des chaînes sont remplacés par diverses formes en chinois ; 6 % de SN sans déterminant français sont présents, contre 32 % de SN sans déterminant chinois. Les maillons non nominaux semblent pouvoir démarrer beaucoup de chaînes, mais leurs référents sont réservés aux groupes de personnages fusionnés et aux suites des pronoms personnels indéfinis dans les deux langues : 26 % en français et 22 % en chinois.

Dans les chaînes locales des référents non humains, ce sont toujours ces deux catégories qui présentent les plus grandes différences : les 45 % de SN définis français sont remplacés par diverses autres catégories en chinois ; 6 % de SN sans déterminant français initient les chaînes des référents non humains, contre 56 % de SN sans déterminant en chinois.

Tableau (8.5) : Premier maillon des chaînes de référence en français et en chinois

Référents	Chaînes globales des humains		Chaînes locales des humains		Chaînes des référents non humains	
Langue	Fr	Ch	Fr	Ch	Fr	Ch
Noms propres	-	-	16 %	13 %	3 %	1 %
SN définis	75 %	-	16 %	-	45 %	-
SN sans déterminant	-	75 %	6 %	32 %	6 %	56 %
SN démonstratifs	-	-	4 %	5 %	5 %	7 %
SN indéfinis	-	-	19 %	21 %	21 %	22 %
SN possessifs	25 %	25 %	13 %	7 %	12 %	7 %
Non nominaux	-	-	26 %	22 %	8 %	7 %

8.2.2. Noms propres : interprétations sémantique et cognitive selon différents contextes linguistiques et culturels

Au niveau sémantique, les noms propres français issus de *Jean-Christophe* sont des expressions référentielles relativement neutres, sans apports relatifs aux traits ontologiques spécifiques des personnages. Contrairement aux noms communs, ces noms propres du roman-fleuve (« Christophe », « Louisa », « Melchior », par exemple) ne reflètent pas directement ni les traits physiques ni les identités sociales et familiales des personnages.

Lorsque ces noms propres français sont traduits vers le chinois, comme 克利斯朵夫, par exemple, avec la transcription en pinyin /ke-li-si-duo-fu/, cette suite de caractères n'est qu'une représentation phonétique de la prononciation de « Ch-ri-s-to-phe », dont le sens étymologique en français, « qui porte le Christ », est complètement perdu : les 5 caractères de ce nom propre chinois constituent visuellement le nom d'une personne étrangère, mais ils n'apportent aucune signification équivalente à celle du français. Dispensés de contraintes morphologiques et syntaxiques, les noms propres sont ainsi transcrits, dans ce roman-fleuve, de manière phonétique au lieu d'être traduits de manière sémantique. De même, pour les autres équivalents chinois des noms propres français, les significations présentes dans les noms des personnages ne peuvent plus être exprimées à travers la stratégie phonétique du traducteur : 鲁意莎 'Louisa' /lu-yi-sha/ ne porte pas de sens « illustre » et « glorieux » ; 约翰·米希尔 'Jean-Michel' /yue-han-mi-xi-er/, deux prénoms combinés, ne signifie ni « Dieu est miséricordieux » dans le sens de « Jean », ni « qui est comme Dieu » dans le sens de « Michel ». Malgré les pertes de signification que cela entraîne, les noms propres d'œuvres étrangères sont *a priori* traduits en fonction de leur prononciation au lieu de leur signification dans la langue d'origine. Cette stratégie de traduction adoptée par Lei Fu dans notre corpus est ainsi celle qui est le plus souvent utilisée pour traduire les noms de personnages dans les autres œuvres étrangères vers le chinois.

Sans parler des formes non distinctes entre les noms communs et les noms propres en chinois, l'usage de ces derniers est influencé non seulement par des conditions discursives spécifiques, mais également par des conditions culturelles et parfois temporelles, à la différence des noms communs. Citons Hamon (1972), qui explique la *fonctionnalité différentielle* du nom propre de personnage par l'approche sémiologique :

> *Le héros est ici, en quelque sorte, enregistré comme tel à partir d'un corpus déterminé, et a posteriori ; une référence à la globalité de la narration et à la somme ordonnancée des prédicats fonctionnels dont il a été le support, et que la culture de l'époque valorise, est nécessaire. (Hamon, 1972 : 92)*

Cette spécificité sémiotique mise à part, Landragin (2011 : 84) propose une explication par le *critère culturel* qui peut servir à comprendre la notion de saillance : l'auteur souligne la saillance visuelle culturelle lorsque l'on voit une peinture ou une photographie. Selon l'individu et selon le contexte culturel, les gens ne sont pas sensibilisés et stimulés de la même manière en raison de leur appartenance à différentes communautés culturelles. Cette approche culturelle nous semble également pertinente pour interpréter la saillance linguistique en fonction de la communauté de lecteurs et de leur contexte culturel : lorsque les noms propres français, visés aux lecteurs francophones, sont transcrits vers le chinois, destinés aux lecteurs sinophones, ils perdent, dans un premier temps, leur forme saillante en majuscule, et, dans un second temps, leur signification inhérente saillante contenue dans leur étymologie.

Un roman-fleuve est lu avant tout par les lecteurs de différentes communautés culturelles et de différentes langues, d'où le contexte extra-linguistique. *Jean-Christophe*, étant œuvre littéraire française traduite vers le chinois, constitue une passerelle entre deux cultures, deux communautés de lecteurs ainsi que deux langues. Les référents familiers d'une communauté peut être exotiques pour l'autre, d'où un des facteurs extra-linguistique qui influence la formulation des expressions renvoyant aux mêmes référents. Si « Jean-Christophe » est un prénom familier pour les lecteurs français et francophones, la transcription en caractères chinois est en elle-même un moyen de conférer un sens exotique aux lecteurs chinois et sinophones. Ces derniers, à la seule lecture du titre du roman-fleuve, peuvent ainsi

entrevoir qu'aussi bien les personnages que les référents peuvent s'inscrire dans un contexte culturel exotique, et, par conséquent, que leur interprétation pourra leur demander plus d'efforts en tant que lecteurs non francophones. Dans les textes traduits du français vers le chinois, les noms propres de personnes et d'endroits paraissent complètement exotiques pour les lecteurs chinois. Par opposition, dans le roman-fleuve original rédigé en français, les lecteurs francophones peuvent avoir de moins fort sentiment exotique par rapport aux lecteurs non francophones. Quant à la traduction des expressions référentielles et coréférentielles, l'égalité du sens transcende celle de la forme. Cela constitue une différence importante des référents entre les textes en français et ceux en chinois.

8.2.3. Pronoms personnels de troisième personne et zéros dans la continuité référentielle

Les pronoms personnels de troisième personne sont très présents dans les deux langues, avec respectivement 46 % en français et 31 % en chinois dans la totalité des expressions référentielles du corpus (voir le tableau (8.3) en plus haut). Ils occupent la première place de toutes les formes de maillons. L'emploi de ces pronoms peut signifier un haut degré d'accessibilité du référent (*cf*. Ariel, 1990). Les pronoms permettent de maintenir la continuité du même topique dans le discours (*cf*. Givón, 1983, Kleiber, 1994a, Combettes, 1986 ; Charolles, 1997, 2002 ; Corblin, 2005 ; Rohde & Kehler, 2014). Combettes (1986) et Kleiber (1994a) indiquent que le pronom « il » français peut être considéré comme marqueur de continuité thématique parce qu'il représente une entité déjà mentionnée dans le texte ; lors de l'apparition d'un pronom personnel, les lecteurs comprennent que l'on continue à parler d'une même entité.

> *« Il » peut être considéré comme un marqueur du topique discursif, c'est-à-dire une expression qui indique quel est le thème discursif en vigueur au moment de son apparition. […] Dire que « il » est une expression thématique ne signifie pas que « il » reprend le thème en vigueur avant son apparition. Il pourrait fort bien indiquer quel est le thème discursif qui vaut pour la phrase dans laquelle il se trouve, alors que la phrase précédente en possède un autre. (Kleiber, 1994a : 110)*

Par rapport aux formes nominales, ces pronoms n'expriment pas de traits sémantiques et qualitatifs concrets ; ils exigent donc, de la part des locuteurs, de garder à l'esprit les informations antérieures mentionnées par les autres expressions dans le discours. C'est pour cette raison que ces pronoms garantissent la continuité référentielle. D'après ces interprétations, il n'est pas surprenant que la chaîne du protagoniste soit constituée à 60 % de cette catégorie en français et 44 % en chinois, ce qui représente la plus grande proportion de maillons parmi toutes les formes nominales et non nominales dans les chaînes de référence.

La proportion du français est beaucoup plus élevée que celle du chinois puisqu'une partie considérable de la fonction de maintien de la continuité référentielle des pronoms explicites est partagée avec les zéros en chinois : 15 % d'expressions référentielles et 25 % de maillons des chaînes sont zéros en chinois, alors que ces proportions en français n'atteignent que 3 % et 5 %.

8.2.4. Expressions référentielles univoques et équivoques

La forme d'expression référentielle est un facteur décisif pour rendre limpide l'identité de référent. Manifestement, les vides risquent de provoquer plus d'ambiguïté que les formes explicites. Par rapport au français, le chinois présente beaucoup de vides, ceux des déterminants dans les SN et les zéros à proprement parler. Il convient donc de se demander ici quels types de vides entraînent une interprétation référentielle équivoque ou univoque.

- Déterminant zéro dans les SN

Les SN sans déterminant n'ont pas du tout la même interprétation référentielle en français et en chinois. 6 % de SN sans déterminant français sont constatés dans le corpus, mais seuls 2 % sont constituants des chaînes de référence. Cela implique que ce groupe de maillons est peu utilisé pour référer à des entités spécifiques et récurrentes dans le roman-fleuve. Pourtant, 39 % de SN sans déterminant chinois sont repérés dans le corpus. Cette proportion diminue fortement pour atteindre 22 % comme constituants des chaînes, mais elle demeure toujours très haute par rapport à toutes les autres catégories de maillons. Cela implique que les SN sans déterminant chinois peuvent référer à la fois à des entités récurrentes spécifiques des chaînes et à des entités génériques ou indéfinies des singletons. Lorsque les SN sans déterminant français ont tendance à référer de manière non spécifique, ceux du chinois ne présentent pas de propension significative : les déterminants vides dans les SN chinois peuvent référer éventuellement de manière équivoque si le contexte ne contient pas d'information suffisante pour identifier le référent unique.

- Les zéros

Les zéros en chinois occupent un pourcentage beaucoup plus élevé, à savoir 25 % de maillons, contre celui du français à seulement 5 %, cette petite partie correspondant aux emplois parallèles de zéros dans les deux langues sans différence, comme le montrent les extraits parallèles [5] et [6], alors que dans la plupart de situations, les zéros chinois sont équivalents des pronoms explicites français, ce qui est le cas montré dans les extraits parallèles [7] et [8].

[5] Quant à **[Rodolphe]**, **[il]** était d'une malice de singe ; […] **[il]** cassait les jouets, **[Ø]** *zéro-1* renversait l'eau, **[Ø]** *zéro-2* salissait sa robe, et **[Ø]** *zéro-3* faisait tomber les plats, en fouillant dans le placard.

[6] **[洛陶夫]** 却 象 猴子 一样的 狡猾， […]： **[Ø]** *zéro-4* 砸破 玩具， **[Ø]** *zéro-5* 倒翻 水， **[Ø]** *zéro-6* 弄脏 衣服， **[Ø]** *zéro-7* 在 壁橱 里 乱掏， **[Ø]** *zéro-8* 把 碟子 都 掉 在 地 下。

Luotaofu	*que*	*xiang*	*houzi*	*yiyangde*	*jiaohua*			*zapo*	*wanju*			*daofan*	*shui*			*nongzang*	*yifu*		*zai*	*bichu*	*li*	*luantao*		*ba*	*diezi*	*dou*	*diao*	*zai*	*di*	*xia*
Rodolphe	mais	comme	singe	pareil	malin			casser	jouet			renverser	eau			salir	robe		PREP	placard	dedans	fouiller		BA	assiette	tous	tomber	PREP	sol	dessous

[7] **[Il]** ne pouvait tourner la clef, **[il]** faiblissait, **[il]** appelait au secours.

[8] **[他]** 锁 不 了 门， **[Ø]** *zéro-9* 没有 气力 了， **[Ø]** *zéro-10* 只 能 喊 救命。

ta	*suo*	*bu*	*le*	*men*		*meiyou*	*qili*	*le*		*zhi*	*neng*	*han*	*jiuming*
3P	fermer	NEG	LE	porte		NEG	force	LE		seulement	pouvoir	crier	secours

Parmi ce grand nombre de zéros chinois, très peu d'occurrences provoquent des risques d'ambiguïté référentielle. Sauf les zéros concernés dans les suites des expressions équivoques que nous avons listés dans la section *7.3.2. Succession des expressions personnelles équivoques en chinois* du *Chapitre 7*, les zéros réfèrent assez clairement à des entités univoques en tant que constituants des chaînes des référents d'humains et non humains.

8.2.5. Critères relatifs à la variation des composants : hétérogénéité, stabilité, longueur et modifieurs

Les chaînes de référence dans le roman-fleuve sont *a priori* composites car constituées de différentes formes de maillons. Pour décrire ces variations, Schnedecker (1997 : 16, 2005 : 113) propose de distinguer les chaînes en deux groupes : celles qui sont *homogènes*, à savoir les chaînes composées massivement par des noms propres et des pronoms ; et celles qui sont au contraire *hétérogènes* : composées d'une variété de SN avec divers déterminants et toute autre forme de maillons. Selon le recensement des formes des composants, comme le montre le tableau (8.6), les proportions des noms propres sont relativement proches dans toutes les chaînes d'humains des deux langues (entre 9 % et 12 %). Les chaînes les plus homogènes sont celles du protagoniste, dont les proportions de maillons non nominaux sont les plus élevées (84 % en français et 83 % en chinois), malgré les petites proportions d'autres formes de maillons nominaux (5 % en français et 6 % en chinois). Les autres chaînes sont hétérogènes : celles des personnages épisodiques et éphémères contiennent des maillons nominaux en proportions relativement hautes, malgré les proportions de noms propres proches de celles du protagoniste ; les chaînes des référents non humains comportent très peu de noms propres (1 % en français et 0,1 % en chinois) mais beaucoup de maillons nominaux (62 % en français et 77 % en chinois). Pour ces deux derniers groupes, les chaînes en chinois semblent plus hétérogènes que celles du français.

Tableau (8.6) : Hétérogénéité des chaînes de référence en français et en chinois

Composants	Noms propres		Autres maillons nominaux		Maillons non nominaux	
Langue	Fr	Ch	Fr	Ch	Fr	Ch
Chaînes du protagoniste	11 %	11 %	5 %	6 %	84 %	83 %
Chaînes des personnages secondaires	12 %	12 %	22 %	23 %	66 %	65 %
Chaînes des personnages épisodiques et éphémères	11 %	9 %	23 %	30 %	66 %	61 %
Chaînes des non humains	1 %	0,1 %	62 %	77 %	37 %	23 %

Sachant que la variation des déterminants dans les expressions référentielles nominales en chinois est moins remarquable que celle du français (4 formes contre 5), nous nous nous posons la question de savoir si la variation au niveau de la tête lexicale ne serait pas plus ou moins variable en chinois qu'en français. Peut-on dire que les expressions nominales en chinois sont très stables, voire même monotones au niveau formel ? Nous avons donc

recensé le coefficient de stabilité des quatre chaînes globales, à savoir celles du protagoniste et des trois personnages secondaires, comme le montre le tableau (8.7).

En effet, le coefficient de stabilité ne montre pas de différence significative entre les deux langues. Les chaînes en français montrent simplement une légère tendance à être moins stables que celles du chinois : le coefficient de stabilité de la chaîne du protagoniste est de 23,5 en français (ses 706 maillons nominaux contiennent 30 noms différents comme têtes lexicales), légèrement moins élevé que celui du chinois à 25,6 ; la même tendance est présentée par les chaîne de « Louisa » et de « Jean-Michel » ; la chaîne « Melchior » est légèrement plus stable en français qu'en chinois.

Tableau (8.7) : Coefficient de stabilité des quatre chaînes de référence globales

	Nbre de maillons nominaux		Nbre de têtes lexicales		Coefficient de stabilité	
	Fr	Ch	Fr	Ch	Fr	Ch
JC	706	640	30	25	23,5	25,6
L	164	174	11	10	14,9	17,4
M	224	282	14	20	16,0	14,1
JM	248	275	11	11	22,5	25,0

Le troisième critère pour décrire la variation des composants des chaînes est celui de la longueur des expressions référentielles, qui dépend principalement des modifieurs de nom. D'après le recensement des longueurs de différents types d'expressions référentielles dans les deux langues, comme le montre le tableau (8.8), deux remarques sont obtenues. D'un côté, les maillons de la chaîne du protagoniste sont les plus courts en moyenne, avec seulement 1,6 mots français et 1,9 caractères chinois. Les singletons se composent de plus de mots dans les deux langues, avec la moyenne de 3,7 mots français et 4,3 caractères en chinois. La longueur réduit fortement lorsqu'il s'agit de celle des maillons dans la chaîne du protagoniste en raison de la différence de leur composition. Les maillons nominaux, composés souvent de plusieurs mots, n'occupent que 16 % des maillons en français et 17 % en chinois, tandis que cela n'est pas le cas pour les singletons. D'un autre côté, les longueurs recensées dans les trois colonnes du tableau (8.8) paraissent toujours relativement courte en français par rapport à celle du chinois. Pour la chaîne du protagoniste, les noms propres occupent une proportion relativement importante : « Christophe » est compté comme un mot en français, mais *Kelisiduofu* 克利斯朵夫 est compté comme cinq caractères en chinois. Malgré 27 % de zéros chinois contre 4 % de zéros français, les maillons du protagoniste sont en moyenne plus longues que ceux du français.

Tableau (8.8) : Longueurs moyennes des expressions référentielles en français et en chinois

Longueur moyenne de	ER		Maillons		Maillon JC	
Langue	Fr	Ch	Fr	Ch	Fr	Ch
Total	3,7	4,3	2,1	2,4	1,6	1,9

Le dernier critère concerne les modifieurs dans les maillons nominaux. Ils sont relativement peu utilisés dans les composants des chaînes, mais les modifieurs présentent une forte proportion d'emploi dans le corpus pour identifier le référent des singletons, comme le montrent les chiffres du tableau (8.9). Cette proportion descendante confirme les remarques

sur les modifieurs dans l'échelle d'accessibilité d'Ariel (1990), selon lesquelles la forme d'une expression reflète l'accessibilité du référent : les SN avec modifieurs renvoient aux référents les moins accessibles[1].

Tableau (8.9) : Maillons nominaux modifiés en français et en chinois

Pourcentage	ER		Maillons		Maillons JC	
Langue	Fr	Ch	Fr	Ch	Fr	Ch
SN modifiés	21 %	20 %	6 %	6 %	1 %	1 %
Noms propres modifiés	1 %	0,7 %	1 %	0,9 %	0,8 %	0,7 %

Les modifieurs peuvent être seuls ou plusieurs, longs ou courts, restrictifs ou non restrictifs. Pour ceux qui se trouvent dans les composants des chaînes, les modifieurs longs tendent à être non restrictifs pour spécifier le référent, alors que les modifieurs courts ne présentent pas cette tendance : ils peuvent être restrictifs ou non restrictifs selon le contexte. Les deux types de modifieurs le plus souvent utilisés avec les noms propres sont les adjectifs et les propositions subordonnées relatives. Les premiers peuvent exprimer aussi bien des qualités objectives que subjectives ; les seconds apportent des informations sur le plan de la narration au lieu d'avoir une influence au niveau de la référence.

La différence quantitative des maillons nominaux modifiés est peu significative entre le français et le chinois, ce en raison, principalement, des modifications des structures syntaxiques : certaines structures de modifieurs sont conservées dans les deux langues, alors que certaines propositions subordonnées relatives dans les maillons nominaux français sont remplacées par des maillons nominaux sans modifieur dans les structures syntaxiques linéaires en chinois. Ce type de changement est constaté dans toutes les chaînes :

[9] **[Louisa, qui comprenait l'enfant]**, fit : – Chut ! laisse-le dormir !
npr-mod

'**[鲁意莎]** 很 了解 孩子，说道："别 闹 了！让 他 睡觉！"'
Luyisha hen liaojie haizi shuodao bie nao le rang ta shuijiao
Louisa très comprendre enfant parler NEG troubler LE laisser 3P dormir
npr-sans-mod

[10]Melchior avait fait pressentir les intentions **[du prince, qui s'était montré gracieusement disposé**
SN-mod

à accepter l'hommage].
'原来 曼希沃 先 设法 探听 **[亲王]** 的 意思，**[亲王]** 表示 很乐意 接受 这
yuanlai Manxiwo xian shefa tanting qinwang de yisi qinwang biaoshi henleyi jieshou zhe
en-effet Melchior d'abord chercher pressentir prince DE intention prince montrer gracieux accepter ce
SN-sans-mod *SN-sans-mod*

个 敬意。'
ge jingyi
CL hommage

Par ailleurs, les modifieurs restrictifs et non restrictifs peuvent servir à la distinction des noms propres chinois. Nous avons vu que les modifieurs des noms propres français tendent à être non restrictifs lorsqu'il s'agit de spécifier le référent, alors que ceux des noms communs ne manifestent pas cette tendance : leurs modifieurs peuvent être restrictifs ou non restrictifs. Cette remarque nous permet de distinguer les noms propres chinois des noms communs étant donné que les deux groupes de noms ne disposent pas de formes distinctes : si les modifieurs

[1] Voir la figure (5.6) du *Chapitre 5*.

sont restrictifs, le nom est probablement commun ; s'ils sont non restrictifs, le nom peut être commun ou propre.

8.2.6. Expressions référentielles dans le cadre phrastique

Les expressions référentielles de chaque fonction syntaxique ne jouent pas le même rôle pour la continuité référentielle et pour la progression textuelle. Leur pourcentage est récapitulé dans le tableau (8.10).

Le sujet est la fonction syntaxique la plus essentielle où se localisent les expressions référentielles, cela étant le cas de 36 % d'occurrences en français et de 43 % en chinois dans le corpus. Aucune autre fonction syntaxique ne dépasse ces proportions. C'est la seule fonction syntaxique pour laquelle la proportion augmente en fonction de la saillance du référent : plus saillant est le référent, plus souvent il apparaît à la place de sujet de phrase. Cette tendance est reflétée par l'augmentation de la proportion des maillons du protagoniste (56 % en français et 65 % en chinois) par rapport à celle des maillons (53 % en français et 59 % en chinois) et des expressions référentielles (36 % en français et 43 % en chinois). Le premier maillon des chaînes est majoritairement à la place de sujet. Le sujet récurrent d'un même référent permet la continuité référentielle dans le roman-fleuve. Cette tendance est plus remarquable en chinois qu'en français, puisque les sujets sont, dans tous les cas, proportionnellement plus significatifs en chinois qu'en français.

Les compléments d'objet et les circonstants affichent une tendance similaire et décroissante suivant la saillance référentielle dans les deux langues. Cette tendance s'oppose à celle des sujets : les maillons du protagoniste apparaissent moins souvent aux places de complément d'objet et de circonstant par rapport aux constituants de toutes les chaînes, qui apparaissent encore moins souvent par rapport aux expressions référentielles en général dans le corpus. Ainsi, les expressions situées à ces deux fonctions syntaxiques tendent à référer à des entités moins saillantes que celles de sujet.

Les compléments du nom présentent des proportions relativement stables et indépendantes de l'influence de la saillance du référent en français et en chinois. Quel que soit le protagoniste ou le référent non humain, quel que soit le maillon d'une chaîne globale ou le singleton, les référents peuvent être mentionnés à la place de circonstant sans différence.

Les attributs concernent les plus petites proportions, avec seulement 2 % en français et 3 % en chinois, avec une légère tendance à la baisse quand la saillance référentielle augmente.

Tableau (8.10) : Fonctions syntaxiques des expressions référentielles en français et en chinois

Pourcentage de	ER		Maillons		JC	
Langue	Fr	Ch	Fr	Ch	Fr	Ch
Sujet	36 %	43 %	53 %	59 %	56 %	65 %
Complément d'objet	24 %	28 %	19 %	21 %	19 %	20 %
Complément du nom	15 %	14 %	18 %	15 %	19 %	14 %
Circonstant	19 %	12 %	7 %	2 %	5 %	0,8 %
Attribut	2 %	3 %	1 %	2 %	0,3 %	0,3 %
Autres	4 %	2 %	3 %	1 %	2 %	0,9 %

8.2.7. Chaînes de référence dans les séquences narratives et dialogales dans le cadre d'un roman-fleuve

De manière générale, le roman-fleuve est un genre narratif. De fait, *Jean-Christophe* se compose principalement de séquences narratives, avec 8 % de séquences dialogales en français et 10 % en chinois dans le corpus.

Les séquences narratives proprement dites sont dominées par les chaînes des humains, et plus précisément par celles des personnages spécifiques et inventés du roman-fleuve. Le déroulement des événements est rythmé tout d'abord par les marqueurs de changement spatio-temporel, et ensuite par les formes nominales des référents.

Dans les séquences dialogales, nous n'avons observé aucune chaîne de notion spatio-temporelle, ni aucune expression déictique de ce type de référent : ces repères spatio-temporels sont mentionnés comme informations contextuelles des séquences dialogales dans les séquences narratives précédente et suivante, au lieu de l'être par les interlocuteurs des séquences dialogales. Les seules expressions déictiques présentes sont les pronoms personnels de première et deuxième personnes : l'emploi de ce groupe d'expressions est identique en français et en chinois. Ces pronoms déictiques sont très récurrents dans les séquences dialogales et très peu utilisés dans les narratives. Par ailleurs, les SN sans déterminant français sont réservés particulièrement pour les appellatifs dans les séquences dialogales, ce qui n'est pas du tout le cas dans les narratives.

8.2.8. Modes de cohabitation des chaînes de référence dans les épisodes multi-référentiels

Notre corpus est composé de 50 épisodes extraits du roman-fleuve *Jean-Christophe*, dont 4 sont mono-référentiels, avec la seule chaîne du protagoniste comme dominante sans concurrent, et 46 multi-référentiels. Nous cherchons ainsi à savoir quels sont les modes de cohabitation des chaînes concurrentes avec celle du protagoniste dans ces épisodes multi-référentiels en français et en chinois. Les modes proposés par Schnedecker (1992, 1997, 2019, 2021) [2] sont la *succession*, l'*entrecroisement*, la *dérivation*, la *partition / fusion* et le *déroulement parallèle*. Ces modes sont appropriés pour décrire la cohabitation des chaînes en français et en chinois. Par exemple, les extraits parallèles [11] et [12] contiennent une cohabitation entre la chaîne du protagoniste et celle de « Jean-Michel » en mode de déroulement parallèle[3] :

[11][**Grand-père**] toussotait. [**Christophe**] savait bien ce que cela voulait dire. [**Le vieux**] brûlait
réf1 *réf-2* *réf-1*

[2] Voir aussi les définitions des modes dans la section *3.2.2.1. Chaînes de référence dans divers genres en français* du *Chapitre 3*.

[3] Et une cohabitation potentielle est en mode de fusion au moment de l'expression « ils » (« réf-3 ») à la condition que ces deux personnes soient reprises comme une seule entité dans la suite du texte.

d'envie de raconter une histoire ; mais **[il]** voulait que **[l'enfant]** la **[lui]** demandât. **[Christophe]** n'y
réf-1 réf-2 réf-1 réf-2

manquait pas. **[Ils]** s'entendaient ensemble. **[Le vieux]** avait une immense affection pour **[[son]**
réf-3 réf-1 réf-1

petit-fils] ; et ce **[lui]** était une joie de trouver en **[lui]** **[un public complaisant]**.
réf-2 réf-1 réf-2 réf-2

[12]**[祖父]** 咳 了 几 声。**[克利斯朵夫]** 很 明白 这 个 意思。**[老人]** 极想
zufu *ke* *le* *ji* *sheng* *Kelisiduofu* *hen* *mingbai* *zhe* *ge* *yisi* *laoren* *jixiang*
grand-père tousser LE quelques coup Christophe bien comprendre ce CL sens vieux-homme désirer
réf-1 réf-2 réf-1

讲 故事，但 **[Ø]** 要 **[孩子]** 向 **[他]** 请求。**[克利斯朵夫]** 立刻 凑上去。
jiang *gushi* *dan* *yao* *haizi* *xiang* *ta* *qingqiu* *Kelisiduofu* *like* *coushangqu*
raconter histoire mais devoir enfant vers 3P demander Christophe immédiatement se-rapprocher
réf-1 réf-2 réf-2 réf-2

[他们俩] 很 投机。**[老人]** 非常 喜欢 **[孙子]**；有 **[个 愿意 听 [他]** 说话 的
tamen-lia *hen* *touji* *laoren* *feichang* *xihuan* *sunzi* *you* *ge* *yuanyi* *ting* *ta* *shuohua* *de*
3P-deux bien s'entendre vieux-homme bien aimer petit-fils avoir CL vouloir écouter 3P parler DE
réf-3 réf-1 réf-2 réf-2 réf-1

人] 更 使 **[他]** 快乐。
ren *geng* *shi* *ta* *kuaile*
personne plus rendre 3P joyeux
réf-1

Pourtant, d'après la répartition des chaînes à l'intérieur de chaque épisode, nous avons constaté de très grandes différences en ce qui concerne la longueur des chaînes qui cohabitent. Le mode de cohabitation montré dans l'extrait [11] ne peut pas toujours concerner des paragraphes et des épisodes, et aucune des deux chaînes ne cohabitent dans l'intégralité d'un épisode multi-référentiel. Par exemple, l'épisode 17 comporte 3 chaînes : celle de « Jean-Christophe » contient 54 maillons, alors que celles de « Melchior » et de « Melchior & Louisa » en comportent respectivement 3 et 4 (voir le tableau (6.2) du *Chapitre 6*). Dans quelques paragraphes, les trois chaînes cohabitent, mais elles ne le font pas dans tout l'épisode. Étant donné cette différence de longueur des chaînes co-présentes dans un même épisode, la notion de *cohabitation*, utilisée ici par nous, semble flexible et appropriée. En effet, lorsqu'un épisode comporte quelques extraits avec des chaînes qui cohabitent, il est considéré comme multi-référentiel, dans lesquels le référent dominant possède un ou plusieurs référents concurrents.

L'observation du corpus révèle que tous les modes ne sont pas aussi présents les uns que les autres : l'entrecroisement (extraits [13] et [14]) et la succession (extraits [15] et [16]) sont plus récurrents que l'alternance, le parallélisme et la dérivation, ce lorsque l'extrait comporte deux chaînes :

[13]Ce jour-là, **[[sa] mère]** **[lui]** avait mis **[ses]** habits les plus propres, de vieux habits donnés, dont
réf-1 réf-2 réf-1 réf-1

l'ingénieuse patience de **[Louisa]** avait su tirer parti. **[Il]** alla **[la]** rejoindre, comme **[elle]** le **[lui]**
réf-2 réf-1 réf-2 réf-2 réf-1

avait dit, dans la maison où **[elle]** travaillait. **[Il]** était intimidé, à l'idée d'entrer seul.
réf-2 réf-1

[14]那 天，**[母亲]** 替 **[他]** 穿 了 最 干净的 衣服，那 是 人家 布施的 旧 衣衫，由
na *tian* *muqin* *ti* *ta* *chuan* *le* *zui* *ganjingde* *yifu* *na* *shi* *renjia* *bushide* *jiu* *yishan* *you*
ce-là jour mère pour 3P habiller LE le-plus propre vêtement cela être on donnant vieux vêtement par
réf-1 réf-2

[鲁意莎] 很 巧妙 很 耐性的 改过了的。**[Ø]** 依着 **[她的]** 吩咐，**[他]** 到 **[她]** 工作 的 人家
Luyisha *hen* *qiaomiao* *hen* *naixingde* *gaiguolede* *yizhe* *tade* *fenfu* *ta* *dao* *ta* *gouzuo* *de* *renjia*
Louisa très ingénieux très patient modifié selon son ordre 3P à 3P travailler DE famille
réf-1 réf-2 réf-1 réf-2 réf-1

去 接 **[她]**。**[他]** 一 想 要 自个儿 进去，**[Ø]** 不免 有点儿 胆小。
qu *jie* *ta* *ta* *yi* *xiang* *yao* *zigeer* *jinqu* *bumian* *youdianer* *danxiao*
aller chercher 3P 3P dès penser devoir seul entrer inévitablement un-peu intimidé
réf-1 réf-2 réf-2

[15][**Elle**] arriva, toute bouleversée encore. [**Elle**] commença par de violents reproches, mêlés de
réf-1 réf-1

nouvelles gifles, auxquelles [**Melchior**] joignit, sitôt qu'[**il**] eut compris, – et probablement avant, –
réf-2 réf-2

des claques à assommer un bœuf.

[16]临了 [**鲁意莎**] 也 急急忙忙 上楼 了。 [**她**] 还 象 刚才 一样的 神魂不定， [**Ø**]
linliao Luyisha ye jijimangmang shanglou le ta hai xiang gangcai yiyangde shenhunbuding
près Louisa aussi précipité monter LE 3P encore CMP tout-à-l'heure CMP bouleversé
réf-1 réf-1 réf-1

一 进来 就 大骂， 又 加上 几 个 嘴巴， [**曼希沃**] 听明白 了， [**Ø**] 也 帮着
yi jinlai jiu dama you jiashang ji ge zuiba Manxiwo tingmingbai le ye bangzhe
dès entrer donc insulter aussi ajouter quelques CL gifle Melchior comprendre LE aussi aidant
réf-2 réf-2

揍 他，（或许 [**Ø**] 没有 明白 之前 已经 动手 了），那 股 狠劲 差不多
zou ta huoxu meiyou mingbai zhiqian yijing dongshou le na gu henjin chabuduo
taper 3P probablement NEG comprendre avant déjà commencer LE ce-là CL violence presque
réf-2

可以 打死 一 条 牛。
keyi dasi yi tiao niu
pouvoir assommer un CL bœuf

Comme nous l'avons vu dans la section *6.3. Chaînes des personnages épisodiques et éphémères* du *Chapitre 6*, la fusion et la partition ont régulièrement trait aux chaînes des humains épisodiques et éphémères : deux personnages constituent respectivement deux chaînes et sont ensuite fusionnés en une autre de deux personnages. Ceux-ci peuvent être le protagoniste, les personnages secondaires, les personnages spécifiques épisodiques et éphémères ; mais ce mode n'est jamais observé dans les chaînes des référents non humains. Comme le montrent les extraits [17] et [18], « ils » (« réf-3 ») est une fusion de « grand-père » (« réf-1 ») et « le petit » (« réf-2 ») :

[17][**Grand-père**] [**le**] prenait souvent avec [**lui**], dans [**ses**] promenades du soir. [**Le petit**] trottinait
réf-1 réf-2 réf-1 réf-1 réf-2

à [**ses**] côtés, en [**lui**] donnant la main. [**Ils**] allaient par les chemins, au travers des champs
réf-1 réf-1 réf-3

labourés, qui sentaient bon et fort. Les grillons crépitaient. Des corneilles énormes, posées de profil en travers de la route, [**les**] regardaient venir de loin et s'envolaient lourdement à leur approche.
réf-3

[18][**祖父**] 在 黄昏 散步 的 时候 常常 带着 [**他**] 一块儿 去。 [**孩子**] 拉着
zufu zai huanghun sanbu de shihou changchang daizhe ta yikuaier qu haizi lazhe
grand-père PREP soir promenade DE temps souvent apportant 3P ensemble aller enfant tenant
réf-1 réf-2 réf-2

[**老人**] 的 手 在 旁边 急急忙忙的 搬着 小 步。 [**他们**] 走着 乡下 的 路，
laoren de shou zai pangbian jijimangmangde banzhe xiao bu tamen zouzhe xiangxia de lu
vieux-homme DE main PREP côté précipité trottinant petit pas 3Ps allant campagne DE chemin
réf-1 réf-3

[**Ø**] 穿过 锄松的 田， [**Ø**] 闻到 又 香 又 浓 的 味道。 蟋蟀 叫着。 很大的 乌鸦
chuanguo chusongde tian wendao you xiang you nong de weidao xishuai jiaozhe hendade wuya
traverser labouré champ sentir et bon et fort DE odeur grillon crépitant énorme corneille
réf-3 réf-3

斜 蹲 在 路 上 远远的 望着 [**他们**]， [**他们**] 一 走近， 就 笨重的
xie dun zai lu shang yuanyuande wangzhe tamen tamen yi zoujin jiu benzhongde
en-travers poser PREP route dessus loin regardant 3Ps 3Ps dès approcher donc lourdement
réf-3 réf-3

飞走 了。
feizou le
s'envoler LE

Outre les situations d'un seul mode de cohabitation, la plupart des extraits multi-référentiels de trois chaînes ou plus présentent un mode hybride qui combine plusieurs modes basiques de cohabitation. Par exemple, les extraits parallèles [19] et [20] contiennent plusieurs

chaînes [4] qui cohabitent les unes avec les autres selon différents modes : la chaîne « Christophe » (« réf-1 ») et « ses voisins » (« réf-2 ») cohabitent en mode de dérivation ; la chaîne « Christophe » (« réf-1 ») et la chaîne « le maître » (« réf-3 ») s'entrecroisent ; la chaîne « certains » (« réf-5 ») est une partition de la chaîne « tous les enfants » (« réf-4 »).

[19]À quelque temps de là, à l'école, où [**Christophe**]$_{réf-1}$ passait [**son**]$_{réf-1}$ temps à regarder les mouches au plafond et à donner des coups de poing à [[**ses**]$_{réf-1}$ **voisins**]$_{réf-2}$, pour [**les**]$_{réf-2}$ faire tomber du banc, [**le maître qui [l'] avait pris en grippe**]$_{réf-3}$, parce qu'[**il**]$_{réf-1}$ remuait toujours, parce qu'on [**l'**]$_{réf-1}$entendait toujours rire, et parce qu'[**il**]$_{réf-1}$ n'apprenait jamais rien, fit une allusion inconvenante, un jour que [**Christophe**]$_{réf-1}$ s'était [**lui-même**]$_{réf-1}$ laissé choir, à certain personnage bien connu dont [**il**]$_{réf-1}$ semblait vouloir suivre brillamment les traces. [**Tous les enfants**]$_{réf-4}$ éclatèrent de rire ; et [**certains**]$_{réf-5}$ se chargèrent de préciser l'allusion, en des commentaires aussi clairs qu'énergiques.

[20]过了 一 晌， [克利斯朵夫]$_{réf-1}$ 上学 了； [他]$_{réf-1}$ 老 望着 天花板 上 的 苍蝇， [Ø] 把
guole yi shang Kelisiduofu shangxue le ta lao wangzhe tianhuaban shang de cangying ba
après un temps Christophe aller-école LE 3P toujours regardant plafond dessus DE mouche BA

拳头 捶着 [旁边 的 孩子]$_{réf-2}$， [Ø]$_{réf-1}$ 推 在 地 下； [他]$_{réf-1}$ 动个 不停， [Ø] 笑个
quantou chuizhe pangbian de haizi tui zai di xia ta dongge buting xiaoge
poing tapant voisin DE enfant pousser PREP terre dessous 3P remuant sans-cesse riant

不停， [Ø]$_{réf-1}$ 从来不 念书。有 一 天， [克利斯朵夫]$_{réf-1}$ 自己 摔 在 了 地 下，
buting conglaibu nianshu you yi tian Kelisiduofu ziji shuai zai le di xia
sans-cesse ne-jamais lire avoir un jour Christophe soi-même tomber PREP LE sol dessous

[讨厌 [他]$_{réf-1}$ 的 老师]$_{réf-3}$ 便 说 了 句 难听的 话 隐射 某 个 大家 知道 的
taoyan ta de laoshi bian shuo le ju nantingde hua yinshe mou ge dajia zhidao de
détester 3P DE maître donc dire LE CL inconvenant mot faire-allusion certain CL tout-le-monde connaître DE

人， [Ø]$_{réf-3}$ 说 [他]$_{réf-1}$ 大概 要 青出于蓝的 走上 那 条 路 了。[所有的 孩子]$_{réf-4}$ 听着
ren shuo ta dagai yao qingchuyulande zoushang na tiao lu le suoyoude haizi tingzhe
personne dire 3P peut-être aller brillamment suivant ce-là CL trace LE tout enfant entendant

都 哈哈大笑； [有些 同学]$_{réf-5}$ 还 揭穿 隐喻， [Ø]$_{réf-5}$ 加上 一些 又 明白 又
dou hahadaxiao youxie tongxue hai jiechuan yinyu jiashang yixie you mingbai you
tout éclater-rire certain camarade même révéler allusion précisant certains et clair et

有分量的 注解。
youfenliangde zhujie
énergique commentaire

Les nombreux exemples des chaînes issues de notre corpus manifestent la complexité des chaînes de référence en français et en chinois. Ce n'est pas un seul référent topical qui permet le développement du texte, mais également les très nombreuses chaînes d'autres référents qui assurent chacune la continuité référentielle et qui établissent ensemble la progression du roman-fleuve.

[4] Il s'agit de quelques vraies chaînes de plus de trois maillons ou du début de quelques chaînes qui présentent moins de trois maillons dans cet extrait ; mais les reprises ont lieu plus loin dans la suite du roman-fleuve.

Conclusion

Notre thèse a consisté à étudier la notion linguistique dite *chaîne de référence*. Cette notion concerne des phénomènes complexes, multiples et variables, dans différents genres discursifs et dans différentes langues. Comment les chaînes de référence se réalisent-elles dans un roman-fleuve en français ? Quelles sont les variations de ces chaînes dans les textes parallèles en chinois ? Ce sont à ces deux questions primordiales que nous avons tenté de répondre dans le présent travail.

- Trois axes d'exploration des chaînes de référence

Celui-ci se divise en trois parties ayant pour objectif de développer les différents aspects relatifs aux chaînes de référence selon trois axes.

Le premier axe est celui des définitions et des théories.

Les études comparatives franco-chinoises sur les chaînes de référence et les phénomènes associés sont moins nombreuses que celles qui ont trait aux aspects comparatifs anglo-chinois et franco-anglais. Dans beaucoup de parties de notre thèse, nous avons effectué un détour en nous référant à des théories et à des exemples en anglais pour ensuite formuler des questions et des hypothèses, et proposer des pistes de recherche dans le cadre de notre approche franco-chinoise. Parmi celles-ci figurent la définition de *chaîne anaphorique* de Chastain, la notion de *chaîne topicale* de Givón, l'*Échelle d'accessibilité* d'Ariel, etc., qui sont basées sur l'observation des faits en anglais. Ces études nous ont permis de développer notre réflexion, dans le sens où nous avons cherché à les adapter au français et au chinois. Cependant, d'autres travaux, consacrés aux chaînes de référence et à leurs notions connexes en français, ont été encore plus cruciaux pour nous ; il s'agit surtout de ceux de Corblin, de Kleiber et de Schnedecker, qui proposent des théories et des définitions fondamentales que nous avons pu adapter à notre étude comparative.

Il existe déjà de nombreuses recherches relatives aux chaînes de référence pour de nombreuses langues. Outre l'anglais et le français, on en trouve aussi pour l'allemand (Baumer *et al.*, 2021), l'italien et le danois (Skytte & Korzen, 2000, cité par Schnedecker & Landragin, 2014 : 11 ; Lundquist, 2005), le coréen (Kim, 2023) ainsi que notre article pour le chinois (Guo, 2021) dans les textes non narratifs. Ce sont ces études antérieures dans diverses langues qui nous ont permis de définir notre perspective de recherche et qui nous ont conduit à proposer notre propre définition de la notion de chaîne de référence pour le chinois. Nous nous avons tâché de mettre en exergue les spécificités de cette notion en français et en chinois dans des textes parallèles appartenant au genre du roman-fleuve.

Le deuxième axe est méthodologique.

En tant que notion linguistique complexe avec de nombreux composants très variés, les chaînes de référence ne sont pas faciles à détecter et à localiser en grande quantité ou dans un long texte. Il a donc fallu nous interroger sur la manière de repérer ces entités linguistiques dans un grand corpus, et par conséquent résoudre des problèmes d'ordre méthodologique.

Pour constituer notre propre corpus de thèse, nous avons choisi le roman-fleuve *Jean-Christophe* de Romain Rolland et sa traduction en chinois par Lei Fu pour leur notoriété dans les domaines de la littérature et de la traduction et pour leur statut représentatif du genre de roman-fleuve. Le langage littéraire et narratif comporte des chaînes de référence fort différentes de celles des autres genres, narratifs et non narratifs ; mais les études antérieures n'ont, à notre connaissance, pas abordé la question des chaînes de référence dans des textes aussi longs. Les extraits de ce roman-fleuve nous ont permis de constituer des textes parallèles, sur lesquels nous avons réalisé notre annotation des chaînes de référence et effectué notre observation des faits.

Pour que l'analyse de ces textes longs soit plus efficace et ses résultats plus riches, nous avons choisi un traitement outillé pour repérer les chaînes de référence dans les deux langues. Nous avons tout d'abord comparé les résultats de l'annotation automatique dans AiHanYu pour le chinois, en synthétisant quelques types d'erreurs récurrentes. Il nous a paru indispensable de faire une typologie des erreurs systématiques avant d'entamer les annotations et les rectifications manuelles pour que nous ne prêtions plus attention à ces mêmes types d'erreurs produites dans les annotations automatiques. Après les tests, nous avons annoté manuellement toutes les expressions référentielles et toutes les chaînes de référence en opérant de nombreuses rectifications.

Pendant ce travail d'annotation et de rectification, nous n'avons pas repris directement les guides d'annotation déjà utilisés pour les autres projets ou pour annoter les corpus monolingues. Nous avons élaboré notre propre guide des conventions d'annotation en l'adaptant particulièrement à la récupération et à la caractérisation des expressions référentielles et des chaînes de référence dans un corpus bilingue franco-chinois. Tous nos choix ont été justifiés par des définitions de dictionnaires, par des théories solides de linguistes et par l'illustration d'exemples. Nous avons cherché à mettre en exergue des arguments linguistiques et informatiques pertinents dans le but de rendre notre méthode de constitution de corpus plus fiable et reproductible, de sorte que les études ultérieures puissent s'appuyer sur la nôtre. Quelques théories cognitives nous ont également servi à fixer des étiquettes d'annotation, comme l'*Échelle d'accessibilité référentielle* et la *Hiérarchie du donné*.

Le troisième axe, enfin, porte sur l'observation des faits.

À partir des analyses quantitatives de la totalité du corpus, nous avons saisi les caractéristiques générales et spécifiques des chaînes de référence des deux langues. L'observation des emplois majoritaires et minoritaires a permis de découvrir les tendances et les possibilités des composants de différents types de chaînes de référence. Les emplois prototypiques et les emplois exceptionnels des expressions référentielles ont été illustrés par des extraits du corpus, qui montrent chacun leur valeur et leur rôle dans la continuité référentielle et la progression du discours. Nous avons constaté, par une approche comparative franco-chinoise, que plusieurs facteurs déterminent ensemble la référence de chaque catégorie d'expression. Le critère formel est sans aucun doute un des facteurs décisifs, mais les autres sont loin d'être négligeables, comme le nombre de référents concurrents, la distance inter-maillonnaire, la fonction syntaxique, le type de séquence textuelle, etc.

Dans notre corpus constitué par des extraits parallèles d'un roman-fleuve, les différences majeures des chaînes de référence sont dues aux impacts des deux langues. De nombreux points sont similaires lorsqu'il s'agit de la structure générale de l'œuvre, comme les types de séquences textuelles, la répartition des chaînes de référence, les types de référents,

etc. ; en revanche, les différences sont significatives lorsqu'il est question des formes et des structures spécifiques, comme notamment celles des déterminants, des formes non nominales, etc.

- Applications potentielles des résultats dans la pratique

Nous souhaitons que les apports de notre thèse ne se limitent pas seulement au niveau théorique, mais qu'ils puissent s'appliquer dans des pratiques comme l'enseignement des langues, la traduction et le traitement automatique des textes.

Pour ce qui est de l'enseignement, il est important de faire prendre conscience aux apprenants de la question de la cohésion discursive lors de la production écrite, que ce soit dans leur langue maternelle ou dans une langue étrangère. Cela ne veut pas dire qu'il faut leur présenter la définition des notions linguistiques ; il est plus pédagogique de leur proposer des exemples et de leur montrer les régularités inhérentes à des langues, pour qu'ils puissent ensuite rendre leurs écrits plus cohésifs, surtout s'il est question d'écrire des textes plus ou moins longs.

En ce qui concerne la traduction, nous avons constaté que celle de Lei Fu n'est pas une transcription littérale de chaque expression ni un transfert des structures syntaxiques identiques du français vers le chinois. Par la comparaison des chaînes de référence entre les deux langues, ni les noms, ni les déterminants, ni les formes non nominales ne peuvent être systématiquement et directement traduits de manière complètement identique d'une langue à l'autre. L'intérêt est de savoir choisir à quel endroit dans l'œuvre et de quelle manière la traduction est le plus pertinente, pour que les textes se présentent comme une intégralité cohésive dans la langue ciblée, au lieu d'être une succession d'expressions isolées ou une juxtaposition d'éléments décousus.

Les exemples typiques et les spécificités des chaînes de référence obtenues dans notre corpus peuvent également servir à améliorer l'annotation automatique des chaînes de référence, la traduction automatique des expressions référentielles et l'identification des référents topicaux dans les discours bilingues.

- Limites et perspectives

En dehors des apports souhaitables, nous sommes consciente des nombreuses limites que comporte notre travail, qui demande à être amélioré dans la perspective de nos prochaines études.

Nous avons en effet rencontré des difficultés à chaque étape de la constitution du corpus.

La première d'entre elles n'est autre que la taille de ce corpus. D'un côté, vu la taille limitée des extraits du roman-fleuve que nous avons pris pour corpus, il est impossible d'inventorier et d'observer tous les phénomènes de cette œuvre ; les quatre premières sections de l'œuvre risquent de ne pas présenter les mêmes caractéristiques que celles de son intégralité. D'un autre côté, nous avons présenté certains résultats chiffrés qui reposent sur des échantillons du corpus puisque certains groupes d'expressions sont très nombreux et que nous n'avons pas eu l'occasion de les analyser avec toutes les occurrences. Cette opération d'échantillonnage n'a malheureusement pas démontré l'avantage de notre corpus en assez

grande quantité. Il est donc souhaitable de continuer à explorer les phénomènes par des annotations plus exhaustives et plus détaillées pour obtenir des chiffres issus de toutes les occurrences et de l'intégralité du corpus.

Pendant l'élaboration du guide des conventions d'annotation, certaines manières d'annoter et d'étiqueter les expressions référentielles ont pu donner lieu à des compromis éventuels pour des raisons linguistiques et opérationnelles. Ces compromis sont inévitables lorsque l'on sort des cadres théoriques idéalisés. Nous avons régulièrement constaté des usages exceptionnels qui ne s'accordent pas parfaitement aux critères prédéfinis dans un guide des conventions. Ces compromis ont été rencontrés lors de l'annotation d'un autre annotateur et lors de la revue de l'annotation par nous-même. D'ailleurs, cela a aussi été le cas pendant l'annotation du corpus DEMOCRAT : lorsqu'un texte est annoté par plusieurs personnes, aucune d'entre elles n'annote de manière identique. Des désaccords sont donc permanents entre les annotateurs et entre la première annotation et les revues réalisées par le même annotateur. Malgré tout cela, nos conventions définitives (après de nombreuses modifications et rectifications) ont été respectées au maximum pendant notre première annotation, pendant les nombreuses rectifications selon les conventions et pendant l'évaluation de la deuxième annotation d'un membre anonyme du projet DEMOCRAT, ce en dépit du fait que les annotations obtenues n'étaient jamais identiques. Les outils de traitement linguistiques se renouvellent jour après jour ; ceux réalisés et accessibles au commencement de notre thèse étaient moins pertinents, ayant moins de fonctions et générant plus d'erreurs systématiques que les outils réalisés par la suite. Il est par conséquent indispensable pour nous d'utiliser des outils plus récents afin d'améliorer notre méthode dans de futurs travaux.

Les questions méthodologiques nous ont paru primordiales pour notre étude, puisque différentes méthodes ont permis d'obtenir différents résultats et de nous diriger vers différentes conclusions par l'observation du même corpus. Pour cette raison, nos réflexions sur le présent travail portent non seulement sur les analyses des faits génériques ou individuels, mais également sur la recherche d'une méthode plus acceptable, plus adéquate et plus scientifique pour analyser les faits du langage.

Bibliographie

Références en français et en anglais

ABBOTT, Barbara. Definite and indefinite. *Encyclopedia of language and linguistics*, 2006, vol. 3, no 392, p. 99-109.

ABEILLÉ, Anne & GODARD, Danièle. *Grande Grammaire du français*. Arles : Actes Sud. 2021.

ACHARD-BAYLE, Guy. Référents évolutifs et point de vue. *Linx*, 2000, 43, p. 1-12.

ADAM, Jean-Michel. Types de séquences élémentaires. *Pratiques : linguistique, littérature, didactique*, 1987, 56, p. 54-79.

ADAM, Jean-Michel. *Le texte narratif – Traité d'analyse pragmatique et textuelle*. Paris : Nathan. 1994.

ADAM, Jean-Michel. Genres, textes, discours : pour une reconception linguistique du concept de genre. *Revue belge de philologie et d'histoire*, 1997, 75, p. 665-681.

ADAM, Jean-Michel. Types de textes ou genres de discours ? Comment classer les textes qui disent de et comment faire ? *Langages*, 2001, 141, p. 10-27.

ADAM, Jean-Michel. Texte, contexte et discours et questions. *Pratiques*, 2006, 129-130, p. 21-34.

ADAM, Jean-Michel. *La linguistique textuelle*. Paris : Armand Colin. 2015 [2011].

ADAM, Jean-Michel. *Les textes : types et prototypes*. Paris : Armand Colin. 2017.

ADAM, Jean-Michel. La notion de texte. *Encyclopédie Grammaticale du Français* (http://encyclogram.fr), 2019.

ADAM, Jean-Michel & HEIDMANN, Ute. *Le texte littéraire : pour une approche interdisciplinaire*. Louvain-la-Neuve [Belgique] : Academia Bruylant. 2009.

ADAM, Jean-Michel & VIPREY, Jean-Marie. Corpus de textes, textes en corpus. Problématique et présentation. *Corpus*, 2009, 8, p. 5-25.

AiHanYu : http://www.aihanyu.org/cncorpus/CpsWParser.aspx. Dernière consultation : novembre 2018.

ANALEC : http://www.lattice.cnrs.fr/ressources/logiciels/analec/, http://explorationdecorpus.corpusecrits.huma-num.fr/analec-2/. Dernière consultation : août 2019.

ANSCOMBRE, Jean-Claude. L'article zéro en français : un imparfait du substantif ? *Langue française*, 1986, 72, p. 4-39.

ARIEL, Mira. *Referring and accessibility*. Journal of linguistics, 1988, 24, p. 65-87.

ARIEL, Mira. *Accessing noun-phrase antecedents*. Londres [Royaume-Uni] : Routledge. 1990.

ARIEL, Mira. Centering, Accessibility and the Next Mention. *Theoretical linguistics*, 2013, 39, p. 39-58.

BAUMER, Emmanuel. *Noms propres et anaphores nominales en anglais et en français : étude comparée des chaînes de référence*. Thèse doctorale en Linguistique anglaise, Université Paris Diderot (Paris 7), 2012 (a).

BAUMER, Emmanuel. Étude contrastive français / anglais des anaphores lexicales dans la presse et dans la fiction littéraire ». *Corela*, 2012 (b), 11.

BAUMER, Emmanuel. *Noms propres et anaphores nominales en anglais et en français : étude comparée des chaînes de référence*. Paris : Harmattan. 2015.

BAUMER, Emmanuel. Chaînes de référence et point de vue dans la fiction littéraire : le cas des nouvelles courtes. *Langue française*, 2017, 195, p. 73-90.

BAUMER, Emmanuel, DIAS, Dominique & SCHNEDECKER, Catherine. Les chaînes de référence dans un corpus contrastif (allemand – anglais – français) de romans réalistes du XIXème siècle : analyse quantitative et qualitative. 2020, p. 211-229.

BAUMER, Emmanuel, DIAS, Dominique, GARDELLE, Laure & PRAK-DERRINGTON, Emmanuelle. Quelles régularités pour les chaînes de référence dans le genre « nouvelles d'actualité » ? Exploration contrastive allemand / anglais / français. *Travaux de Linguistique : Approches contrastives des chaînes de référence*, 2021, 82, p. 17-44.

BOUCHEIX, Jean-Michel & FAYOL, Michel. Initiation, maintien et changement de références : le cas des recettes. Dans : KLEIBER, Georges *et al.* (Éds), *La continuité référentielle*. Paris : Librairie Klincksieck, 1997, p. 53-70.

BOUDREAU, Sylvie & KITTREDGE, Richard. Résolution d'anaphores et identification des chaînes de coréférence : une approche 'minimaliste'. *8ème Journées internationales d'Analyse statistique des Données Textuelles (JADT 2006)*, Besançon, 2006.

BRANCY, Jean-Yves. Romain Rolland, Stefen Zweig et l'Europe (1919-1933). *Cahier de Brèves, Chancellerie des Universités de Paris, Etudes Romain Rolland*, 2005, p. 8-13.

BRANCY, Jean-Yves. S'engager pour la paix en 1914 : l'éclairage de Romain Rolland. *Congrès du Mouvement de la Paix*, Toulouse, 2014.

CAPIN, Daniéla. Chaînes de référence dans les textes médiévaux non-narratifs : les year books ou l'élaboration d'une écriture juridique. *Langages*, 2014, 195, p. 61-78.

CAPIN, Daniéla, GLIKMAN, Julie, SCHNEDECKER, Catherine & TODIRASCU, Amalia. Le rôle des chaînes de référence dans la structuration textuelle : étude diachronique de l'ancien français au français moderne. *Langages*, 2021, 224-4, p. 87-107.

CHARAUDEAU, Patrick & MAINGUENEAU, Dominique. *Dictionnaire d'analyse du discours*. Paris : Seuil. 2002.

CHAROLLES, Michel. Contraintes pesant sur la configuration des chaînes de référence comportant un nom propre. *Travaux du Centre de Recherches Sémiologiques*, 1987, 53, p. 29-55.

CHAROLLES, Michel. Les plans d'organisation textuelle : périodes, chaînes, portées et séquences. *Pratiques : linguistique, littérature, didactique*, 1988, 57, p. 3-13.

CHAROLLES, Michel. Cohésion, cohérence et pertinence du discours. *Travaux de Linguistique*, 1995, 29, p. 125-151.

CHAROLLES, Michel. Identité, changement et référence pronominale. Dans : KLEIBER, Georges *et al.* (Éds), *La continuité référentielle*. Paris : Librairie Klincksieck, 1997, p. 71-96.

CHAROLLES, Michel. Référents évolutifs et évolution de la référence. Dans : DE MULDER, Walter & SCHNEDECKER, Catherine (Éds), *Les référents évolutifs entre linguistique et philosophie*. Paris : Klincksieck, 2001, p. 39-97.

CHAROLLES, Michel. *La référence et les expressions référentielles en français*. Paris : Éditions Ophrys. 2002.

CHAROLLES, Michel. Comment évaluer les effets des relatives en « qui » sur les chaînes de référence ? Dans : CHAROLLES, Michel et al. (Éds), *Parcours de la phrase - Mélanges offerts à Pierre Le Goffic*. Paris : Éditions Ophrys, 2007, p. 193-212.

CHAROLLES, Michel & SCHNEDECKER, Catherine. Coréférence et identité : le problème des référents évolutifs. *Langages*, 1993, 27, p. 106-126.

CHASTAIN, Charles. Reference and Context. Dans : GUNDERSON, Keith (Éd.), *Language, Mind and Knowledge*. Minnesota Studies in the Philosophy of Science (volume 7). Minneapolis [États-Unis] : University of Minnesota Press, 1975, p. 194-269.

COMBETTES, Bernard. *Pour une grammaire textuelle*. Bruxelles [Belgique] : De Boeck-Duculot. 1983.

COMBETTES, Bernard. Introduction et reprise des éléments d'un texte. *Pratiques : linguistique, littérature, didactique*, 1986, 49, p. 69-84.

COMBETTES, Bernard. Facteurs discursifs et contraintes syntaxiques : aspects diachroniques de la relation de cataphore. *Langue française*, 2017, 195, p. 111-128.

CORBLIN, Francis. Les chaînes de référence : analyse linguistique et traitement automatique. *Intellectica : Revue de l'Association pour la recherche cognitive*, 1985, 1, p. 123-143.

CORBLIN, Francis. *Indéfini, défini et démonstratif : constructions linguistiques de la référence*. Genève [Suisse] : Librairie Droz S.A. 1987.

CORBLIN, Francis. *Les formes de reprises dans le discours : anaphores et chaînes de référence*. Rennes : Presses Universitaires de Rennes. 1995.

CORBLIN, Francis. Les indéfinis : variables et quantificateurs. *Langue française*, 1997, 116, p. 8-32.

CORBLIN, Francis. Les chaînes de la conversation et les autres. Dans : GOUVARD, Jean-Michel (Éd.), *De la langue au style*. Lyon : Presses universitaires de Lyon, 2005, p. 233-254,

CORBLIN, Francis. *Cours de sémantique : Introduction*. Paris : Armand Colin. 2013.

CORNISH, Francis. Anaphore pragmatique, référence, et modèles du discours. Dans : KLEIBER, Georges & TYVAERT, Jean-Emmanuel (Éds), *L'anaphore et ses domaines*. Paris : Librairie Klincksieck, 1990, p. 81-96.

CORNISH, Francis. Les chaînes topicales : leur rôle dans la gestion et la structuration du discours. *Cahiers de grammaire*, 1998, 23, p. 19-40.

CORNISH, Francis. L'accessibilité cognitive des référents, le centrage d'attention et la structuration du discours : une vue d'ensemble. *Verbum*, 2000, 22, p. 7-30.

CORNISH, Francis. 'Focus of attention' in Discourse. Dans : MACKENZIE, Lachlan *et al.* (Éds), *A New Architecture for Functional Grammar*. Berlin [Allemagne] : Mouton de Gruyter, 2004, p. 117-150.

CORNISH, Francis. English demonstratives: Discourse deixis and anaphora. A discourse-pragmatic account. Dans : Nilsen, Randi Alice, Nana Aba Appiah Amfo & Kaja Borthen (Éds), *Interpreting utterances: Pragmatics and its interfaces – Essays in Honour of Thorstein Fretheim*. Missouri [États-Unis] : Novus Press. 2007, p. 147-166.

CORNISH, Francis. « Strict » Anadeixis, Discourse Deixis and Text Structuring. *Language Sciences*, 2011, 33 (5), p. 753-767.

CORNISH, Francis. Référence indexicale et structuration du discours : incidences du choix de « procédure indexicale », de la / des relation(s) de cohérence invoquée(s), et de l'aspectualité, *Symposuim S'caladis « Sences dessus dessous : niveaux et domaines de réalisation du sens »*. Toulouse, 2013, p. 1-14.

CORNISH, Francis. Procédures référentielles indexicales, relations rhétoriques et structuration du discours. *Discours*, 2015, 17, p. 3-28.

CORNISH, Francis. SN démonstratifs et anadeixis : sens 'spatial', ou valeurs tributaires d'une stratégie pragmatique potentielle ? *Journal of French Language Studies*, 2017, 27, p. 215-239.

Corpus BCC-CCFT (*BLCU Corpus Center*) : http://bcc.blcu.edu.cn/CCFTv2. Dernière consultation : décembre 2020.

Corpus DEMOCRAT : https://www.ortolang.fr/market/corpora/democrat. Dernière consultation : décembre 2019.

CRViewer : http://boberle.com/projects/crviewer/index.html. Dernière consultation : novembre 2018.

CULIOLI, Antoine. A propos de la particule *de* en chinois. *Cahiers de Linguistique Asie Orientale*, 1994, 23, p.77-86.

DAVID, Sophie. Quelques remarques à propos du traitement automatique des relations anaphoriques. Dans : KLEIBER, Georges & TYVAERT, Jean-Emmanuel (Éds), *L'anaphore et ses domaines*. Paris : Librairie Klincksieck, 1990, p. 112-141.

DE CRISTOFARO, Jonathan, STRUBE, Michael & F. MCCOY, Kathleen. Building a Tool for Annotating Reference in Discourse. Dans : *The Relation of Discourse / Dialogue Structure and Reference*. 1999.

DE MULDER, Walter. La création du monde par l'article défini. Le marqueur évidentiel ? *Langue française*, 1994, 102, p. 108-120.

DE MULDER, Walter. Du sens des démonstratifs à la construction d'univers. *Langue française*, 1998, p. 21-32.

DEBRAS, Camille & BEAUPOIL-HOURDEL, Pauline. Gestualité et construction des chaînes de référence dans un corpus d'interactions tandem. *Cahiers de praxématique*, 2019, 72, p. 1-25.

DELABORDE, Marine. La coréférence floue dans les chaînes du corpus DEMOCRAT. *Langages*, 2021, 224, p. 47-65.

DÉSOYER, Adèle, LANDRAGIN, Frédéric, TELLIER, Isabelle, LEFEUVRE, Anaïs, ANTIONE, Jean-Yves & DINARELLI, Marco. Coreference Resolution for French Oral Data: Machine Learning Experiments with ANCOR. *International Conference on Intelligent Text Processing and Computational Linguistics*, Springer, Cham [Suisse], 2016, p. 507-519.

DODDINGTON, George, MITCHELL, Alexis, PRZYBOCKI, Mark, RAMSHAW, Lance, STRASSEL, Stephanie & WEISCHEDEL, Ralph. Automatic Content Extraction (ACE) Program - Task Definitions and Performance Measures. *LREC 2004 : 4th International Conference on Language Resources and Evaluation*, Lisbon [Portugal], 2004, p. 24-30.

DUBOIS, Jean & LAGANE, René. *La nouvelle grammaire du français*. Paris : Librairie Larousse. 1973.

DUBOIS, Jean, GIACOMO, Mathée, GUESPIN, Louis, MARCELLESI, Christiane, MARCELLESI, Jean-Baptiste & MEVEL, Jean-Pierre. *Le Dictionnaire de Linguistique et des Sciences du Langage*. Paris : Librairie Larousse. 2012.

DUCROT, Oswald & TODOROV, Tzvetan. *Dictionnaire encyclopédique des sciences du langage*. Paris : Éditions du Seuil. 1972.

DURANTI, Alessandro. Sociocultural dimensions of discourse. Dans : VAN DIJK, Teun (Éd.), *Handbook of discourse analysis, vol. 1*. Cambridge [États-Unis] : Academic Press, 1985, p. 193-230.

DVORAK, Jan. Le démonstratif et autres marqueurs anaphoriques dans les chaînes de référence en français et en tchèque. Une étude exploratoire du roman *Le brave soldat Chvéïk*. *Travaux de Linguistique : Approches contrastives des chaînes de référence*, 2021, 82, p. 129-152.

FENG, Li-Jun, ELHADAD, Noémie & HUENERFAUTH, Matt. Cognitively motivated features for readability assessment. *EACL 2009 : Proceedings of the 12th Conference of the European Chapter of the ACL*. Athènes [Grèce], 2009, p. 229-237.

FLAUX, Nelly & VAN DE VELDE, Danièle. *Les noms en français : esquisse de classement*. Paris : Éditions Ophrys. 2000.

France Archives : https://francearchives.fr/fr/commemo/recueil-2015/38720. Dernière consultation : juillet 2022.
FOUGEROUSE, Marie-Christine, KAMBER, Alain & PELLAT, Jean-Christophe. *Grammaire du Français : FLE C1-C2 Perfectionnement*. Louvain-la-Neuve [Belgique] : De Boeck Supérieur. 2021.
FUCHS, Catherine. Les relatives et la construction de l'interprétation. *Langages*, 1987, 88, p. 95-127.
GALMICHE, Michel. Les ambiguïtés référentielles ou les pièges de la référence. *Langue française*, 1983, 57, p. 60-86.
GALMICHE, Michel. A propos de la définitude. *Langages*, 1989, 94, p. 7-37.
GALLAGHER, John. Traduction littéraire et études sur corpus. Dans : BALLARD, Michel & PINEIRA-TRESMONTANT, Carmen (Éds), *Les corpus en linguistique et en traductologie*. Artois : Artois Presses Université, 2007, p. 199-230.
GARDELLE, Laure. "Anaphora", "anaphor" and "antecedent" in nominal anaphora: definitions and theoretical implications. *Cercles*, 2012 (a), 22, p. 25-40.
GARDELLE, Laure. Pronominal gender in references to animals: a statistical analysis of the influence of six variables in discourse. *Anglophonia / Sigma*, 2012 (b), 16-32, p. 11-23.
GARDELLE, Laure, ROSSI, Caroline & VINCENT-DURROUX, Laurence. La gestion de l'anaphore en discours : complexités et enjeux. *Cahiers de praxématique* [en ligne], 2019, 72, p. 1-17.
GERDES, Kim & MIAO, Jun. Donner accès à l'œuvre de Fu Lei. In XU, Yun (Éd.), *Le monde spirituel de Fu Lei et sa signification dans le temps* (傅雷的精神世界及其时代意义), *Colloque international Fu Lei et traduction, traductions chinoises*. Nanjing [Chine] : Université de Nanjing, 2008, p. 15-18.
GARY-PRIEUR, Marie-Noëlle. Figurations de l'individu à travers différentes constructions du nom propre en français. *Cahiers de praxématique* [en ligne], 1996, 27, p. 1-11.
GARY-PRIEUR, Marie-Noëlle. La dimension cataphorique du démonstratif : étude de constructions à relative. *Langue française*, 1998, 120, p. 44-50.
GARY-PRIEUR, Marie-Noëlle. *Les déterminants du français*. Paris : Éditions Ophrys. 2011.
GARY-PRIEUR, Marie-Noëlle & LEONARD, Martine. Le démonstratif dans les textes et dans la langue. *Langue française*, 1998, 120, p. 5-20.
GARY-PRIEUR, Marie-Noëlle & NOAILLY, Michèle. Le nom propre. In : *Encyclopédie Grammaticale du Français* : http://encyclogram.fr. 2019. Dernière consultation : décembre 2019.
GERNSBACHER, Morton Ann. Coherence cues mapping during comprehension. Processing interclausal relationships. *Studies in the production and comprehension of text*, 1997, p. 3-22.
GERNSBACHER, Morton Ann & HARGREAVES, David. Accessing Sentence Participants: The Advantage of First Mention. *Journal of Memory and Language*, 1988, 27, p. 699-717.
GIVÓN, Talmy. Topic, Pronoun and Grammatical Agreement. Dans : GIVÓN, Talmy (Éds), *Subject and Topic*. New York [États-Unis] : Academic Press, 1976, p. 149-188.
GIVÓN, Talmy. Topic Continuity In Discourse: An Introduction. Dans : GIVÓN, Talmy (Éd.), *Topic Continuity In Discourse: A Quantitative Cross-Language Study*. Amsterdam [Pays-Bas] : John Benjamins Publishing Company, 1983, p. 5-42.
GIVÓN, Talmy. *Syntax: An Introduction (Volume II)*. Amsterdam [Pays-Bas] : John Benjamins Publishing Company. 2001.

GIVÓN, Talmy. *The story of zero*. Amsterdam [Pays-Bas] : John Benjamins Publishing Company. 2017.

GLIKMAN, Julie, GUILLOT, Céline & OBRY, Vanessa. Les chaînes de référence dans un corpus de textes narratifs médiévaux : traits généraux et facteurs de variation. *Langages*, 2014, 195, p. 43-60.

GRANGER, Sylviane. The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies. *Corpus-based approaches to contrastive linguistics and translation studies*, 2003, p. 17-29.

GROBOL, Loïc. Exploitation du corpus DEMOCRAT par apprentissage artificiel. *Langages*, 2021, 224, p. 129-145.

GROBOL, Loïc, LANDRAGIN, Frédéric & HEIDEN, Serge. Interoperable annotation of (co) references in the Democrat project. *Thirteenth Joint ISO-ACL Workshop on Interoperable Semantic Annotation*, Montpellier, 2017, p. 99-105.

GROSS, Gaston. *Manuel d'analyse linguistique : approche sémantico-syntaxique du lexique*. Lille : Presses Universitaires du Septentrion. 2012.

GUILLOT-BARBANCE, Céline & QUIGNARD, Matthieu. Chaînes de référence et structure textuelle dans les Essais sur la peinture de Diderot. *Discours*, 2019, 25, p. 3-26.

GUO, Chang. Chaînes de référence dans les textes encyclopédiques : étude contrastive en français et en chinois. *Travaux de Linguistique : Approches contrastives des chaînes de référence*, 2021, 82, p. 129-152.

GUNDEL, Jeanette & FRETHEIM, Thorstein (2004). Topic and Focus. In HORN, Laurence & WARD, Gregory (Éds), *The Handbook of Pragmatics* (pp. 175-196). Malden [États-Unis] : Blackwell Publishing.

HABERT, Benoît. Des corpus représentatifs : de quoi, pour quoi, comment. *Cahiers de l'Université de Perpignan*, 2000, 31, p. 11-58.

HABERT, Benoît. *Instruments et ressources électroniques pour le français*. Paris : Éditions Ophrys. 2005.

HABERT, Benoît. *Construire des bases de données pour le français*. Paris : Éditions Ophrys. 2009.

HALLIDAY, M. A. K. Spoken and Written modes of meaning. *Media Texts: Authors and Readers*, 1994, 7, p. 41-58.

HALLIDAY, M. A. K. & HASAN, Ruqaiya. *Cohesion in English*. New York [États-Unis] : Routledge. 2013 [1976].

HAMON, Philippe. Pour un statut sémiologique du personnage. *Littérature*, 1972, 6, p. 86-110.

HE, Yuanxiu. Chinese Topic Chains: An Analysis of the Topic Chains Used in Fortress Besieged. *Open Journal of Modern Linguistics*, 2013, 3-4, p. 344-347.

HEIDEN, Serge. The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme. *24th Pacific Asia Conference on Language, Information and Computation*, Sendai [Japan] : Institute for Digital Enhancement of Cognitive Development, Waseda University, 2010, p. 389-398.

HEIDEN, Serge, MAGUÉ, Jean-Philippe & PINCEMIN, Bénédicte. TXM : une plateforme logicielle open-source pour la textométrie – conception et développement. *10th International Conference on the Statistical Analysis of Textual Data (JADT 2010)*, Rome [Italie], 2010, p. 1021-1032.

HOBBS, Jerry. Coherence and Coreference. *Cognitive Science*, 1979, 3, p. 67-90.

HOU, Jiaqi. *Étude de la saillance référentielle dans un corpus de textes narratifs ---- analyse contrastive des chaînes de référence en français et en chinois*. Thèse doctorale en Sciences du langage, Université de la Sorbonne nouvelle – Paris 3, 2021.

HUANG, Yan. *The syntax and pragmatics of anaphora: A study with special reference to Chinese*. Cambridge [États-Unis] : Cambridge University Press. 1994.

HUANG, Yan. Anaphora: *A Cross-linguistic Study*. New York [États-Unis] : Oxford University Press. 2000.

HUYGHE, Richard. Noms d'objets et noms d'événements : quelles frontières linguistiques ? *Scolia*, 2012, 26, p. 81-103.

JENKINS, Christina. Les procédés référentiels dans les portraits journalistiques. *XV Skandinaviske romanistkongress*, Oslo [Norvège], 2002, p. 507-516.

JONASSON, Kerstin. *Le nom propre : constructions et interprétations*. Louvain-la-Neuve [Belgique] : Duculot. 1994.

JONASSON, Kerstin. Formes lexicales et pronominales dans des chaînes de référence en français et suédois. *Syntaxe et sémantique*, 2009, 1, p. 111-130.

KARTTUNEN, Lauri. Discourse referents. Dans : *Notes from the linguistic underground*. New York [États-Unis] : Academic Press, 1976, p. 363-385.

KEENAN, Edward & COMRIE, Bernard. Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry*, 1977, 8, p. 63-99.

KEENAN, Edward & HAWKINS, Sarah. The Psychological Validity of the Accessibility Hierarchy. Dans : KEENAN, Edward (Éd.), *Universal Grammar: 15 Essays*. Londres [Royaume-Uni] : Croom Helm, 1987, p. 60-88.

KERBRAT-ORECCHIONI, Catherine & TRAVERSO, Véronique. Types d'interactions et genres de l'oral. *Langages*, 2004, 153, p. 41-51.

KESIK, Marek. *La cataphore*. Paris : Presses universitaires de France. 1989.

KIM, Seo-Yeoung. *Etude contrastive des chaînes de référence et problèmes de la coréférence : en français et en coréen*, Thèse doctorale en Sciences du langage, Université de Strasbourg, 2023.

KLEIBER, Georges. *Problèmes de référence : descriptions définies et noms propres*. Paris : Klincksieck. 1981.

KLEIBER, Georges. Remarques sur la généricité et la spécificité. *Français (Le) Moderne*, 1983, 51, p. 36-49.

KLEIBER, Georges. Déictiques, embrayeurs, token-réflexives, symboles indexicaux, etc. : comment les définir ? *L'Information grammaticale*, 1986, 30, p. 3-22.

KLEIBER, Georges. Peut-on définir une catégorie générale de l'anaphore ? *Vox Romanica*, 1988, 47, p. 1-14.

KLEIBER, Georges. Référence, texte et embrayeurs. *Semen* [en ligne], 1989, 4.

KLEIBER, Georges. Anaphore-déixis : où en sommes-nous ? *Information grammaticale*, 1991, 51, p. 3-18.

KLEIBER, Georges. Cap sur les topiques avec le pronom *il*. *L'information grammaticale*, 1992, 54, p. 15-25.

KLEIBER, Georges. *Anaphores et pronoms*. Louvain-la-Neuve [Belgique] : Duculot. 1994 (a).

KLEIBER, Georges. Contexte, interprétation et mémoire : approche standard vs approche cognitive. *Langue française*, 1994 (b), p. 9-22.

KLEIBER, Georges. Sens, référence et existence : que faire de l'extra-linguistique ? *Langages*, 1997 (a), 127, p. 9-37.

KLEIBER, Georges. Référents évolutifs et pronoms : une suite. Dans : KLEIBER, Georges *et al.* (Éds), *La continuité référentielle*. Paris : Librairie Klincksieck, 1997 (b), p. 115-148.
KLEIBER, Georges. Marqueurs référentiels et théorie du centrage. *Linx*, 2002, 47, p. 107-119.
KRIPKE, Saul. Naming and necessity. Dans : DAVIDSON, Donald & HARMAN, Gilbert (Éds), *Semantics of natural language*. Dordrecht [Pays-Bas] : Reidel Publishing Company, 1972, p. 253-355.
KRIPKE, Saul. Speaker's reference and Semantic reference. *Midwest studies in philosophy*, 1977, 2, p. 255-276.
LEBART, Ludovic & SALEM, André. Statistique textuelle. Paris : Dunod, 1994.
LANDRAGIN, Frédéric. Saillance physique et saillance cognitive. *Corela, Cognition, représentation, langage* [en ligne], 2004, 2-2, p. 1-30.
LANDRAGIN, Frédéric. Une procédure d'analyse et d'annotation des chaînes de coréférence dans des textes écrits. *Varia*, 2011, 10, p. 61-80.
LANDRAGIN, Frédéric. Analyse, visualisation et identification automatique des chaînes de coréférences : des questions interdépendantes ? *Langue française*, 2014 (a), 195, p. 17-34.
LANDRAGIN, Frédéric. Anaphores et coréférences : analyse assistée par ordinateur. *Nouvelles perspectives sur l'anaphore. Points de vue linguistique, psycholinguistique et acquisitionnel*, Berne [Suisse] : Peter Lang, 2014 (b).
LANDRAGIN, Frédéric. Description, modélisation et détection automatique des chaînes de référence (DEMOCRAT). *Bulletin de l'AFIA*, 2016 (a), p. 11-16.
LANDRAGIN, Frédéric. Conception d'un outil de visualisation et d'exploration de chaînes de coréférences. *13ème Journées internationales d'Analyse statistique des données textuelles (JADT 2016)*, Nice, 2016 (b).
LANDRAGIN, Frédéric. Le corpus DEMOCRAT et son exploitation. Présentation. *Langages*, 2021, 224, p. 11-24.
LANDRAGIN, Frédéric & TANGUY, Noalig. Référence et coréférence du pronom indéfini *on*. *Langages*, 2014, 195, p. 99-115.
LANDRAGIN, Frédéric. Analyse, visualisation et identification automatique des chaînes de coréférence : des questions interdépendantes ? *Langue française*, 2017, 195, p. 17-34.
LANDRAGIN, Frédéric. Approches contrastives des chaînes de référence : Présentation. *Travaux de Linguistique : Approches contrastives des chaînes de référence*, 2021, 82, p. 7-16.
LANDRAGIN, Frédéric, POIBEAU, Thierry & VICTORRI, Bernard. ANALEC : a New Tool for the Dynamic Annotation of Textual Data. *International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul [Turquie], p. 357-362.
LANDRAGIN, Frédéric, POTIER, Juliette & BOTHUA, Meryl. Annotation manuelle d'expressions référentielles : expérimentations pour simplifier les prises de décisions et optimiser le processus. *9ème Journées Internationales de la Linguistique de corpus (JMC 2017)*, Grenoble, 2017.
LASSALLE, Emmanuel. *Structured Learning with Latent Trees : a Joint Approach to Coreference Resolution*. Thèse doctorale en Informatique, Université Paris Diderot (Paris 7). 2015.
LE GOFFIC, Pierre & WANG, Xiuli. Les pronoms interrogatifs – indéfinis du chinois : l'exemple de *shéi* 'qui/quelqu'un/quiconque'. *Verbum*, 2002, 24, p. 451-471.
LEBLOND, Aude. *Poétique du roman-fleuve de* Jean-Christophe *à* Maumort. Thèse doctorale en Littérature française, Université Sorbonne Nouvelle – Paris 3, 2010.

LEBLOND, Aude. Genèses du roman-fleuve : réduire le cycle à l'opus ? *Genesis*, 2016, 42, p. 53-64.
LEEMAN, Danielle. *Les déterminants du nom en français : syntaxe et sémantique*. Paris : Presses Universitaires de France. 2004.
LEROY, Sarah. *Le nom propre en français*. Paris : Éditions Ophrys. 2004.
LI, Charles & THOMPSON, Sandra. Subject and topic: A new typology of language. Dans : LI, Charles (Éd.), *Subject and topic*. New York [États-Unis] : Academic Press, 1976, p. 457-489.
LI, Charles & THOMPSON, Sandra. Third-Person Pronouns and Zero-Anaphora in Chinese Discourse in Discourse and Syntax. Dans : *Discourse and syntax*. Leyde [Pays-Bas] : Brill, 1979, p. 311-335.
LI, Charles & THOMPSON, Sandra. *Mandarin Chinese: A functional reference grammar. Vol. 3*. Berkeley et Los Angeles [États-Unis] : University of California Press. 1989.
LONGO, Laurence. Un corpus pour optimiser l'identification automatique des chaînes de référence. *Cahiers de praxématique*, 2014, 54, p. 246-262.
LONGO, Laurence & TODIRASCU, Amalia. RefGen : un module d'identification des chaînes de référence dépendant du genre textuel. *TALN 2010*, Montréal [Canada], 2010.
LONGO, Laurence & TODIRASCU, Amalia. Chaînes de référence et genre textuel pour la détection automatique de thèmes. *Journée ATALA*, Paris, 2012.
LONGO, Laurence & TODIRASCU, Amalia. Une étude de corpus pour la détection automatique de thèmes. *Texte et Corpus*, 2013, 4, p. 143-155.
LONGO, Laurence & TODIRASCU, Amalia. Vers une typologie des chaînes de référence dans des textes administratifs et juridiques. *Langages*, 2014, 195, p. 79-98.
LUNDQUIST, Lita. Un cas d'ambiguïté référentielle, aspects pragmatiques. Dans : KLEIBER, Georges & TYVAERT, Jean-Emmanuel (Éds), *L'anaphore et ses domaines*. Paris : Librairie Klincksieck, 1990, p. 229-249.
LUNDQUIST, Lita. Noms, verbes et anaphores (in)fidèles. Pourquoi les Danois sont plus fidèles que les Français. *Langue française*, 2005, 1, p. 73-91.
LYONS, John. Remarques sur les phrases possessives, existentielles et locatives. *Langages*, 1974 [1967], 34, p. 47-53.
LYONS, John. *Sémantique linguistique*. Traduit par DURANT, Jacques & BOULONAIS, Dominique. Paris : Librairie Larousse. 1990 [1978].
LYONS, John. *Linguistic Semantics: An Introduction*. Cambridge [États-Unis] : Cambridge University Press. 1995.
Manuel d'annotation du Projet Democrat, version 2.4.2., 7 février 2018.
MANUÉLIAN, Hélène & SCHNEDECKER, Catherine. Introduction des référents et premier maillon de suites coréférentielles : la réalité des textes. *Langages*, 2021, 224, p. 67-86.
MARQUE-PUCHEU, Christiane. Entre statut phrastique et statut textuel : l'exemple des énoncés situationnels. *Discours*, 2011, 9, p. 3-19.
MARTIN, Robert. Le fantôme du nom abstrait. Dans : FLAUX, Nelly & GLATIGNY, Michel (Éds), *Les Noms Abstraits, Histoire et Théorie*. Paris : Didier Samain. 1996, p. 41-50.
MASSERON, Caroline & SCHNEDECKER, Catherine. Production des chaînes de référence et problèmes de structuration textuelle : la redénomination comme catalyseur d'une conjoncture discursive problématique. *Linx*, 1995, hors-série 6-2, p. 423-442.
MATHET, Yann & WIDLÖCHER, Antoine. Évaluation des annotations : ses principes et ses pièges. *TAL*, 2016, 57, p. 73-98.

MÉLANIE-BECQUET, Frédérique & LANDRAGIN, Frédéric. Linguistique outillée pour l'étude des chaînes de référence : questions méthodologiques et solutions techniques. *Langages*, 2014, 195, p. 117-137.

MEYER-PLANTUREUX, Chantal. Romain Rolland, prix Nobel de littérature. *France Archives* : https://francearchives.fr/fr/commemo/recueil-2015/38720. 1994. Dernière consultation : novembre 2020.

MIAO, Jun. Analyses textométriques de corpus parallèles français-chinois : l'Aube de Jean-Christophe et ses trois traductions chinoises. *9ème Journées internationales d'Analyse statistique des Données Textuelles (JADT 2008)*, Lyon, 2008.

MIAO, Jun. *Fu Lei en chiffre – les pronoms personnels dans la traduction de* Jean-Christophe. Bloc personnel de Jun Miao : http://www.miaojun.net/index.php/about-event/traductologie. 2011. Dernière consultation : octobre 2018.

MIAO, Jun. *Approches textométriques de la notion de style du traducteur ---- Analyses d'un corpus parallèle français-chinois : Jean-Christophe de Romain Rolland et ses trois traductions chinoises*. Thèse doctorale en Traductologie, Université Sorbonne nouvelle – Paris 3, 2012.

MIAO, Jun & SALEM, André. *Comparaisons textométriques de traductions franco-chinoises*. Bloc personnel de Jun Miao : http://www.miaojun.net/index.php/about-event/traductologie. 2015. Dernière consultation : octobre 2018.

MILNER, Jean-Claude. Réflexions sur la référence. *Langue française*, 1976, 30, p. 63-73.

MILNER, Jean-Claude. *De la syntaxe à l'interprétation : quantités, insultes, exclamations*. Paris : Éditions du Seuil. 1978.

MILNER, Jean-Claude. *Ordres et raisons de langue*. Paris : Éditions du Seuil. 1982.

MITKOV, Ruslan. Towards More Comprehensive Evaluation in Anaphora Resolution. *International Conference on Language Resources and Evaluation* (*LREC*). 2000.

MITKOV, Ruslan. Towards a More Consistent and Comprehensive Evaluation of Anaphora Resolution Algorithms and Systems. *Applied Artificial Intelligence*, 2001, 15-3, p. 253-276.

MITKOV, Ruslan, EVANS, Richard, ORASAN, Constantin, BARBU, Catalina, JONES, Lisa & SOTIROVA, Violeta. Coreference and anaphora: Developing annotating tools, annotated resources and annotation strategies. *Proceedings of the Discourse, Anaphora and Reference Resolution Conference (DAARC2000)*. Barcelona [Espagne], 2000, p. 49-58.

MOIRAND, Sophie & REBOUL-TOURÉ, Sandrine. Nommer les événements à l'épreuve des mots et de la construction du discours. *Langue française*, 2015, 188, p. 105-120.

OBERLÉ, Bruno. *Étude des chaînes de référence dans les articles de recherche de format IMRad : problème d'annotation, analyses quantitative et qualitative*. Mémoire de master en Sciences du langage, Université de Strasbourg, 2016.

OBERLÉ, Bruno. SACR: A Drag-and-Drop Based Tool for Coreference Annotation. *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC 2018)*, Miyazaki [Japan], 2018.

OBERLÉ, Bruno, SCHNEDECKER, Catherine, BAUMER, Emmanuel, CAPIN, Daniéla, GLIKMAN, Julie, GUO, Chang, REVOL, Thierry, TODIRASCU, Amalia & TUSHKOVA, Julia. Les chaînes de référence dans les textes encyclopédiques du XIIème au XXIème siècle : étude longitudinale. *Travaux de linguistique*, 2018, 2, p. 67-141.

OBRY, Vanessa, GLIKMAN, Julie, GUILLOT, Céline & PINCEMIN, Bénédicte. Les chaînes de référence dans les récits brefs en français : étude diachronique (XIIIème – XVIème siècle). *Langue française*, 2017, 195, p. 91-110.

Onto Notes Chinese Co-reference Guidelines, version 4.0., BBN Technologies, 2008.

ORĂSAN, Constantin. PALinkA: A highly customisable tool for discourse annotation. Dans : *Proceedings of the Fourth SIGdial Workshop of Discourse and Dialogue*. 2003. p. 39-43.

PARIS, Marie-Claude. Le morphème « de » et la relativation en mandarin. *Cahier de Linguistique Asie Orientale*, 1977, 2, p. 65-76.

PARIS, Marie-Claude. *Problèmes de syntaxe et de sémantique en linguistique chinoise*. Paris : Presses universitaires de France. 1981.

PASSONNEAU, Rebecca. Computing reliability for coreference annotation. *Proceedings of the Language Resources and Evaluation Conference (LREC 2004)*, Lisbon [Portugal], 2004.

PINCEMIN, Bénédicte. Hétérogénéité des corpus et textométrie. *Langages*, 2012, 187, p. 13-26.

Penn Chinese Treebank Project : https://verbs.colorado.edu/chinese/ctb.html. Dernière consultation : juin 2020.

PERRET, Michèle. Quelques remarques sur l'anaphore nominale aux XIVe et XV siècles. *L'information grammaticale*, 2000, 87, p. 17-23.

PEYRAUBE, Alain. Adverbiaux et compléments de lieu en chinois. *Cahiers de linguistique – Asie orientale*, 1977, 1, p. 43-60.

POIBEAU, Thierry. The Multilingual Named Entity Recognition Framework. *European Conference of the Association for Computational Linguistics (EACL 2003)*. Budapest [Hongrie], 2003, p. 155-158.

POIBEAU, Thierry. Sur le statut référentiel des entités nommées. *Conférence Traitement Automatique des Lanugue 2005*. Dourdan, 2005, p. 173-183.

POTAU, Marta Recasens. *Coreference: Theory, Annotation, Resolution and Evaluation*. Thèse doctorale en Linguistique et Communication, University of Barcelona, 2010.

PRINCE, Ellen. Toward a Taxonomy of Given – New Information. *Syntax and semantics (vol 14)*. New York [États-Unis] : Academic Press. 1981. p. 223-255.

Projet ALLuSIF (*AnaLyse de la Lisibilité et SImplification pour le Français*) : https://lilpa.unistra.fr/fdt/projets/anciens-projets/allusif/ . 2013-2014. Dernière consultation : octobre 2019.

Projet DEMOCRAT (*DEscription et MOdélisation des Chaînes de Référence : outils pour l'Annotation de corpus (en diachronie et en langues comparées) et le Traitement automatique*) : http://www.lattice.cnrs.fr/democrat/. 2016-2020. Dernière consultation : novembre 2018.

Projet MC4 (*Modélisation Contrastive et Computationnelle des Chaînes de Coréférence*) : http://www.lattice.cnrs.fr/projets/projets-passes/projet-peps-mc4-modelisation-contrastive-et-computationnelle-des-chaines-de-coreference/. 2011-2012. Dernière consultation : juillet 2019.

QUIGNARD, Matthieu, LE MENE, Marine & LANDRAGIN, Frédéric. Elaboration du corpus DEMOCRAT ; procédures d'annotation et d'évaluation. *Langages*, 2021, 224, p. 25-46.

RABUT, Isabelle, WU, Yongyi, LIU, Hong, DROCOURT, Zhitang. *Méthode de chinois : premier niveau*. Paris : L'Asiathèque. 2012.

RAGHUNATHAN, Karthik, LEE, Heeyoung, RANGARAJAN, Sudarshan, CHAMBERS, Nathanael, SURDEANU, Mihai, JURAFSKY, Dan & MANNING, Christopher. A multi-pass sieve for

coreference resolution. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, MIT, Massachusetts [États-Unis], 2010, p. 492-501.
REICHLER-BÉGUELIN, Marie-José. Anaphore, cataphore et mémoire discursive. *Pratiques : linguistique, littérature, didactique*, 1988, 57, p. 15-43.
REICHLER-BÉGUELIN, Marie-José. Stratégies référentielles et variation. *Langue française*, 1997, 115, p. 101-110.
REICHLER-BÉGUELIN, Marie-José. L'usage des SN démonstratifs dans les *Fables* de La Fontaine. *Langue française*, 1998, 120, p. 95-109.
REUTER, Yves. L'importance du personnage. *Pratiques*, 1988, 60, p. 3-22.
RIEGEL, Martin, PELLAT, Jean-Christophe & RIOUL, René. *Grammaire méthodique du français*. Paris : Presses Universitaires de France. 2016 [1994].
ROHDE, Hannah & KEHLER, Andrew. Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience*, 2014, 29, p. 912-927.
ROUSSARIE, Laurent. Sens et dénotation. *Sémanticlopédie* : http://www.semantique-gdr.net/dico/index.php/Sens_et_d%C3%A9notation#D.C3.A9notation_.28Bedeutung. 29. 2007. Dernière consultation : avril 2019.
SACR (*Script d'Annotation des Chaînes de Référence*) : http://boberle.com/projects/sacr/index.html. Dernière consultation : novembre 2018.
SCHNEDECKER, Catherine. *Référence et discours : chaînes de référence et redénomination (Essai sur l'emploi en seconde mention du nom propre)*. Thèse doctorale en Philosophie, Sciences du langage et Communication, Département de Linguistique Générale de l'Université des Sciences humaines de Strasbourg, 1992.
SCHNEDECKER, Catherine. *Nom propre et Chaînes de référence*. Paris : Librairie Klincksieck. 1997 (a).
SCHNEDECKER, Catherine. Comment transformer une grenouille en (UN/Ø) prince charment ... ou l'alchimie des prédicats transformateurs hyperonymes. Dans : KLEIBER, Georges *et al.* (Éds), *La continuité référentielle*. Paris : Librairie Klincksieck, 1997 (b), p. 181-208.
SCHNEDECKER, Catherine. Les chaînes de référence dans les portraits journalistiques : éléments de description. *Travaux de linguistique*, 2005 (a), 51, p. 85-133.
SCHNEDECKER, Catherine. La notion de saillance : problèmes définitoires et avatars. Dans : INKOVA, Olga et al. (Éds), *Saillance : aspects linguistiques et communicatifs de la mise en évidence dans un texte*. Franche-Comté : Presses Universitaires de Franche-Comté, 2011, p. 21-48.
SCHNEDECKER, Catherine. Chaîne de référence et variation selon le genre. *Langages*, 2014, 195, p. 23-42.
SCHNEDECKER, Catherine. Les chaînes de référence : une configuration d'indices pour distinguer et identifier les genres textuels. *Langue française*, 2017, 195, p. 53-72.
SCHNEDECKER, Catherine. De l'intérêt de la notion de chaîne de référence par rapport à celles d'anaphore et de coréférence. *Cahier de praxématique : la gestion de l'anaphore en discours : complexité et enjeux* [en ligne], 2019, 72, p. 1-19.
SCHNEDECKER, Catherine. *Les chaînes de référence en français*. Paris : Éditions Ophrys. 2021.
SCHNEDECKER, Catherine & CHAROLLES, Michel. Les référents évolutifs : points de vue ontologique et phénoménologique. *Cahiers de linguistique française*, 1993, 14, p. 197-227.
SCHNEDECKER, Catherine & LANDRAGIN, Frédéric. Les chaînes de référence : présentation. *Langages*, 2014, 195, p. 3-22.

SCHNEDECKER, Catherine & LONGO, Laurence. Impact des genres sur la composition des chaînes de référence : le cas des faits divers. *3ème Congrès Mondial de Linguistique Française Lyon*, 2012, p. 1957-1972.
SCHNEDECKER, Catherine, GLIKMAN, Julie & LANDRAGIN, Frédéric. Les chaînes de référence : annotation, application et questions théoriques. *Langue française*, 2017, 195, p. 5-16.
SCHNEUWLY, Bernard. Genres et types de discours : considérations psychologiques et ontogénétiques. Dans : REUTER, Yves (Éd.), *Les interactions lecture-écriture*. Berne [Suisse] : Peter Lang. 1998, p. 155-173.
SCHNEUWLY, Bernard, BRONCKART, Jean-Paul, PASQUIER, Auguste, BAIN, Daniel & DAVAUD, Clairette. Typologie de texte et stratégies d'enseignement. *Le français aujourd'hui*, 1985, 69, p. 63-71.
STRAWSON, Peter. On referring. *Mind*, 1950, 59, p. 320-344.
STRAWSON, Peter. Singular Terms, Ontology and Identity. *Mind*, 1956, 65, p. 133-454.
STRAWSON, Peter. Singular Terms and Predication. *The Journal of Philosophy*, 1961, 58, p. 393-412.
Sémanticlopédie : http://www.semantique-gdr.net/dico/index.php/Accueil. Dernière consultation : avril 2019.
TENG, Shou-Hsin. Deixis, anaphora, and demonstratives in Chinese. *Cahiers de linguistique – Asie orientale*, 1981, 10, p. 5-18.
TEUBERT, Wolfgang & ČERMÁKOVÁ, Anna. Directions in corpus linguistics. *Lexicology and corpus linguistics*, 2004, p. 113-166.
TEUBERT, Wolfgang. Corpus Linguistics and Lexicography. *International Journal of Corpus Linguistics*, 2001, 6, p. 125-153.
TLFi (Trésor de la Langue Française informatisée) : http://atilf.atilf.fr/dendien/scripts/tlfiv4/showps.exe?p=combi.htm;java=no;. Dernière consultation : août 2022.
TODIRASCU, Amalia, FRANÇOIS, Thomas, BERNHARD, Delphine, GALA, Nuria, LIGOZAT, Anne-Laure & KHOBZI, Ratiba. Chaînes de référence et lisibilité des textes : le projet ALLuSIF. *Langue française*, 2017, 195, p. 35-52.
TODOROV, Tzvetan. *Les genres du discours*. Paris : Éditions du Seuil. 1978.
TUTIN, Agnès. A corpus-based study of pronominal anaphoric expressions in French. *Corpus,* 2002, 24, p. 516-521.
VICTORRI, Bernard. Le calcul de la référence. Dans : ENJALBERT, Patrice (Éd.), *Sémantique et traitement automatique du langage naturel*. Paris : Hermès, 2005, p. 133-172.
VINET, Marie-Thérèse & ZHOU, Huijun. La possession inaliénable en chinois mandarin et en français. *Cahiers de linguistique – Asie orientale*, 2003, 32, p. 157-193.
WANG, Xiuli. *La cohésion transphrastique : les compléments locatifs antéposés en chinois et en français*. Thèse doctorale en Sciences du langage, Université de la Sorbonne nouvelle – Paris 3, 2004.
WIDLOCHER, Antoine & MATHET, Yann. La plate-forme Glozz : environnement d'annotation et d'exploration de corpus. *TALN 2009*, Senlis, 2009.
WILKENS, Rodrigo, OBERLE, Bruno, LANDRAGIN, Frédéric & TODIRASCU, Amalia. French coreference for spoken and written language. *Proceedings of the 12th Edition of the Language Resources and Evaluation Conference (LREC 2020)*, Marseille, France. 2020.
WILMET, Marc. Les déterminants du nom en français : essai de synthèse. *Langue française*, 1983, 57, p. 15-33.
WILMET, Marc. *La détermination nominale*. Paris : Presses Universitaires de France. 1986.

XIA, Fei. *The Segmentation Guidelines for the Penn Chinese Treebank (3.0)*. 2000 (a)

XIA, Fei. *The Part-Of-Speech Tagging Guidelines for the Penn Chinese Treebank (3.0)*. 2000 (b).

XIA, Fei, PALMER, Martha, XUE, Nianwen, OKUROWSKI Mary Ellen, KOVARIK, John, CHIOU, Fudong, HUANG, Shizhe, KROCH, Tony, MARCUS, Mitch. Developing Guidelines and Ensuring Consistency for Chinese Text Annotation. *LREC*. 2000.

XU, Liejiong. Definiteness Effects on Chinese Word Order. *Cahiers de linguistique – Asie orientale*, 1995, 241, p. 29-48.

XU, Liejiong. A special use to the third singular pronoun. *Cahiers de linguistique – Asie orientale*, 1999, 281, p. 3-22.

XUE, Nianwen, CHIOU, Fudong & PALMER, Martha. Building a large-scale annotated Chinese corpus. *COLING: The 19th International Conference on Computational Linguistics*, Taipei [Chine], 2002.

YANG-DROCOURT, Zhitang, LIU, Hong, RABUT, Isabelle. *Méthode de chinois : deuxième niveau*. Paris : L'Asiathèque. 2011.

YEH, Chinglong & CHEN, Yichun. Zero Anaphora Resolution in Chinese with Shallow Parsing. *Journal of Chinese Language and Computing*, 2007, 17-1, p. 41-56.

ZHAO, Shanheng & NG, Hwee-Tou. Identification and resolution of Chinese zero pronouns: A machine learning approach. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, Prague [République tchèque], 2007, p. 541-550.

ZIMINA-POIROT, Maria. *Approches quantitatives de l'extraction de ressources traductionnelles à partir de corpus parallèles*. Thèse doctorale en Sciences du langage, Université de la Sorbonne nouvelle – Paris 3, 2004.

ZIMINA-POIROT, Maria. Exploration textométrique de corpus de traduction. *Meta*, 2005, 50-4.

ZIMINA-POIROT, Maria. Corpus multilingues : exploration textométrique de l'espace intertextuel. Dans : BALLARD, Michel & PINEIRA-TRESMONTANT, Carmen (Éds), *Les corpus en linguistique et en traductologie*. Artois : Artois Presses Université, 2007, p. 107-121.

ZRIBI-HERTZ, Anne. NP lui-même. Dans : KLEIBER, Georges & TYVAERT, Jean-Emmanuel (Éds), *L'anaphore et ses domaines*. Paris : Librairie Klincksieck, 1990, p. 377-402.

Référence en chinois

*Les notes sur les types de référence en chinois :
[A] Chapitres dans un ouvrage imprimé (article from anthology)
[C] Communication (collected papers)
[CP] Outils informatiques (computer program)
[D] Travaux universitaires (dissertation)
[DB] Bases de données (database)
[J] Articles de revues imprimés (journal article)
[J/OL] Articles de revues électroniques (journal article on line)
[M] Ouvrage imprimé (major work)
[N] Articles de journaux imprimés (newspaper article)
[N/OL] Articles de journaux électroniques (newspaper article on line)
[R] Rapport et autres types de documents (report)
[S] Spécification (specification)

曹长远 (CAO, Changyuan). 论专有名词 (Sur les noms propres en chinois) [N]. 临沂师专学报 (*Journal scientifique de l'Ecole Normale de Linyi*), 1988 : 101-103.

曹道根 (CAO, Daogen). 再论汉语是否有限定和非限定区分 (Revisiting the Finite vs Non-Finite Distinction in Mandarin Chinese) [J]. 当代语言学 (*Contemporary Linguistics*), 2018, 20, p. 19-39.

陈平 (CHEN, Ping). 释汉语中与名词性成分相关的四组概念 (Quatre paires de notions sur les groupes nominaux en chinois) [J]. 中国语文 (*Studies of the Chinese Language*), 1987, 2, p. 81-92.

陈平 (CHEN, Ping). 汉语双向名词句与话题-陈述结构 (Structure co-topicale et structure topique-commentaire en chinois) [J]. 中国语文 (*Studies of the Chinese Language*), 2004, 6, p. 493-507.

陈平 (CHEN, Ping). 汉语零形回指的话语分析 (Analyse discursive de l'anaphore zéro en chinois) [J]. 中国语文 (*Studies of the Chinese Language*), 1987, 5, p. 363-378. (à trouver)

陈平 (CHEN, Ping). 语言学的一个核心概念 « 指称 » 问题研究 (Référence, une notion essentielle pour la linguistique) [J]. 当代修辞学 (*Rhétorique contemporaine*), 2015, 189, p. 1-15.

陈平 (CHEN, Ping). 话语的结构与意义及话语分析的应用 (Structure discursive et application de l'ananlyse du discours) [J]. 当代修辞学 (*Rhétorique contemporaine*), 2017, 2, p. 1-11.

陈思和 (CHEN, Sihe). 傅雷的精神遗产 (Patrimoine littéraire de Lei Fu) [N]. 中国翻译 (*Chinese Translators Journal*), 2008, 4.

成柏泉 (CHENG, Boquan). 《约翰·克利斯朵夫》在中国 (Jean-Christophe en Chine) [J]. 读书 (*Lire*), 1980.

崔卫 (CUI, Wei). 俄汉话语零形回指之比较 (Etude contrastive de l'anaphore zéro en russe et en chinois) [J]. 中国俄语教学 (*Teaching Russian In China*), 1998, 2, p. 2-5.

董秀芳 (DONG, Xiufang). 汉语光杆名词指称特性的历时演变 (Diachronic Changes of the Referential Properties of Bare Nouns in Chinese) [J]. 语言研究 (*Studies in Language and Linguistics*), 2010, 30, p. 11-20.

方梅 (FANG, Mei). 关于复句中分句主语省略的问题 (Question sur l'ellipse du sujet dans les propositions subordonnées en chinois) [N]. 延边大学学报 : 社会科学版 (*Journal of Yanbian University: Social Sciences Edition*), 1985, 1, p. 44-54.

方梅 (FANG, Mei). 篇章语法与汉语篇章语法研究 (La grammaire discursive générale et la grammaire discursive de chinois) [J]. 中国社会科学 (*Social Sciences in China*), 2005, 6, p. 165-172.

方梅 (FANG, Mei). 汉语叙事语体中零形主语小句的句法整合功能 (Phrases sans sujet dans le genre narratif en chinois) [A]. 篇章分析——对汉语、法语指称链条分布规律的实证研究及其标注 (*Analyse textuelle : Etudes sur la répartition et la méthodologie d'annotation des chaînes de référence en chinois et en français*) [M]. 北京 (Pékin [Chine]) : 北京语言大学出版社 (Beijing Language and Culture University Press), 2018, p. 15-36.

傅雷 (FU, Lei). 傅雷谈艺录 (*Selected Writings on Arts*) [M]. 天津 (Tianjin [Chine]) : 天津人民出版社 (Tianjin People's Press). 2017 [1933 – 1966].

高彦梅 (GAO, Yanmei). 语篇语义框架研究 (*Constructing Semantic Frames in Discourse*) [M]. 北京 (Pékin [Chine]) : 北京大学出版社 (Peking University Press). 2015.

郭畅 (GUO, Chang). 百科文体中指称链条的法汉比较研究 (Étude contrastive sur les chaînes de référence dans les textes encyclopédiques) [A]. 篇章分析——对汉语、法语指称链条分布规律的实证研究及其标注 (*Analyse textuelle : Etudes sur la répartition et la méthodologie d'annotation des chaînes de référence en chinois et en français*) [M]. 北京 (Pékin [Chine]) : 北京语言大学出版社 (Beijing Language and Culture University Press), 2018, p. 53-59.

郭畅 (GUO, Chang). 对汉法关系从句的指称问题的对比研究——以 « 约翰·克里斯多夫 » 与傅雷译本为例 (Référence des subordonnées relatives en français et en chinois ---- étude de cas à partir des textes dans Jean-Christophe et sa traduction de Lei Fu) [A]. 建立在语篇自动标注平行语料库之上的汉法衔接方式对比研究 (*Études contrastives sur la cohérence et la cohésion en français et en chinois dans les corpus parallèles annotés automatiques*) [M]. 吉林 (Jilin [Chine]) : 吉林大学出版社 (Jilin University Press), 2021, p. 128-139.

郭纯洁 (GUO, Chunjie). 现代语言学研究方法 (*Méthodes de la recherche sur la linguistique contemporaine*) [M]. 北京 (Pékin [Chine]) : 科学出版社 (Science Publishing). 2015.

郭锐 (GUO, Rui). 汉语词类划分的论证 (Discussion sur la catégorisation grammaticale des mots en chinois) [J]. 中国语文 (*Studies of the Chinese Language*), 2001, 285, p. 494-560.

郭锐 (GUO, Rui). 现代汉语词类研究 (*Catégories grammaticales du chinois contemporain*) [M]. 北京 (Pékin [Chine]) : 商务印书馆 (The Commercial Press). 2002.

何卫, 钱军 & 姜望琪 (HE, Wei, QIAN, Jun & JIANG, Wangqi). 从语词到篇章 (Du lexique au discours) [A]. 语言学教程 (*Méthode de la linguistique*) [M]. 北京 (Pékin [Chine]) : 北京大学出版社 (Peking University Press), 2013, p. 68-88.

侯敏 & 孙建军 (HOU, Min & SUN, Jianjun). 汉语中的零形回指及其在汉英机器翻译中的处理对策 (Zero anaphora in Chinese and How to process it in Chinese-English MT) [J]. 中文信息学报 (*Journal of Chinese Information Processing*), 2004, 19, p. 14-20.

胡壮麟 & 彭宣维 (HU, Zhuanglin & PENG, Xuanwei). 语言与计算机 (Langue et informatique) [A]. 语言学教程 (*Méthode de la linguistique*) [M]. 北京 (Pékin [Chine]) : Peking University Press (北京大学出版社), 2013, p. 221-241.

黄娴 & 张克亮 (HUANG, Xian & ZHANG, Keliang). 汉语零形回指研究综述 (Zero anaphora in Chinese ---- the state of art) [J]. 中文信息学报 (*Journal of Chinese Information Processing*), 2009, 23, p. 10-15.

蒋平 & 承云 (JIANG, Ping & CHENG, Yun). 零形回指在汉语不同题材语篇中的分布 (Distribution of Zero anaphora in Chinese Texts of Different Genres) [J]. 现代语言学 (*Modern Linguistics*), 2014, 2, p. 179-185.

姜望琪 (JIANG, Wangqi). 主题, 主位与篇章 (Topique, sujet et discours) [J]. 外语教学与研究 (*Foreign Language Teaching and Research*), 2005, 37, p. 387-392.

姜望琪 (JIANG, Wangqi). 从句子语法到篇章语法 (The Turn from Sentence Grammar to Discourse Grammar) [J]. 中国外语 (*Foreign Languages in China*), 2007, 19, p. 20-27.

姜望琪 (JIANG, Wangqi). 语篇语言学研究 (*Discours Linguistics*) [M]. 北京 (Pékin [Chine]) : 北京大学出版社 (Peking University Press). 2011.

姜望琪 (JIANG, Wangqi). 意义 (Sens) [A]. 语言学教程 (*Méthode de la linguistique*) [M]. 北京 (Pékin [Chine]) : 北京大学出版社 (Peking University Press), 2013, p. 89-109.

教育部语言文字应用研究所计算语言学研究室 (l'Institut de la Linguistique appliquée du Ministre de l'Éducation de Chine). 国家语委语料库科研成果简介 (*Présentation du CNCorpus*) [R]. 2006.

李虹 (LI, Hong). 汉语指称问题的语义-语用界面研究 (*Etude sur les aspects sémantique et pragmatique de la référence en chinois*) [M]. 北京 (Pékin [Chine]) : 科学出版社 (Science Publishing). 2015.

李先银 (LI, Xianyin). 定名组合的指称功能与汉语多项定语的顺序 (A Study on the Function of Attribute and the Word Order of Multiple Attribute in Chinese) [J]. 语言与翻译 (*Language and Translation*), 2016, 1, p. 11-18.

李元瑞 (LI, Yuanrui). 现代汉语零形主语小句研究综述 (Rapport des études sur le sujet zéro en chinois contemporain) [J]. 现代语文 (*Modern Chinese*), 2016, 2, p. 14-19.

刘丹青 (LIU, Danqing) (2002). 汉语类指成分的语义属性和句法属性 (La référence générique en chinois : aspect sémantique et aspect syntaxique) [J]. 中国语文 (*Studies of Chinese Language*), 2002, 290, p. 411-422.

刘丹青 (LIU, Danqing). 汉语关系从句标记类型初探 (Etude initiative sur les marqueurs grammaticaux des propositions subordonnées relatives en chinois) [J]. 中国语文 (*Studies of Chinese Language*), 2005 (a), 304, p. 3-15.

刘丹青 (LIU, Danqing). 语法调查与研究中的从属小句问题 (Questions des propositions subordonnées dans les études linguistiques) [J]. 当代语言学 (*Contemporary Linguistics*), 2005 (b), 7, p. 193-212.

刘丹青 & 徐烈炯 (LIU, Danqing & XU, Liejiong) (1999). 焦点与背景, 话题及汉语 « 连 » 字句 (Focus, arrière-plan, topique et la structure LIAN en chinois) [J]. 中国语文 (*Studies of the Chinese Language*), 1999, 4, p. 243-252.

柳鸣九 (LIU, Mingjiu). 罗曼·罗兰与 « 约翰·克利斯多夫 » 的评价问题 (Sur les critiques littéraires de Romain Rolland et de Jean-Christophe) [J]. 外国文学 (*Littérature étrangère*), 1993, 1, p. 270-275.

柳鸣九 (LIU, Mingjiu). 纪念翻译巨匠傅雷 (A la mémoire du grand traducteur Lei Fu) [N]. 中国翻译 (*Chinese Translators Journal*), 2008, 4.

柳鸣九 (LIU, Mingjiu). 中国翻译文化灿烂历程之回顾 (Histoire des travaux de la traduction en Chine) [N]. 文艺报 (*Literature and Art Newspaper*), 2019, 5, p. 1-7.

陆俭明 (LU, Jianming). 现代汉语语法研究教程 (*Méthode de la grammaire du chinois contemporain*) [M]. 北京 (Pékin [Chine]) : 北京大学出版社 (Peking University Press). 2013.

罗常培 & 吕叔湘 (LUO, Changpei & LV, Shuxiang). 现代汉语规范化问题 (Sur la standardisation de chinois contemporain) [C]. 现代汉语规范问题学术会议文件汇编 (*Congrès de la Standardisation de chinois contemporain*) [R]. 北京 (Pékin [Chine]): 科学出版社 (Science Press). 1956.

吕叔湘 (LV, Shuxiang). 吕叔湘文集. 第一卷 : 中国文法要略 (*Essais de Lv Shu-Xiang. Volume I : Aperçu de la Grammaire du chinois*) [M]. 北京 (Pékin [Chine]): 商务印书馆 (The Commercial Press). 1990.

吕叔湘 (LV, Shuxiang). 现代汉语八百词 (*Huit cents mots en chinois contemporain*) [M]. 北京 (Pékin [Chine]) : 商务印书馆 (The Commercial Press). 1999.

屈承熹 (QU, Chengxi). 汉语篇章语法 : 理论与方法 (Mandarin Chinese Discourse Grammar: Theory and Practice) [J]. 俄语语言文学研究 (*Russian Language and Literature Studies*), 2006, 13, p. 1-15.

沈家煊 (SHEN, Jiaxuan). 谓语的指称性 (Can predicates be nominals ?) [J]. 外文研究 (*Foreign Studies*), 2013, 1, p. 1-13.

王秀丽 (WANG, Xiuli). 当代法国语言学研究动态 (*Etudes contemporaines de la linguistique française*) [M]. 北京 (Pékin [Chine]) : 北京语言大学出版社 (Beijing Language and Culture University Press). 2010.

魏慧萍 (WEI, Huiping). 也谈现代汉语的分期 (Nouvelles réflexions sur le chinois contemporain en diachronie) [N]. 常州工学院学报(社科版) (*Journal of Changzhou Institute of Technology: Social Science Edition*), 2008, 26, p. 75-78.

吴岳添 (WU, Yuetian). 百年回顾——法国小说在我国的译介和研究 (Cent ans d'histoire : les romans français traduits et introduits en Chine) [EB/OL]. 中国社会科学院, 中国外国文学网 (*Académie des sciences sociales de Chine : Littérature étrangère en Chine*) : http://ifl.cssn.cn/xshs/201211/t20121108_2404981.shtml. 2004. Dernière consultation : juillet 2019.

肖航 (XIAO, Hang). 国家语委现代汉语通用平衡语料库：标注语料库数据及使用说明 (*CNCorpus : Tagged Corpus Specification, Institute of Applied Linguistice Ministry of Education*) [S]. 2009.

肖红 & 许均 (XIAO, Hong & XU, Jun). 试论傅雷的翻译观 (On Fu Lei's viewpoint of translation) [N]. 四川外语学院学报 (*Journal of Sichuan International Studies University*), 2002 (18).

谢天振 (XIE, Tianzhen). 翻译文学——争取承认的文学 (La littérature de traduction : une littérature qui reste à être reconnue) [A]. 翻译思考录 (*Reflections On Translation*) [M]. 武汉 (Wuhan [Chine]) : 湖北教育出版社 (Hubei Education Press), 1998, p. 193-204.

熊岭 (XIONG, Ling). 现代汉语指称范畴研究 (*Study on Modern Chinese Referring Category*) [D]. Thesis Linguistics and Applied Linguistics. 华中师范大学文学院 (Central China Normal University), 武汉(Wuhan [Chine]). 2012.

熊岭 (XIONG, Ling). 汉语定指关系从句定指机制及定指度 (Mechanism and Definite Degree of Chinese Definite-Referring Clause) [N]. 汉语学习 (*Chinese Language Learning*), 2016, 2-1, p. 21-29.
徐赳赳 (XU, Jiujiu). 叙述文中 « 他 » 的话语分析 (Analyse discursive de ta (il) dans les textes narratifs) [J]. 中国语文 (*Chinese Language*), 1990, 218, p. 325-337.
徐赳赳 (XU, Jiujiu). 多动词小句中的零形式 (Forme zéro dans les propositions de plusieurs verbes prédicatifs) [J]. 中国语文 (*Chinese Language*), 1993, 5, p. 236 : 332-342.
徐赳赳 (XU, Jiujiu). 现代汉语篇章回指研究 (*Anaphora in Chinese Texts*) [M]. 北京 (Pékin [Chine]) : 中国社会科学出版社 (China Social Sciences Press). 2003.
徐赳赳 (XU, Jiujiu). 关系小句的语法和篇章特征分析 (Grammatical and Textual Features of Relative Clause) [J]. 汉语学习 (*Chinese Language Learning*), 2008, 5, p. 3-8.
徐赳赳 (XU, Jiujiu). 现代汉语篇章语言学 (*Text Linguistics in Contemporary Chinese*) [M]. 北京 (Pékin [Chine]) : 商务印书馆 (The Commercial Press). 2014.
徐赳赳 (XU, Jiujiu). 新媒体篇章的词频分析 (Fréquence des mots dans les textes sur les nouveaux médias : étude de cas à partir des conversations sur QQ) [A]. 篇章分析——对汉语、法语指称链条分布规律的实证研究及其标注 (*Analyse textuelle : Etudes sur la répartition et la méthodologie d'annotation des chaînes de référence en chinois et en français*) [M]. 北京 (Pékin [Chine]) : 北京语言大学出版社 (Beijing Language and Culture University Press), 2018, p. 37-42.
徐赳赳 (XU, Jiujiu) & WEBSTER, Jonathan. 叙述文中名词回指分析 (Anaphores nominales dans les textes narratifs en chinois) [N]. 语言教学与研究 (*Language Teaching and Linguistic Studies*), 1999, 4, p. 92-109.
许余龙 (XU, Yulong). 对比语言学的定义与分类 (Définition et typologie de la linguistique contrastive) [J]. 外国语 (*Journal of Foreign Languages*), 1992, 80, p. 12-17.
许余龙 (XU, Yulong). 汉语主从句间的回指问题 (Inter-clausal anaphora in Chinese complex sentences) [J]. 当代语言学 (*Contemporary Linguistics*), 2003, 5, p. 97-107.
许余龙 (XU, Yulong). 语篇回指实证研究中的数据库建设 (Database Development for the Testing of a Discourse Anaphora Resolution Model) [N]. 外国语 (*Journal of Foreign Languages*), 2005 (a), 156, p. 23-29.
许余龙 (XU, Yulong). 从回指确认的角度看汉语叙述体篇章中的主题标示 (Topic marking in Chinese narrative discourse: A perspective in discourse anaphora resolution) [J]. 当代语言学 (*Contemporary Linguistics*), 2005 (b), 7, p. 122-131.
许余龙 (XU, Yulong). 对比功能分析与翻译 (Contrastive Functional Analysis and Translation) [N]. 山东外语教学 (*Shandong Foreign Language Teaching Jornal*), 2006, 113, p. 3-8.
许余龙 (XU, Yulong). 再论语言对比基础的类 (Tertium Comparationis Revisited) [N]. 外国语 (*Journal of Foreign Languages*), 2007, 172, p. 21-27.
许余龙 (XU, Yulong). 对比语言学 (*Contrastive Linguistics*) [M]. 上海 (Shanghai [Chine]) : 上海外语教育出版社 (Shanghai Foreign Language Education Press). 2010.
许余龙 (XU, Yulong). 释几对与语篇回指相关的概念 (Quelques notions essentielles de l'anaphore dans le discours) [N]. 外文研究 (*Foreign studies*), 2013, 1, p. 1-8.
许余龙 (XU, Yulong). 英汉指称词语的语篇回指功能对比研究 (Referring Expressions and Their Anaphoric Function in Chinese and English Discourse. A contrastive Study) [N]. 外国语 (*Journal of Foreign Languages*), 2018, 21, p. 26-34.

荀恩东, 饶高琦, 肖晓悦 & 臧娇娇 (XUN, Endong, RAO, Gaoqi, XIAO, Xiaoyue & ZANG, Jiaojiao). 大数据背景下 BCC 语料库的研制 (The construction of the BCC Corpus in the age of Big Data) [J]. 语料库语言学 (*Corpus Linguistics*), 2016, 3, p. 93-118.

俞士汶, 朱学锋, 王惠 & 张芸芸 (YU, Shiwen, ZHU, Xuefeng, WANG, Hui & ZHANG, Yunyun). 现代汉语语法信息词典规格说明书 (Modern Chinese Grammar Information Dictionary Specification) [N]. 中文信息学报 (*Journal of Chinese Information Processing*), 1996, 10.

乐耀 (YUE, Yao). 从汉语书面叙事体的语篇结构看人物指称的分布和功能 (On the distribution and function of person reference in Chinese written narratives) [J]. 当代语言学 (*Contemporary Linguistics*), 2010, 4, p. 338-347.

张伯江 (ZHANG, Bojiang). 汉语名词怎样表现无指成分 (Sur les noms en chinois en emplois non référentiel) [A]. 庆祝中国社会科学院语言研究所建所 45 周年学术论文集 (*Recueil pour la célébrité du 45ème Anniversaire de l'Institut de linguistique de l'Académie chinoise des sciences sociales de Chine*) [M]. 北京 (Pékin [Chine]) : 商务印书馆 (The Commercial Press), 1997, p. 192-199.

张伯江 (ZHANG, Bojiang). 现代汉语的双及物结构式 (Construction bitransitive en chinois contemporain) [J]. 中国语文 (*Chinese Language*), 1999, 3-270, p. 175-184.

张伯江 (ZHANG, Bojiang). 汉语话题结构的根本性 (Sur la nature de la structure topicale dans le discours en chinois) [A]. 木村英树教授还历记念, 中国语文法论丛 (*A la mémoire du professeur Hideki Kimura, Grammaire du chinois*) [M]. Tokyo [Japon] : 日本白帝社 (Hakuteisha), 2013, p. 130-141.

张汉民 & 徐赳赳 (ZHANG, Hanmin & XU, Jiujiu). Structure "zai + endroit" dans le discours (篇章中的 "在+处所" 结构) [N]. 浙江师大学报 (社会科学版) (*Journal of Zhejiang Normal University: Social Sciences*), 1992, 2, p. 67-71.

张卫中 (ZHANG, Weizhong). 20 世纪中国文学语言演变述略 (Evolution du langage littéraire chinois au XXème siècle) [J]. 兰州学刊 (*Journal de Lanzhou*), 2009, 185, p. 74-77.

郑茗元 & 常霜林 (ZHENG, Mingyuan & CHANG, Shuanglin). 语料库语言学与机器翻译导论 (*Initiation à la linguistique de corpus et à la traduction automatique*) [M]. 北京 (Pékin [Chine]) : 中国水利水电出版社 (Waterpub). 2015.

中华人民共和国国家标准 (Standardization Administration of the People's Republic of China), 信息处理用现代汉语词类标记规范 (*Standard of POS Tag of Contemporary Chinese for CIP*) [S], 2006.

Corpus

ROLLAND, Romain. *Jean-Christophe*. Paris : Éditions Albin Michel. 2007 [1912].
ROLLAND, Romain. 约翰·克利斯朵夫 (*Jean-Christophe*) [M]. Traduit par FU, Lei (傅雷). 合肥 (Hefei [Chine]) : 安徽文艺出版社 (Anhui Literature and Art Publishing House). 2017 [1912].

Table des matières

Remerciements **1**

Abréviations **3**

Sommaire **5**

Introduction **7**

Partie I : Problèmes théoriques relatifs aux chaînes de référence dans les approches contrastives franco-chinoises **15**

Chapitre 1 : Revue des définitions de la notion de chaîne de référence et notre position relativement aux aspects franco-chinois **17**

1.1. Chaîne anaphorique de Chastain (1975) 17

1.2. Chaîne topicale de Li & Thompson (1979) et de Givón (1983) 18

1.2.1. Définitions : topique de discours et chaîne topicale 18

1.2.2. Deux positions face aux notions de topique et de chaîne topicale 20

1.2.3. Chaîne topicale en linguistique chinoise 21

1.3. Chaîne de référence de Corblin (1985, 1995) 23

1.4. Chaîne référentielle de Charolles (1988) 24

1.5. Chaîne pronominale de Xu (徐赳赳, 1990) 26

1.6. Chaîne de référence de Schnedecker (1992, 1997, 2021) 28

1.6.1. Apports textuels 28

1.6.2. Définition selon trois critères 29

1.6.3. Configuration selon la saillance du référent 31

1.7. Chaîne de coréférence des linguistes informaticiens 32

1.8. Synthèse des définitions 33

1.8.1. Domaines et approches de départ 35

1.8.2. Nombres de maillons 35

1.8.3. Formes des composants 36

1.8.4. Langues de base des études 37

1.9. Notre position relativement aux aspects textuels franco-chinois 37

1.10. Questions et discussions à la base des comparaisons des textes 39

1.10.1. Comparaison entre une copie d'élève et un extrait de roman-fleuve en français 39
1.10.2. Comparaison entre deux extraits de roman-fleuve en français et en chinois 41

Chapitre 2 : Trois notions-clés fondamentales pour examiner les problèmes des chaînes de référence relatifs à l'approche franco-chinoise 43
2.1. Notion de référence 44
2.2. Références endophoriques : anaphore et cataphore 45
2.2.1. Expressions endophoriques typiques : nominales, pronominales et zéros 48
2.2.2. Expressions endophoriques exceptionnelles : verbales et adverbiales 50
2.3. Référence exophorique : déixis 52
2.4. Entre l'endophore et l'exophore : l'ana-déixis 54
2.5. Notion de coréférence 55
2.5.1. Les noms propres : désignateurs autonomes 57
2.5.2. Les SN indéfinis 59
2.6. Bilan 61
Bilan de la Partie I 63

Partie II : Problèmes méthodologiques relatifs aux approches contrastives et à la constitution d'un corpus 65

Chapitre 3 : Motivations et représentativité en termes du choix d'un corpus 67
3.1. Caractéristiques de Jean-Christophe selon les aspects littéraires et culturels 68
3.1.1. Jean-Christophe et la traduction de Lei Fu dans le domaine de la littérature 68
3.1.2. Corpus parallèle souple 72
3.1.3. Influence des contextes culturels sur les expressions référentielles 73
3.2. Caractéristiques de Jean-Christophe en termes de chaînes de référence 78
3.2.1. Définition et caractéristiques d'un roman-fleuve 79
3.2.2. Chaînes de référence dans un roman-fleuve en français vs en chinois 82
3.2.2.1. Chaînes de référence dans divers genres en français 83
3.2.2.2. Chaînes pronominales et expressions anaphoriques zéro en chinois 89
3.2.3. Chaînes de référence au vu des séquences textuelles 93

3.2.3.1. Séquences narratives et dialogales dans le genre narratif 93
3.2.3.2. Séquences mono-référentielles et multi-référentielles 96
3.2.4. Chaînes de référence au vu de la longueur du texte 98
3.3. Bilan et hypothèses 100

Chapitre 4 : Outils pour la constitution d'un corpus bilingue annoté en expressions référentielles et en chaînes de référence 103
4.1. Outil de tokenisation et de catégorisation automatique du chinois : AiHanYu 103
4.1.1. Particularité du traitement des textes en chinois : tokenisation 103
4.1.2. AiHanYu : fonctions et opérations 104
4.1.3. Deux tests sur les erreurs systématiques dans les annotations automatiques 106
4.2. Outil d'annotation manuelle : SACR 109
4.2.1. Problèmes relatifs à l'annotation automatique 109
4.2.2. Problèmes soulevés spécifiquement en français et en chinois 110
4.2.3. Opérations en SACR 112
4.3. Outil de calcul et de visualisation : CRViewer 121
4.4. Bilan 123

Chapitre 5 : Processus du traitement outillé des expressions référentielles et des chaînes de référence en français et en chinois 125
5.1. Phase 1 – préparation 125
5.1.1. Étape 1 : nettoyage des textes 126
5.1.2. Étape 2 : partition des textes en épisodes 126
5.1.2.1. Trois principes de la partition des textes en épisodes 127
5.1.2.2. Application des trois principes dans le corpus 129
5.1.3. Étape 3 : tokenisation des textes en chinois 132
5.2. Phase 2 – annotation manuelle d'expressions référentielles et de chaînes de référence 133
5.2.1. Annotation des mentions : singletons et maillons 134
5.2.2. Annotation des caractéristiques des mentions 135
5.2.2.1. Place de la mention dans la chaîne 137
5.2.2.2. Catégories sémantiques du référent 138
5.2.2.2.1. Référents concrets 139
5.2.2.2.2. Référents abstraits 142

5.2.2.2.3. Référents coordonnés 144
5.2.2.3. Mode de la référence 144
5.2.2.4. Fonction syntaxique 147
5.2.2.5. Forme : déterminant, tête lexicale et modifieur 153
5.2.2.5.1. Mentions nominales : déterminants et noms propres 157
5.2.2.5.2. Mentions non nominales 163
5.2.2.5.3. Modifieur et tête lexicale des mentions nominales 168
5.2.2.5.4. Échelles hypothétiques d'accessibilité des expressions référentielles en chinois 170
5.2.3. Bilan 171
5.3. Phase 3 – évaluations et calculs 172
5.4. Bilan 173
Bilan de la Partie II 178

Partie III : Observations sur le corpus Jean-Christophe : régularités et spécificités contrastives des chaînes de référence en français et en chinois 179

Chapitre 6 : Régularités des chaînes de référence dans un roman-fleuve français : aspects quantitatifs et qualitatifs 181
6.1. Chaîne du personnage « Jean-Christophe » 185
6.1.1. Présentation générale 186
6.1.1.1. Catégories grammaticales des maillons 186
6.1.1.2. Densité référentielle et corrélation avec la catégorie des maillons ...187
6.1.1.3. Titre, premier maillon et deuxième maillon 189
6.1.1.4. Redémarrage de la chaîne de référence après une longue distance ..193
6.1.1.5. Répartition des maillons 195
6.1.2. Divers maillons nominaux 198
6.1.2.1. Noms propres : répartition quantitative et tactique 202
6.1.2.1.1. Coïncidences avec les changements textuels 202
6.1.2.1.2. Coïncidence avec les changements de point de vue 204
6.1.2.2. SN définis 206
6.1.2.2.1. Distance inter-maillonnaire 207
6.1.2.2.2. SN défini comme premier maillon de la chaîne et des séquences textuelles 207

6.1.2.2.3. Noms propres et SN définis : contextes multi-référentiels et fonctions syntaxiques 208
6.1.2.3. SN possessifs 209
6.1.2.4. SN démonstratifs : marqueurs de trois sous-types de rupture 211
6.1.2.5. SN indéfinis : typologie de modes de référence et caractéristiques de divers aspects 213
6.1.2.6. SN sans déterminant : référence dans la narration et dans le discours rapporté direct 215
6.1.2.7. Bilan 217
6.1.3. Modifieurs des maillons nominaux : longueur et restriction 219
6.1.3.1. Noms communs modifiés 220
6.1.3.2. Noms propres modifiés 224
6.1.4. Divers maillons non nominaux 227
6.1.4.1. Les longues suites des maillons non nominaux 228
6.1.4.2. Pronoms déictiques « je » et « tu » : séquences dialogales vs séquences narratives 219
6.1.4.3. Zéro : diverses contraintes syntaxico-discursives 230
6.2. Chaînes des trois personnages secondaires en français 231
6.2.1. Premier maillon : introduction des personnages secondaires 233
6.2.2. Noms propres et SN définis : différences suscitées par l'identité familiale 234
6.2.3. SN possessifs : reflet de divers degrés de dépendance et de relation familiale 236
6.2.4. SN sans déterminant : appellatifs dans les séquences dialogales 236
6.2.5. SN indéfinis et SN démonstratifs : maillons de quantités non significatives 237
6.2.6. Redémarrage des chaînes 238
6.2.7. Comparaison des chaînes des personnages secondaires avec celle du protagoniste 239
6.3. Chaînes des personnages épisodiques et éphémères 240
6.3.1. Des chaînes épisodiques aux chaînes éphémères 240
6.3.2. Chaînes épisodiques et éphémères selon les sous-types sémantiques 241
6.3.2.1. Désignation par les noms propres des sous-types d'humains 242
6.3.2.2. Introduction et redémarrage 243
6.3.3. Composition des chaînes épisodiques et éphémères 244
6.3.4. Chaînes vraisemblables de « on » : référents équivoques 245
6.4. Chaînes des référents non humains 248

6.4.1. Chaînes du « fleuve » et de « la musique » 248

6.4.2. Composants selon la catégorie grammaticale 250

6.5. Bilan 251

6.5.1. Chaînes de référence au niveau global du corpus : catégorie sémantique, répartition, longueur et composants 253

6.5.2. Chaînes de référence au premier niveau local : la première chaîne et la chaîne la plus longue dans chaque épisode 253

6.5.3. Chaînes de référence au deuxième niveau local : épisodes mono-référentiels et multi-référentiels 256

6.5.4. Chaînes de référence au troisième niveau local : séquences narratives et dialogales à l'intérieur des épisodes 258

6.5.5. Caractérisation globale des chaînes de référence dans le de roman-fleuve 259

Chapitre 7 : Propriétés des chaînes de référence dans un roman-fleuve traduit en chinois : approches contrastives franco-chinoises 261

7.1. Chaîne du protagoniste 'Christophe' Kelisiduofu 克利斯朵夫 263

7.1.1. Maillons nominaux selon la catégorie du nom et le déterminant 267

7.1.1.1. Noms propres : interprétations similaires et différentes selon les approches comparatives 268

7.1.1.2. SN sans déterminant 270

7.1.1.2.2. Dans le cadre phrastique : les SN sans déterminant à différentes positions syntaxiques 272

7.1.1.2.3. Dans le cadre inter-phrastique : influence des facteurs discursifs 274

7.1.1.2.4. Dans le cadre inter-phrastique : référence des SN sans déterminant à la place de sujet 275

7.1.1.3. SN démonstratifs : 'ce (classificateur)' zhe(ge) 这 (个) / 'ce-là (classificateur)' na(ge) 那 (个) + nom 276

7.1.1.3.1. Trois fonctions syntaxiques des SN démonstratifs 277

7.1.1.3.2. SN démonstratifs 278

7.1.1.3.3. Démonstratif endophorique et démonstratif endo-exophorique 280

7.1.1.4. Deux catégories nominales de plus faible quantité : SN possessifs et SN indéfinis 280

7.1.1.4.1. SN possessifs : wode 我的 / nide 你的 / tade 他的 + nom 281

7.1.1.4.2. SN indéfinis : '(un) classificateur' yi(ge) (一)个 + nom 284

7.1.1.5. Bilan 286
7.1.2. Modifieurs des maillons nominaux 288
7.1.2.1. Trois composants des maillons nominaux modifiés 288
7.1.2.2. Différences entre les maillons modifiés dans les deux langues 291
7.1.3. Quatre formes de maillons non nominaux 293
7.1.3.1. Pronoms personnels de troisième personne : composants principaux de la chaîne 293
7.1.3.2. Pronoms déictiques : emploi dans les séquences dialogales et narratives 295
7.1.3.3. Déterminants possessifs 297
7.1.3.4. Zéro : typologie et différences sino-françaises 299
7.1.4. Bilan 303
7.2. Chaînes des personnages secondaires 303
7.2.1. Démarrage et redémarrage 304
7.2.2. Catégories grammaticales des composants 305
7.2.2.1. Groupes de maillons nominaux dominants : noms propres et SN sans déterminant 306
7.2.2.2. Maillons nominaux de faible quantité : SN possessifs, SN démonstratifs et SN indéfinis 308
7.2.2.3. Pronoms personnels, déterminants possessifs et zéros 309
7.2.3. Bilan 310
7.3. Chaînes d'autres humains : épisodiques, éphémères et indéfinis 311
7.3.1. Chaînes des référents humains épisodiques et éphémères parallèles 311
7.3.2. Succession des expressions personnelles équivoques en chinois 312
7.4. Chaînes des référents non humains 313
7.4.1. Composants des chaînes parallèles du « fleuve » et de « la musique » 314
7.4.2. Chaînes des notions spatio-temporelles 315
7.4.3. Chaînes des référents relatifs au corps 316
7.5. Bilan 318

Chapitre 8 : Synthèse comparative 321
8.1. Référents du monde réel reflétés dans un roman-fleuve en français et en chinois 321
8.1.1. Nombre et densité des expressions référentielles, des référents et des chaînes de référence 321
8.1.2. Référents communs et distincts 323

8.1.3. Référents indépendants, référents associés et référents inaliénables : possessif obligatoire et facultatif 324

8.1.4. Corrélation entre la catégorie d'expression et la saillance du référent 325

8.2. Continuité des chaînes et progression du roman-fleuve 326

8.2.1. Premier maillon dans la chaîne de référence pour introduire les nouveaux référents 327

8.2.2. Noms propres : interprétations sémantique et cognitive selon différents contextes linguistiques et culturels 328

8.2.3. Pronoms personnels de troisième personne et zéros dans la continuité référentielle 329

8.2.4. Expressions référentielles univoques et équivoques 330

8.2.5. Critères relatifs à la variation des composants : hétérogénéité, stabilité, longueur et modifieurs 331

8.2.6. Expressions référentielles dans le cadre phrastique 334

8.2.7. Chaînes de référence dans les séquences narratives et dialogales dans le cadre d'un roman-fleuve 335

8.2.8. Modes de cohabitation des chaînes de référence dans les épisodes multi-référentiels 335

Conclusion 339

Bibliographie 343

Table des matières 365

Liste des tableaux 373

Liste des figures 377

Liste des annexes 379

Liste des tableaux

Tableau (1.1) : Synthèse des définitions de chaîne 34

Tableau (1.2) : Comparaison entre une copie d'élève et un extrait d'un roman-fleuve en français 40

Tableau (1.3) : Comparaison entre deux extraits d'un roman-fleuve en français et en chinois 41

Tableau (1.4) : Comparaison entre une copie d'élève en français, un extrait d'un roman-fleuve en français et un extrait d'un roman-fleuve en chinois 42

Tableau (3.1) : Interférence réelle ou potentielle des référents challengers (Schnedecker, 1997 : 80) 97

Tableau (6.1) : Catégories sémantiques des référents dans Jean-Christophe 181

Tableau (6.2) : Répartition des chaînes des personnages dans le corpus en français 183

Tableau (6.3) : Proportion des maillons de la chaîne « Jean-Christophe » selon la catégorie morpho-syntaxique 186

Tableau (6.4) : Corrélation entre la densité référentielle et la proportion des catégories non nominales 188

Tableau (6.5) : Maillons de la chaîne « Jean-Christophe » et leur répartition dans les sections et les épisodes en français (début de l'*Annexe 4*) 195

Tableau (6.6) : Caractéristiques des suites de maillons nominaux dans la chaîne « Jean-Christophe » 198

Tableau (6.7) : Fonctions syntaxiques des noms propres et des SN définis dans la chaîne « Jean-Christophe » 208

Tableau (6.8) : Fonctions syntaxiques des SN possessifs dans la chaîne « Jean-Christophe » ..210

Tableau (6.9) : Caractéristiques discursives des SN démonstratifs français dans la chaîne « Jean-Christophe » 211

Tableau (6.10) : Caractéristiques discursives des SN indéfinis français dans la chaîne « Jean-Christophe » 213

Tableau (6.11) : Caractéristiques discursives des SN sans déterminant français dans la chaîne « Jean-Christophe » 216

Tableau (6.12) : Comparaison des indices retenus des différentes catégories des maillons nominaux français dans la chaîne « Jean-Christophe » 217

Tableau (6.13) : Adjectifs qualitatifs et subjectifs français dans les modifieurs des noms communs dans la chaîne « Jean-Christophe » 222

Tableau (6.14) : Modifieurs adjectivaux des noms propres français dans la chaîne « Jean-Christophe » 225

Tableau (6.15) : Composition de la chaîne du protagoniste dans les épisodes mono-référentiels 231

Tableau (6.16) : Nombre de maillons et répartition des chaînes des personnages secondaires en français 232

Tableau (6.17) : Compositions des chaînes des personnages secondaires 232

Tableau (6.18) : Maillons à la place de redémarrage des chaînes des personnages secondaires 238

Tableau (6.19) : Comparaison entre la moyenne des chaînes des personnages secondaires et la chaîne du protagoniste 240

Tableau (6.20) : Sous-types d'humains dans les chaînes des personnages épisodiques et éphémères 242

Tableau (6.21) : Maillons de démarrage et de redémarrage des chaînes épisodiques et éphémères 244

Tableau (6.22) : Composants des chaînes épisodiques et éphémères 245

Tableau (6.23) : Composants des chaînes des référents non humains 250

Tableau (6.24) : Synthèse de quatre groupes de chaînes de référence 253

Tableau (6.25) : Coïncidence entre la première chaîne, la chaîne du protagoniste, et la chaîne la plus longue dans chaque épisode 255

Tableau (6.26) : Chaînes de référence dans les épisodes mono-référentiels 257

Tableau (6.27) : Chaînes de référence dans les séquences dialogales et dans le corpus 258

Tableau (6.28) : Composants des chaînes de référence dans les séquences dialogales et dans le corpus 259

Tableau (7.1) : Répartition des chaînes des personnages en chinois 261

Tableau (7.2) : Catégories des maillons chinois dans la chaîne du protagoniste 263

Tableau (7.3) : Proportions dissimilaires des maillons chinois et français dans la chaîne du protagoniste 264

Tableau (7.4) : Caractéristiques discursives des noms propres chinois dans la chaîne du protagoniste (section II, épisodes 12 et 13) 269

Tableau (7.5) : Fonctions syntaxiques des SN sans déterminant chinois dans la chaîne du protagoniste 273

Tableau (7.6) : Comparaison entre tous les SN sans déterminant chinois et les SN sans déterminant sujets dans la chaîne du protagoniste 275

Tableau (7.7) : Caractéristiques discursives des SN démonstratifs chinois dans la chaîne du protagoniste 277

Tableau (7.8) : Caractéristiques discursives des SN possessifs chinois dans la chaîne du protagoniste 281

Tableau (7.9) : Caractéristiques discursives des SN indéfinis chinois dans la chaîne du protagoniste 285

Tableau (7.10) : Bilan des maillons nominaux chinois dans la chaîne du protagoniste 286

Tableau (7.11) : Synthèse comparative des maillons modifiés en chinois et en français dans la chaîne du protagoniste 292

Tableau (7.12) : Caractéristiques discursives des pronoms personnels chinois dans la chaîne du protagoniste (section IV, épisode 37) 294

Tableau (7.13) : Caractéristiques discursives des déterminants possessifs chinois dans la chaîne du protagoniste (section III, épisodes 25 – 28) 298

Tableau (7.14) : Caractéristiques discursives des sujets zéro chinois dans la chaîne du protagoniste (section III, épisode 25) 300

Tableau (7.15) : Répartition, longueur et densité des chaînes des personnages secondaires en chinois et en français 304

Tableau (7.16) : Catégories des maillons chinois au démarrage et au redémarrage des chaînes des personnages secondaires 305

Tableau (7.17) : Composants des chaînes des personnages secondaires en chinois et en français 306

Tableau (7.18) : Comparaison des noms propres et des SN sans déterminant en chinois dans les chaînes des personnages secondaires (section II) 308

Tableau (7.19) : Caractéristiques discursives des maillons non nominaux chinois dans les chaînes des personnages secondaires (section I, épisode 2) 310

Tableau (7.20) : Comparaison entre les chaînes des personnages secondaires et celle du protagoniste en chinois 312

Tableau (7.21) : Succession des expressions personnelles équivoques en chinois 312

Tableau (7.22) : Chaînes des référents non humains en chinois 314

Tableau (7.23) : Chaînes du « fleuve » et de la « musique » en chinois et en français 315

Tableau (7.24) : Paramètres des chaînes de référence chinoises selon le référent 319

Tableau (7.25) : Composants des chaînes de référence en chinois selon le référent 319

Tableau (8.1) : Nombres, pourcentages et densités des expressions référentielles en français et en chinois 322

Tableau (8.2) : Catégories des référents dans le corpus en français et en chinois 324

Tableau (8.3) : Catégories des expressions référentielles dans le corpus en français et en chinois 325

Tableau (8.4) : Répartition et longueur des chaînes de référence en français et en chinois 326

Tableau (8.5) : Premier maillon des chaînes de référence en français et en chinois 327

Tableau (8.6) : Hétérogénéité des chaînes de référence en français et en chinois 331

Tableau (8.7) : Coefficient de stabilité des quatre chaînes de référence globales 332

Tableau (8.8) : Longueurs moyennes des expressions référentielles en français et en chinois 332

Tableau (8.9) : Maillons nominaux modifiés en français et en chinois 333

Tableau (8.10) : Fonctions syntaxiques des expressions référentielles en français et en chinois 334

Liste des figures

Figure (0.1) : Corrélations entre les catégories grammaticales et les fonctions syntaxiques en français et en chinois selon Lu (陆俭明, 2013 : 10, nous traduisons) 9

Figure (1.1) : Les périodes, les chaînes, les portées et les séquences dans l'organisation textuelle selon Charolles (1988 : 10) 25

Figure (1.2) : Configuration des chaînes pronominales dans les articles journalistiques chinois selon Xu (徐赳赳, 1990 : 330) 26

Figure (2.1) : Sous-types de la référence selon Halliday & Hasan (2013 [1976] : 33) 43

Figure (2.2) : Échelle d'indexicalité encodée par certaines catégories d'expressions indexicales selon Cornich (201 : 219) 55

Figure (2.3) : Échelle d'indexicalité encodée pour le chinois 55

Figure (3.1) : Versions de Jean-Christophe en chinois (non exhaustives) 70

Figure (3.2) : Familiarité des référents dans le roman-fleuve en français et dans sa traduction en chinois 76

Figure (4.1) : Plateforme de tokenisation et de catégorisation du chinois par AiHanYu 105

Figure (4.2) : Interface de l'importation de texte brut dans SACR 113

Figure (4.3) : Interface de délimitation des expressions référentielles en français dans SACR 114

Figure (4.4) : Dénomination des mentions par le nom du référent dans SACR 114

Figure (4.5) : Dénomination des mentions en chinois par des noms en lettre dans SACR 115

Figure (4.6) : Dénomination et construction des chaînes de référence dans SACR 115

Figure (4.7) : Mise en couleur des chaînes de référence dans SACR 116

Figure (4.8) : Importation du schéma des étiquettes dans SACR 117

Figure (4.9) : Ajout des étiquettes des propriétés linguistiques des mentions annotées dans SACR 117

Figure (4.10) : Interface d'annotation des expressions référentielles en chinois dans SACR ...118

Figure (4.11) : Calculs directs sur les textes annotés dans SACR 119

Figure (4.12) : Visualisation d'une chaîne spécifique sélectionnée dans SACR 119

Figure (4.13) : Interface de la recherche des mentions d'une même étiquette dans SACR 120

Figure (4.14) : Recherche des mentions d'une étiquette spécifique 120

Figure (4.15) : Calculs sur la base des textes annotés en tableau de chiffres par CRViewer 122

Figure (4.16) : Visualisation du nombre de mots des épisodes en français en histogramme par CRViewer 122

Figure (4.17) : Visualisation des paramètres en diagramme circulaire par CRViewer 123

Figure (6.1) : Typologie des SN indéfinis français selon Charolles (2002) 214
Figure (6.2) : Maillons nominaux modifiés français dans la chaîne « Jean-Christophe » 227
Figure (6.3) : Caractéristiques des unités textuelles du corpus Jean-Christophe 252
Figure (7.1) : Fonctions syntaxiques des maillons chinois dans la chaîne du protagoniste 267
Figure (7.2) : Interprétation de SN sans déterminant chinois selon Abbott (2006) 272
Figure (8.1) : Échelle de la saillance des référents dans Jean-Christophe en français et en chinois 323

Liste des annexes

Annexe 1 : Signes abrégés des catégories grammaticales dans AiHanYu 383

Annexe 2 : Examen du taux de bonne annotation par AiHanYu 384

Annexe 3 : Examen du taux de bonne annotation des 101 – 200 tokens concernant les expressions référentielles et coréférentielles par AiHanYu 386

Annexe 4 : Maillons français dans la chaîne « Jean-Christophe » 388

Annexe 5 : Suites des maillons nominaux français dans la chaîne « Jean-Christophe » 394

Annexe 6 : Séquences dialogales et épistolaire en français 396

Annexe 7 : Caractéristiques discursives des noms propres français dans la chaîne « Jean-Christophe » 406

Annexe 8 : Caractéristiques discursives des SN définis français dans la chaîne « Jean-Christophe » 414

Annexe 9 : SN possessifs français dans la chaîne « Jean-Christophe » 417

Annexe 10 : SN démonstratifs français dans la chaîne « Jean-Christophe » 420

Annexe 11 : SN indéfinis français dans la chaîne « Jean-Christophe » 420

Annexe 12 : SN sans déterminant français dans la chaîne « Jean-Christophe » 421

Annexe 13 : Noms communs modifiés français dans la chaîne « Jean-Christophe » 421

Annexe 14 : Noms propres modifiés français dans la chaîne « Jean-Christophe » 423

Annexe 15 : Maillons français dans la chaîne « Louisa » 425

Annexe 16 : Maillons français dans la chaîne « Melchior » 427

Annexe 17 : Maillons français dans la chaîne « Jean-Michel » 428

Annexe 18 : SN définis français dans la chaîne « Louisa » 430

Annexe 19 : SN définis français dans la chaîne « Melchior » 431

Annexe 20 : SN possessifs français dans la chaîne « Jean-Michel » 432

Annexe 21 : SN sans déterminant français dans la chaîne « Louisa » 433

Annexe 22 : SN sans déterminant français dans la chaîne « Melchior » 433

Annexe 23 : SN indéfinis français dans les chaînes des personnages secondaires 434

Annexe 24 : Chaînes des personnages épisodiques en français 435

Annexe 25 : Maillons français dans les chaînes des personnages épisodiques 437

Annexe 26 : Maillons français dans les chaînes des personnages éphémères 442

Annexe 27 : Succession des pronoms « on » en français 444

Annexe 28 : Chaînes des référents non humains en français 445

Annexe 29 : Maillons chinois dans la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫 455

Annexe 30 : Échantillon de noms propres chinois dans la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫 (section II, épisode 11) 461

Annexe 31 : Caractéristiques discursives des SN sans déterminant chinois dans la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫 461

Annexe 32 : Séquences dialogales et épistolaire en chinois 466

Annexe 33 : SN démonstratifs chinois dans la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫 ..474

Annexe 34 : Noms communs modifiés chinois dans la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫 475

Annexe 35 : Noms propres modifiés chinois dans la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫 480

Annexe 36 : Échantillon de pronoms personnels chinois dans la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫 (section IV, épisode 37) 483

Annexe 37 : Pronoms déictiques chinois dans la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫 485

Annexe 38 : Pronoms séictiques chinois dans les séquences narratives dans la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫 486

Annexe 39 : Échantillon de déterminants possessifs chinois dans la chaîne 'Christophe' Kelisiduofu 克利斯朵夫 (section III, épisodes 25 – 28) 486

Annexe 40 : Échantillon de zéros chinois dans la chaîne 'Christophe' *Kelisiduofu* 克利斯朵夫 (section III, épisode 25) 487

Annexe 41 : Maillons chinois de la chaîne 'Louisa' *Luyisha* 鲁意莎 488

Annexe 42 : Maillons chinois de la chaîne 'Melchior' *Manxiwo* 曼希沃 489

Annexe 43 : Maillons chinois de la chaîne 'Jean-Michel' *Yunehan-Mixier* 约翰·米希尔 491

Annexe 44 : Échantillon de noms propres chinois dans les chaînes des personnages secondaires (section II, épisodes 11 – 12) 493

Annexe 45 : Échantillon de SN sans déterminant chinois dans les chaînes des personnages secondaires (section II, épisodes 11 – 13) 494

Annexe 46 : SN possessifs chinois dans les chaînes des personnages secondaires 495

Annexe 47 : SN démonstratifs chinois dans les chaînes des personnages secondaires 496

Annexe 48 : SN indéfinis chinois dans les chaînes des personnages secondaires 497

Annexe 49 : Échantillon de maillons non nominaux chinois dans les chaînes des personnages secondaires (section I, épisode 2) 498

Annexe 50 : Maillons chinois dans les chaînes des personnages épisodiques 499

Annexe 51 : Caractéristiques des chaînes des personnages épisodiques en chinois 503

Annexe 52 : Maillons chinois dans les chaînes des personnages éphémères 504

Annexe 53 : Caractéristiques des chaînes des personnages éphémères en chinois 507

Annexe 54 : Chaînes des référents non humains en chinois 510

Université de Strasbourg

Chang GUO

École doctorale des **Humanités** (ED520)
Université de Strasbourg

ÉTUDE CONTRASTIVE DES CHAÎNES DE RÉFÉRENCE EN FRANÇAIS ET EN CHINOIS

ÉTUDE DE CAS À PARTIR D'UN CORPUS PARALLÈLE COMPOSÉ DE *JEAN-CHRISTOPHE* DE ROMAIN ROLLAND ET DE SA TRADUCTION PAR LEI FU

Résumé

Les discussions sur la notion de chaîne de référence et les notions connexes en français et en anglais ont été initiées dans les années quatre-vingt-dix. Cette notion et ses diverses approches relatives en français sont développées dans la thèse de Schnedecker (1992) et ses publications ultérieures ainsi que par de nombreux auteurs supportés par le projet ANR DEMOCRAT (2016-2020). Cette notion textuelle et discursive est fondamentale pour comprendre la continuité d'un texte, en particulier celle des textes longs lorsque les connecteurs ne suffisent pas à garantir seuls la cohérence textuelle. Les chaînes de référence semblent alors être les seules unités linguistiques qui permettent de comprendre ces textes longs comme un ensemble sans rupture de sujet. À partir d'une adaptation de la notion de chaîne de référence par les auteurs français, nous proposons dans un premier temps une définition de cette notion en chinois avec des hypothèses formulées au sujet de divers composants éventuels et divers critères quantitatifs et qualitatifs. Dans la deuxième partie de notre travail, nous cherchons à résoudre des questions méthodologiques détaillées par le biais de la constitution d'un corpus franco-chinois annoté des chaînes de référence dans les extraits du roman-fleuve *Jean-Christophe*, textes narratifs longs. Ce corpus nous permet, dans la troisième partie, d'obtenir les résultats quantitatifs pour toutes les chaînes de référence significatives du roman, et de démontrer, de manière qualitative, les caractéristiques des chaînes de référence en chinois dans le contexte général du roman et dans des conditions textuelles spécifiques.

Mots-clés : chaîne de référence, continuité référentielle, linguistique franco-chinoise

Université de Strasbourg

Chang GUO

École doctorale des **Humanités** (ED520)
Université de Strasbourg

CONTRASTIVE STUDY OF REFERENCE CHAINS IN FRENCH AND CHINESE

CASE STUDY BASED ON A PARALLEL CORPUS CONSTRUCTED BY *JEAN-CHRISTOPHE* OF ROMAIN ROLLAND AND ITS TRANSLATION BY LEI FU

Résumé en anglais

Discussions on the notion of reference chain and related notions in French and English started in the 1990s. This notion and its various relative approaches in French are developed in the thesis of Schnedecker (1992) and her subsequent publications, as well as by many authors supported by the ANR project *DEMOCRAT* (2016-2020). This textual and discursive notion is fundamental to understanding the continuity of text, in particular that of long texts when connectors alone are not enough to guarantee the textual coherence. The reference chains then seem to be the only linguistic units that allow these kind of long texts to be understood as a whole without breaking the subject. Based on an adaptation of the notion of reference chain by French authors, we first propose a definition of this notion in Chinese with hypotheses formulated about various possible components and various quantitative and qualitative criteria. In the second part of our work, we seek to solve detailed methodological questions through the constitution of a French-Chinese corpus annotated with the reference chains in the extracts of the river-novel *Jean-Christophe* which are long narrative texts. This corpus allows us, in the third part, to obtain quantitative results for all the significant reference chains of the novel, and to demonstrate, in the qualitative way, the characteristics of the reference chains in Chinese in the general context of the novel and in specific textual conditions.

Keywords : reference chain, referencial continuity, French-Chinese contrastive linguistics