Contribution de la linguistique de corpus à la constitution de langues contrôlées pour la rédaction technique : l'exemple des exigences de projets spatiaux

Maxime Warnier

▶ **To cite this version:**

Maxime Warnier. Contribution de la linguistique de corpus à la constitution de langues contrôlées pour la rédaction technique : l'exemple des exigences de projets spatiaux. Linguistique. Université Toulouse - Jean Jaurès, 2018. Français. NNT : . tel-02062833

HAL Id: tel-02062833
https://theses.hal.science/tel-02062833

Submitted on 10 Mar 2019

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 Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse - Jean Jaurès

Présentée et soutenue par :
Maxime Warnier
le lundi 10 septembre 2018

Titre :

Contribution de la linguistique de corpus à la constitution de langues contrôlées pour la rédaction technique :
l'exemple des exigences de projets spatiaux

École doctorale et discipline ou spécialité :
ED CLESCO : Sciences du langage

Unité de recherche :
CLLE (UMR 5263)

Directeur/trice(s) de Thèse :

Anne Condamines, Directrice de Recherche, CNRS

Jury :

Thierry Charnois, Professeur des universités, Université Paris 13 : rapporteur
Natalie Kübler, Professeure des universités, Université Paris Diderot : rapporteure
Ulrich Heid, Professeur, Universität Hildesheim : examinateur
Ludovic Tanguy, Maître de conférences HDR, Université Toulouse - Jean Jaurès : examinateur
Daniel Galarreta, Ingénieur, référent CNES pour la thèse : invité

cnes

RÉGION MIDI-PYRÉNÉES

RÉSUMÉ

L'objectif de notre travail, qui émane d'une demande de la sous-direction Assurance Qualité du CNES (Centre National d'Études Spatiales), est d'augmenter la clarté et la précision des spécifications techniques rédigées par les ingénieurs préalablement à la réalisation de systèmes spatiaux. L'importance des spécifications (et en particulier des exigences qui les composent) pour la réussite des projets de grande envergure est en effet désormais très largement reconnue, de même que les principaux problèmes liés à l'utilisation de la langue naturelle (ambiguïtés, flou, incomplétude) sont bien identifiés. Dès lors, de nombreuses solutions, plus ou moins formalisées, ont été proposées et développées pour limiter les risques d'interprétation erronée – dont les conséquences potentielles peuvent se révéler extrêmement coûteuses – lors de la rédaction des exigences, allant des langages logiques aux guides de rédaction, en passant par des outils de vérification semi-automatique.

Nous pensons que pour qu'elle soit réellement adoptée par les ingénieurs du CNES (qui ne sont actuellement pas tenus de suivre de règles de rédaction), la solution que nous nous efforçons de mettre au point se doit d'être à la fois efficace (autrement dit, elle doit limiter sensiblement le risque langagier) et aisée à mettre en place (autrement dit, elle ne doit pas bouleverser trop profondément leurs habitudes de travail, ce qui la rendrait contre-productive). Une langue contrôlée (en anglais : *Controlled Natural Language*), c'est-à-dire un ensemble de règles linguistiques portant sur le vocabulaire, la syntaxe et la sémantique, nous paraît être une réponse idéale à ce double besoin – pour autant qu'elle reste suffisamment proche de la langue naturelle. Or, les langues contrôlées pour la rédaction technique déjà existantes que nous avons envisagées, bien qu'élaborées par des experts du domaine, ne nous semblent pas toujours pertinentes d'un point de vue linguistique : certaines règles sont trop contraignantes, certaines ne le sont pas assez, d'autres encore ne se justifient pas vraiment.

Nous voudrions donc définir une langue contrôlée pour la rédaction des exigences en français au CNES. L'originalité de notre démarche consiste à systématiquement vérifier nos hypothèses sur un corpus d'exigences (constitué à partir d'authentiques spécifications de projets spatiaux) à l'aide de techniques et d'outils de traitement automatique du langage existants, dans l'optique de proposer un ensemble cohérent de règles (nouvelles ou inspirées de règles plus anciennes) qui puissent ainsi être vérifiées semi-automatiquement lors de l'étape de spécification et qui, surtout, soient conformes aux pratiques de rédaction des ingénieurs du CNES. Pour cela, nous nous appuyons notamment sur l'hypothèse de l'existence d'un genre textuel, que nous tentons de prouver par une analyse quantitative, ainsi que sur les notions de normalisation et normaison. Notre méthodologie combine les approches *corpus-based* et *corpus-driven* en tenant compte à la fois des règles imposées par deux autres langues contrôlées (dont l'adéquation avec des données réelles est discutée au travers d'une analyse plus qualitative) et des résultats offerts par des outils de *text mining*.

ABSTRACT

The aim of this work is to improve the clarity and precision of the technical specifications written in French by the engineers at CNES (Centre National d'Études Spatiales / National Centre for Space Studies) prior to the realization of space systems. The importance of specifications (and particularly of the requirements that are part of them) for the success of large-scale projects is indeed widely acknowledged; similarly, the main risks associated with the use of natural language (ambiguity, vagueness, incompleteness) are relatively well identified.

In this context, we would like to propose a solution that would be used by the engineers at CNES (who are currently not asked to follow specific writing rules): in that respect, we believe that this solution should be both effective (i.e. it should significantly limit the above-mentioned risks) and not too disruptive (which would make it counterproductive). A Controlled Natural Language (CNL) – i.e. a set of linguistic rules constraining the lexicon, the syntax and the semantics – seems to be an interesting option, provided that it remains close enough to natural language. Unfortunately, the CNLs for technical writing that we have examined are not always relevant from a linguistic point of view.

Our methodology for developing a CNL for requirements writing in French at CNES relies on the hypothesis of the existence of a textual genre; besides, we make use of existing Natural Language Processing tools and methods to validate the relevance of the rules on a corpus of genuine requirements written for former projects.

Chaque équivoque, chaque malentendu suscite la mort ;
le langage clair, le mot simple, peut seul sauver de cette mort.

— Albert Camus, *L'Homme révolté* (1951)

REMERCIEMENTS

Je tiens en premier lieu à remercier vivement ma directrice de thèse, Anne Condamines, pour la confiance qu'elle m'a accordée et pour sa grande disponibilité tout au long de mon doctorat : qu'il s'agisse de me guider dans ma réflexion, de me conseiller, de m'encourager ou de me relire, à Toulouse ou à distance, elle n'a jamais cessé de m'apporter son aide et ce travail lui doit évidemment beaucoup.

Je remercie sincèrement Natalie Kübler et Thierry Charnois d'avoir très aimablement accepté d'être les rapporteurs de cette thèse et Ulrich Heid et Ludovic Tanguy de m'avoir eux aussi fait l'honneur de faire partie de mon jury.

J'exprime aussi toute ma gratitude envers le Centre National d'Études Spatiales et le Conseil Régional de Midi-Pyrénées (désormais Occitanie) pour avoir co-financé cette thèse. Leur soutien, qui a rendu possible cette recherche, est très apprécié.

Au CNES, je souhaite également remercier toutes les personnes qui m'ont accordé leur temps ou fourni les données nécessaires à ce travail – en particulier Daniel Galarreta, mon référent, ainsi que Jean-François Gory et Nicolas Deslandres pour leur suivi. Tous trois ont fait preuve à mon égard d'une grande gentillesse et de beaucoup de disponibilité. Leur expertise dans un domaine qui m'était encore totalement inconnu il y a quelques années fut précieuse.

Je remercie de même Pierre-Vincent Paubel et Aurélie Mouneyrac pour m'avoir conseillé respectivement sur l'élaboration et l'analyse des résultats de l'enquête en ligne, ainsi qu'Elsa et Wendy pour leur relecture du manuscrit.

Je souhaiterais par ailleurs témoigner ma reconnaissance à tous les membres de CLLE-ERSS : le doctorant que je suis n'aurait pas pu souhaiter une unité plus accueillante et plus stimulante. J'ai également beaucoup apprécié travailler avec mes collègues enseignants des universités de Toulouse et de Tours.

Parmi mes ami(e)s docteur(e)s ou futur(e)s docteur(e)s, je salue l'équipe des JéTou 2015 (Florian, Francesca, Jean-Philippe, Laury, Luce, Olivier et Sophie) et les co-organisatrices du CEPEL 2016-2017 (Julie et Léa), avec qui les bons souvenirs l'emportent largement sur les moments de doute, ainsi qu'Aleksandra, Alienor, Assaf, Bénédicte, Camilla, Cécile, Céline, Caroline, Émilie, Fanny, Florence, François,

Giusi, Guillaume, Hugo, Julie, Karine, Karla, Laetitia, Lama, Manuel, Marianne, Marine, Matilde, Natalia, Nataly, Nikola, Pavel et Stéphanie (en espérant n'oublier personne !). Ce fut un réel plaisir de travailler et de rire avec vous.

Sans pouvoir nommer tout le monde, j'aimerais y ajouter toutes celles et tous ceux qui, à leur façon, de près ou de loin, m'ont apporté leur soutien ou qui ont rendu mon quotidien plus sympathique ces dernières années : professeurs ; co-auteurs ; ami(e)s en Belgique, à Toulouse et à Tours ; partenaires de sport ; membres du personnel de l'université et du restaurant universitaire (avec une mention spéciale à Fouzia pour sa bonne humeur indéfectible).

Enfin, je remercie Elsa pour son soutien sans faille durant les derniers mois de rédaction (malgré mon indéniable manque de disponibilité) et pour tous les savoureux repas qu'elle m'a préparés (et qu'il ne me reste plus qu'à éliminer. . .), et bien sûr ma famille – pour les milliers de kilomètres de route qu'ils ont dû faire pour m'aider à m'installer, pour leur affection, pour avoir toujours cru en moi et pour tant d'autres choses encore. Ils m'ont déjà dit être fiers de moi : qu'ils sachent que c'est réciproque.

Maxime

TABLE DES MATIÈRES

INTRODUCTION 1

I DEMANDE DU CNES ET PROPOSITIONS POUR UNE MÉTHODE DE CONSTITUTION D'UNE LANGUE CONTRÔLÉE POUR LA RÉDACTION DES EXIGENCES 5

1 LANGUE, NORMES ET USAGES 7
1.1 Introduction . . . 7
1.2 La langue naturelle comme moyen de transmettre de l'information . 7
1.3 Les normes langagières . . . 16
1.3.1 Normes subjectives et normes objectives . . . 16
1.3.2 Normes spontanées et normes imposées . . . 20
1.4 Genre textuel et sous-langage . . . 23
1.5 Langues contrôlées . . . 30
1.5.1 Introduction . . . 30
1.5.2 Historique . . . 30
1.5.3 Définition . . . 39
1.5.4 Évaluation . . . 42
1.6 Conclusion . . . 43

2 APPORT DE LA LINGUISTIQUE DE CORPUS POUR LA RÉDACTION DES EXIGENCES 45
2.1 La rédaction des exigences : besoins et enjeux . . . 45
2.1.1 Besoin, exigences et spécifications . . . 45
2.1.2 Enjeux . . . 47
2.1.3 Solutions existantes . . . 49
2.1.4 Conclusion . . . 51
2.2 Vers la constitution d'une langue contrôlée basée sur corpus . . . 52
2.2.1 Premier aperçu de la rédaction des exigences au CNES . . . 53
2.2.2 Propositions méthodologiques . . . 57
2.2.3 Corpus d'étude . . . 62
2.2.4 Conclusion . . . 66

II ANALYSE ET ÉVALUATION DES PREMIERS RÉSULTATS 69

3 ANALYSE DU CORPUS : APPROCHE CORPUS-BASED 71
3.1 Introduction . . . 71
3.2 Choix des langues contrôlées . . . 73
3.3 Choix des phénomènes langagiers . . . 76
3.3.1 Pronoms . . . 76
3.3.2 Conjonctions . . . 78
3.3.3 Longueur des phrases . . . 82
3.3.4 Temps verbaux et modalisation . . . 86
3.3.5 Voix passive . . . 92

3.3.6 Pronom « on » et nominalisations déverbales . . . 97
3.3.7 Conclusion . . . 100
3.4 Choix des textes et outils . . . 104
3.4.1 Textes de comparaison . . . 104
3.4.2 Outils . . . 106
3.5 Analyses . . . 109
3.5.1 Pronoms . . . 109
3.5.2 Conjonctions . . . 119
3.5.3 Longueur des phrases . . . 129
3.5.4 Temps verbaux . . . 134
3.5.5 Modalisation . . . 141
3.5.6 Voix passive . . . 145
3.5.7 Pronom « on » . . . 151
3.5.8 Nominalisations déverbales . . . 156
3.6 Conclusion . . . 159
4 ANALYSE DU CORPUS : APPROCHE CORPUS-DRIVEN 161
4.1 Introduction . . . 161
4.2 Première étude . . . 162
4.2.1 Méthodologie . . . 162
4.2.2 Statut des motifs . . . 165
4.2.3 Analyse des motifs extraits . . . 167
4.2.4 Élargir la notion de *boilerplate* . . . 170
4.2.5 Conclusion . . . 174
4.3 Deuxième étude . . . 175
4.3.1 Introduction . . . 175
4.3.2 Méthodologie . . . 178
4.3.3 Aperçu des résultats . . . 180
4.3.4 Conclusion . . . 182
4.4 Conclusion . . . 184
5 VALIDATION DES PROPOSITIONS : ENQUÊTE 187
5.1 Introduction . . . 187
5.2 Dispositif . . . 188
5.3 Choix des questions . . . 191
5.3.1 Modalisation . . . 192
5.3.2 Nombre de phrases . . . 193
5.3.3 Anaphore . . . 194
5.3.4 Agent . . . 195
5.4 Résultats . . . 196
5.4.1 Présentation générale . . . 196
5.4.2 Choix des analyses . . . 197
5.4.3 Modalisation . . . 199
5.4.4 Nombre de phrases . . . 202
5.4.5 Anaphore . . . 205
5.4.6 Agent . . . 207
5.5 Conclusion . . . 210

CONCLUSION ET PERSPECTIVES 213

ANNEXES 217

A GLOSSAIRE DES SPÉCIFICATIONS CNES 219

B RÉSULTATS DE LA DEUXIÈME ÉTUDE CORPUS-DRIVEN 227

C PRÉAMBULE DE L'ENQUÊTE 229

D EXIGENCES UTILISÉES POUR L'ENQUÊTE 231

D.1 Modalisation . . . 231

D.2 Nombre de phrases . . . 231

D.3 Anaphore . . . 232

D.4 Agent . . . 232

E RÉSULTATS DÉTAILLÉS DE L'ENQUÊTE 233

E.1 Modalisation . . . 233

E.2 Nombre de phrases . . . 240

E.3 Anaphore . . . 243

E.4 Agent . . . 248

BIBLIOGRAPHIE 253

INDEX DES AUTEURS 275

INDEX THÉMATIQUE 281

TABLE DES FIGURES

FIGURE 1 Influence réciproque de la norme et des usages 44
FIGURE 2 Approches *corpus-based* et *corpus-driven* 62
FIGURE 3 Approche *corpus-based* pour la définition de nouvelles règles 73
FIGURE 4 Fréquence des pronoms (par rapport au nombre total de mots) 110
FIGURE 5 Fréquence des conjonctions (par rapport au nombre total de mots) 119
FIGURE 6 Fréquence des phrases longues (par rapport au nombre total de phrases) 129
FIGURE 7 Fréquence du présent et du futur (par rapport au nombre total de formes verbales) 134
FIGURE 8 Fréquence du modal « devoir » (par rapport au nombre total de verbes) 141
FIGURE 9 Fréquence de la voix passive (par rapport au nombre total de verbes, hors participes passés) 145
FIGURE 10 Fréquence du pronom « on » (par rapport au nombre total de mots) 151
FIGURE 11 Fréquence des nominalisations déverbales (par rapport au nombre total de mots) 156
FIGURE 12 Analyse factorielle des correspondances (lignes = POS tags, colonnes = textes) 160
FIGURE 13 Approche *corpus-driven* pour améliorer la rédaction des exigences 161
FIGURE 14 Méthodologie (1re étude *corpus-driven*) 165
FIGURE 15 Méthodologie (2e étude *corpus-driven*) 181
FIGURE 16 Exemple de question posée dans l'enquête en ligne 190

LISTE DES TABLEAUX

Tableau 1 Composition du corpus 65
Tableau 2 Positions respectives de l'agent et du patient par rapport au procès 100
Tableau 3 Composition du corpus et des textes de comparaison 106
Tableau 4 Fréquence des pronoms (par rapport au nombre total de mots) 109
Tableau 5 Effectifs théoriques (attendus) pour les pronoms 111
Tableau 6 Résidus pour les pronoms 111
Tableau 7 Fréquence des conjonctions (par rapport au nombre total de mots) 119
Tableau 8 Longueur des phrases et fréquence des phrases longues (par rapport au nombre total de phrases) 129
Tableau 9 Fréquence de certains modes et temps (par rapport au nombre total de formes verbales) 134
Tableau 10 Fréquence du modal « devoir » (par rapport au nombre total de verbes) 141
Tableau 11 Fréquence de la voix passive (par rapport au nombre total de verbes, hors participes passés) 145
Tableau 12 Fréquence du pronom « on » (par rapport au nombre total de mots) 151
Tableau 13 Fréquence des nominalisations déverbales (par rapport au nombre total de mots) 156
Tableau 14 Profil des utilisateurs ayant répondu à l'enquête 196
Tableau 15 Synthèse : Modalisation 200
Tableau 16 Synthèse : Modalisation (2) 202
Tableau 17 Synthèse : Nombre de phrases 204
Tableau 18 Synthèse : Anaphore 206
Tableau 19 Synthèse : Anaphore (2) 207
Tableau 20 Synthèse : Agent 208
Tableau 21 Synthèse : Agent (2) 210

SIGLES ET ACRONYMES

Pour les sigles utilisés dans les spécifications du CNES, se reporter à l'annexe A.

ACE	Attempto Controlled English
AECMA	Association Européenne des Constructeurs de Matériel Aérospatial
AECMA SE	AECMA Simplified English
AFIS	Association Française d'Ingénierie Système
ASD	AeroSpace and Defence Industries Association of Europe
ASD-STE	ASD Simplified Technical English
CFE	Caterpillar Fundamental English
CNES	Centre National d'Études Spatiales
C(N)L	Controlled (Natural) Language [Langue (Naturelle) Contrôlée]
ECSS	European Cooperation for Space Standardization
GWR	*Guide for Writing Requirements* de l'INCOSE
HTML	HyperText Markup Language
IEEE	Institute of Electrical and Electronics Engineers
INCOSE	International Council on Systems Engineering
ISO	Organisation internationale de normalisation
LC	Langue Contrôlée, Langage Contrôlé
LN	Langue Naturelle, Langage Naturel
SDMC	Sequential Data Mining under Constraints
TAL(N)	Traitement Automatique du Langage (Naturel)
UML	Unified Modeling Language

INTRODUCTION

Le présent travail trouve son origine dans une demande émanant de la Sous-Direction Assurance Qualité du Centre National d'Études Spatiales (CNES) visant à améliorer la façon dont les ingénieurs rédigent les exigences et les spécifications (section 2.1.1) de projets spatiaux [1].

La rédaction des exigences est en effet une activité critique en ingénierie système et logiciel, une mauvaise compréhension par l'une des parties prenantes – en particulier le maître d'œuvre ou fournisseur – pouvant se révéler lourde de conséquences par la suite. Dans ce contexte, il n'est pas étonnant que l'ingénierie des exigences se fixe pour objectif une meilleure gestion des exigences, de leur phase initiale de définition jusqu'à leur validation finale, à travers un ensemble de principes, de bonnes pratiques et de méthodes, qui pour certains sont vus par beaucoup comme une condition nécessaire à la réussite de tout projet de grande ampleur (2.1.2).

Une exigence, en tant qu'expression d'un besoin, se doit ainsi d'être claire et univoque, « stated simply and [...] easy to understand » (ISO/IEC/IEEE29148, 2011 : 11). Mais cette simplicité d'expression et de compréhension n'a malheureusement rien d'évident. Lorsqu'elles sont rédigées en langue naturelle, les exigences s'exposent inévitablement à tous les risques inhérents à cette dernière (1.2) : ambiguïté, flou, imprécision, incomplétude. Partant de ce constat, de nombreuses solutions ont été proposées avec plus ou moins de succès, allant de la formalisation logique à la détection automatique du risque langagier pour assister les rédacteurs dans leur tâche, mais aucune ne semble cependant s'être véritablement imposée au détriment des autres (2.1.3).

Pour le linguiste qui veut bien s'y intéresser, cette situation – loin d'être un cas unique – pose inévitablement la question de la relation (conflictuelle ?) entre normes et usages (1.3). Il est en effet possible de contraindre (aux niveaux lexical, morphosyntaxique et sémantique notamment) la langue naturelle dans l'optique d'en faire un meilleur moyen de transmission de l'information, en particulier en contexte technique : c'est d'ailleurs ce que proposent par exemple les langues contrôlées (1.5), qui sont donc des normes langagières consciemment définies. Pour être réellement efficace et pour être effectivement adoptée, toutefois, une telle norme devrait selon nous nécessairement prendre en compte les usages de celles et ceux à qui elle est destinée. Ces usages sont en effet le plus souvent déjà contraints par d'autres types de normes (spontanées, celles-là) qui apparaissent dans les communautés de discours (notamment professionnelles) caractérisées par

1. [↑]Le sujet de cette thèse était initialement intitulé « Analyse linguistique comparée des guides rédaction technique et des usages réels dans les spécifications de systèmes spatiaux au CNES en vue d'améliorer la rédaction et la compréhension des exigences ».

des situations de communication récurrentes, qui elles-mêmes génèrent des régularités linguistiques pouvant constituer un genre textuel ou un sous-langage (1.4) à part entière.

À l'instar des socioterminologues, pour qui la normalisation devrait s'appuyer sur la normaison (1.3.2), nous pensons qu'une langue contrôlée pour la rédaction technique gagnerait à prendre en compte les régularités langagières caractéristiques du genre qu'elle entend contraindre. Ceci implique de redéfinir la méthode (hélas généralement très peu transparente) utilisée pour concevoir les langues contrôlées, ce qui est d'autant plus nécessaire que les recommandations proposées par ces dernières ne font que trop peu souvent l'objet d'une évaluation (1.5.4) et manquent parfois cruellement d'expertise linguistique (3.3.7).

Nous proposons donc de définir une méthodologie simple et pourtant innovante pour la constitution de langues contrôlées destinées à la rédaction technique qui tiennent compte de leurs utilisateurs finaux. Celle-ci relève de la linguistique de corpus (et de la linguistique outillée) et propose de combiner les approches traditionnellement dites *corpus-based* et *corpus-driven* (2.2.2). Suffisamment générale pour être facilement reproductible dans d'autres domaines, sur d'autres types de textes et dans d'autres langues, elle sera illustrée grâce aux exigences du CNES. Notre objectif, à ce stade, n'est pas d'aboutir à une langue contrôlée complète (ce qui nécessiterait évidemment davantage de moyens), mais de démontrer la faisabilité et l'intérêt de nos propositions. Nous effectuerons donc un premier état des lieux et proposerons quelques pistes concrètes d'améliorations.

Plutôt que de créer des règles *ex nihilo*, nous préférons prendre en compte à la fois l'état de l'art (représenté entre autres par les prescription récurrentes des langues contrôlées existantes) et les usages réels (attestés par le corpus spécialisé que nous avons constitué à cet effet) et vérifions systématiquement nos hypothèses avant de formuler des recommandations. Une analyse quantitative permet de confirmer l'existence d'un genre textuel et de caractériser objectivement ce dernier, tandis qu'une analyse linguistique qualitative permet de s'assurer de la pertinence des règles existantes et de les raffiner si nécessaire.

Par ailleurs, l'utilisation d'outils adaptés (dont beaucoup sont distribués librement à la communauté) permet de faire émerger les régularités directement du corpus, et ainsi de s'affranchir d'hypothèses *a priori*. Ces outils facilitent en particulier l'identification d'unités terminologiques ou phraséologiques qui, après validation, pourraient intégrer des ressources simplifiant ou améliorant la rédaction des exigences ou d'autres textes techniques (nous pensons en particulier aux *boilerplates* (4.2.4)).

L'utilisation d'un corpus permet donc d'obtenir des résultats objectifs, qui confirment ou au contraire remettent en question les prescriptions des langues contrôlées. Toujours dans l'optique de s'assurer de la pertinence de ces dernières et de prendre en considération ceux à qui elles s'adressent, nous estimons qu'une étape

de validation (indépendante) des propositions est également nécessaire avant que celles-ci ne leur soient éventuellement imposées – cette dernière étape devant en outre elle aussi permettre de les raffiner.

Notre démarche, qui part d'un besoin réel et s'appuie exclusivement sur des données authentiques, s'inscrit dans une vision de la linguistique comme science située. L'intérêt attendu est au moins double : d'une part, répondre à une demande concrète aux enjeux importants, et ainsi contribuer à faire (re)connaître les linguistes, dont le rôle (souvent confondu avec celui des grammairiens) est encore relativement méconnu en dehors du milieu académique ; d'autre part, améliorer notre connaissance du fonctionnement de la langue en contexte spécialisé, dont on sait qu'il diffère (parfois fortement) de celui de la langue dite générale.

Cette thèse comporte deux parties, elles-mêmes composées de plusieurs chapitres. Dans la première, nous commencerons par présenter la problématique générale dont relève notre recherche, qui est celle de la langue comme moyen (imparfait) de transmettre de l'information et de l'importance des normes langagières pour la communication, qu'elles soient objectives ou prescriptives, spontanées ou imposées. En lien direct avec celles-ci, nous porterons une attention particulière aux concepts de genre textuel, sous-langage et bien sûr langue contrôlée, qui sont centraux pour notre travail (chapitre 1). Nous exposerons ensuite plus concrètement la demande à laquelle nous répondons et la solution que nous entendons lui apporter (chapitre 2). Pour cela, nous présenterons dans un premier temps les problèmes que posent les exigences techniques ainsi que les solutions existantes, avant de proposer une méthodologie de constitution (ou d'amélioration) des langues contrôlées, que nous voyons comme notre principale contribution.

La seconde partie est consacrée à la mise en œuvre de nos propositions et à la présentation de quelques résultats. Nous analyserons et exploiterons notre corpus d'exigences, d'abord grâce à une approche plutôt *corpus-based* (chapitre 3, dans lequel nous présenterons aussi en détail deux langues contrôlées pour la rédaction technique et les phénomènes linguistiques concernés par quelques-unes de leurs recommandations), puis par une approche plutôt *corpus-driven* (chapitre 4), que nous illustrerons par deux études centrées sur la recherche de motifs fréquents. Enfin, pour évaluer la pertinence de certaines recommandations récurrentes, nous présenterons les résultats d'une enquête visant à recueillir le jugement d'experts et de non experts sur des exemples d'exigences formulées selon ces règles et sur des formulations différentes (chapitre 5).

Première partie

DEMANDE DU CNES ET PROPOSITIONS POUR UNE MÉTHODE DE CONSTITUTION D'UNE LANGUE CONTRÔLÉE POUR LA RÉDACTION DES EXIGENCES

1

LANGUE, NORMES ET USAGES

1.1 INTRODUCTION

Dans cette première partie, nous commencerons par mettre en lien la demande formulée par le CNES (que l'on peut résumer en une volonté d'augmenter la clarté et la précision des informations contenues dans les spécifications techniques rédigées en langue naturelle) avec les travaux déjà consacrés à la question de la langue comme moyen de transmettre de l'information utile – question à laquelle nous souhaitons donc apporter une contribution originale à travers notre travail. En ce sens, nous estimons que ce dernier s'inscrit pleinement dans une vision de la linguistique que l'on peut qualifier de science *située*, telle que CONDAMINES et NARCY-COMBES (2015 : 1) proposent de l'entendre :

> [...] situer la science c'est entrer dans une perspective où la recherche n'est plus appliquée à un projet, mais où elle est une partie de ce projet et où les deux se modifient réciproquement au fur et à mesure que le projet avance.

Loin de se limiter à une simple « application » du savoir linguistique à un cas pratique, cette démarche revêt en fait un intérêt double : d'une part, répondre à des besoins concrets (qui peuvent parfois constituer de véritables enjeux sociétaux) et ainsi démontrer l'intérêt de la discipline en dehors d'un cadre purement académique et, d'autre part, nous renseigner sur le fonctionnement de la langue dans des contextes réels particuliers, où elle a généralement été encore assez peu étudiée, en dépit de son importance non négligeable :

> On parle souvent alors de co-construction de l'objet, qui se nourrit des différentes compétences en jeu et qui permet de faire évoluer la réflexion, c'est-à-dire, de progresser vers un résultat qui prenne en compte à la fois la demande initiale et les connaissances sur le fonctionnement de la langue. In fine, les connaissances linguistiques se trouvent enrichies et le paradigme de la linguistique augmenté. (CONDAMINES et NARCY-COMBES, 2015 : 7)

Nous nous intéresserons dans un premier temps au fonctionnement général de la langue et aux risques qui y sont liés (ainsi qu'aux solutions qui ont été proposées et à leurs limites), avant d'envisager plus en détail le cas de la rédaction technique, qui nous concerne plus particulièrement.

1.2 LA LANGUE NATURELLE COMME MOYEN DE TRANSMETTRE DE L'INFORMATION

La demande qui nous a été adressée par le CNES et ce qu'elle implique – à savoir restreindre l'usage habituellement intuitif de la langue qu'en font ses locu-

teurs pour communiquer – nous amène nécessairement à nous interroger sur le fonctionnement des langues naturelles et sur les risques que l'utilisation de celles-ci peut engendrer, en particulier lorsque ces derniers sont méconnus de celles et ceux qui les utilisent (lesquels sont souvent qualifiés de locuteurs « naïfs »). En effet, elle soulève un constat contradictoire – et pourtant central – qui est à l'origine de cette thèse : d'un côté, la langue s'impose comme le moyen le plus évident (à tel point qu'elle se laisse bien souvent oublier) et le plus puissant (puisqu'elle permet d'exprimer tout ce que permettent d'exprimer tous les autres langages [1]) de transmettre des informations de natures très variées, dans des contextes qui le sont tout autant ; d'un autre côté (et l'on peut y voir le revers de la médaille de cette polyvalence), il semblerait que dans certains de ces contextes, elle ne soit pas toujours aussi optimale qu'on le souhaiterait et qu'elle n'aille pas sans poser quelques problèmes. La rédaction de spécifications techniques est l'un de ces contextes particuliers d'utilisation de la langue (en l'occurrence, de la langue écrite) où ces problèmes se posent avec une acuité certaine, tant une mauvaise compréhension entre interlocuteurs, si elle n'est pas détectée suffisamment tôt, peut se révéler lourde de conséquences.

Il découle de ce constat que si la langue naturelle apparait comme incontournable, il doit néanmoins être possible – du moins dans des contextes bien définis tels que celui que nous évoquions plus haut – d'en contraindre l'usage, et ce dans l'optique d'éviter les situations connues pour être sources de mécompréhension, et ainsi de la rendre plus « adaptée » par rapport à l'objectif poursuivi.

En effet, si la langue permet aux interlocuteurs de communiquer, les énoncés (qui en cela, selon la terminologie consacrée, se distinguent des simples phrases) qui composent un discours nécessitent, pour être correctement interprétés, que soit pris en compte le contexte de l'énonciation. La notion de *contexte* (au sens large) apparaît à la fois vaste et fondamentale, non seulement en pragmatique mais plus généralement en linguistique, où elle est très souvent mobilisée, sans toujours être définie. Duranti et Goodwin (1997 : 2) écrivaient d'ailleurs à son propos :

> Providing a formal – or simply explicit – definition of a concept can lead to important analytic insights. [...] However, it does not seem possible at the present time to give a simple, precise, technical definition of *context*, and eventually we might have to accept that such a definition may not be possible. At the moment, the term means quite different things within alternative research paradigms, and indeed even within particular traditions seems to be defined more by situed practice, by *use* of the concept to work with particular analytic problems, than by formal definition.

1. ↑Les mots nous permettent par exemple de décrire un schéma, une formule mathématique ou encore un geste, mais l'inverse n'est pas forcément vrai. Plus encore, la langue nous permet de parler de ce qui est, mais aussi de ce qui n'existe pas (ou plus, ou pas encore), de ce qui pourrait être, etc.

La situation n'est guère plus simple dans le champ des études francophones. Ainsi, à la même époque, KLEIBER (1999 : 167) faisait-il remarquer de façon plus critique :

> On le sait : point de salut en dehors du contexte ! Mais lorsqu'il s'agit de définir ce talisman, l'enthousiasme avec lequel on l'utilise pour se sortir d'une passe difficile cède bien vite la place à un embarras que l'on étouffe dans une note rapide et commode renvoyant le problème à plus tard et à d'autres. Tant la notion de contexte est protéiforme, flexible. Tant encore elle présente des contours flous ou variables selon les modèles et les emplois. Tant aussi elle donne lieu à des caractérisations divergentes et même contradictoires.

Les recherches sur la notion même de contexte n'ont pas cessé depuis lors et, si les définitions se sont multipliées, l'absence de consensus sur l'une d'elle semble toujours d'actualité (KERBRAT-ORECCHIONI, 2012).

Il faut au minimum distinguer entre un contexte proprement linguistique (généralement nommé *co-texte*), qui peut se définir comme « l'environnement linguistique [...] d'un élément particulier d'un énoncé [...][,] constitué de l'ensemble des unités linguistiques qui précèdent et qui suivent cet élément singulier (Pougeoise 1996 : 156) » (cité par KLEIBER, 1999 : 169) — car toute unité linguistique s'inscrit dans une chaîne plus large, sa « distribution », qui contribue à la production du sens – et un contexte extralinguistique, qui réfère quant à lui à la situation dans laquelle l'énoncé a été produit. On conçoit aisément que ce dernier puisse en fait englober des éléments nombreux et (infiniment) variés, allant de la situation des locuteurs, du lieu et du moment de l'énonciation (qu'il est nécessaire de connaître pour interpréter les indexicaux, par exemple) au contexte socioculturel dans lequel cette énonciation s'inscrit :

> L'usage contemporain de la notion de contexte est tributaire de cette ambiguïté. Tantôt le contexte est décrit comme rapport à la situation d'énonciation, qui tient alors lieu de rapport au « réel ». Faute d'une réflexion sur l'histoire et la tradition, comme la situation d'énonciation et la situation d'interprétation ne sont pas séparées alors, la situation est entendue comme un *hic et nunc*. D'où l'essor considérable de la pragmatique, essentiellement consacrée à l'oral. [...] Or, une situation n'est pas un simple état de choses, ni une relation intersubjective qui échapperait à l'histoire : elle ne se définit et ne se circonscrit que relativement à une légalité culturelle et sociale où elle prend son sens initial. En d'autres termes, une situation est une occurrence d'une pratique sociale. (RASTIER, 1998 : 98-99)

Cette distinction d'usage, généralement respectée dans la littérature (quoique les termes employés puissent légèrement varier), est, de fait, simple à observer et permet en partie de réduire la polysémie du terme *contexte*. Il n'est néanmoins pas exclu de penser que ces deux types de « contexte » soient en réalité liés et doivent idéalement être considérés ensemble (notamment quand on s'intéresse à la question des genres textuels) :

> Faut-il faire du contexte situationnel et du contexte linguistique deux objets d'étude indépendants, le premier relevant de la pragmatique, le second d'une linguistique supraphrastique ? Une sémantique des textes invite à penser qu'il s'agit de deux faces d'un seul objet, que ces deux contextes se correspondent et se révèlent l'un par l'autre. Le genre, qui se manifeste dans la matière linguistique même du texte, est comme l'empreinte significative et caractéristique d'une situation socioculturelle et du rapport au texte qu'elle prévoit. Autrement dit, le genre est le répondant, en tant que pôle intrinsèque du texte, des circonstances de rédaction et de lecture, qui font partie des pôles extrinsèques du texte. (Rastier et Pincemin, 1999 : 84)

Le co-texte, mais aussi le contexte extralinguistique sont en permanence exploités par les interlocuteurs dans leur activité d'interprétation des discours :

> Context is a source of information that can be used (is used, should be used, may be used, must be used) by a language processor to reduce (or completely eliminate) ambiguity, vagueness, or underspecification in its interpretations of the utterances that it processes. That's one of the effects of context. It constraints interpretation. (Hirst, 1997 : 5)

> It is impossible even to think of a sentence independently of context, and when we are asked to consider a sentence for which no context has been specified, we will automatically hear it in the context in which it has been most often encountered. (Fish, 1980 : 310)

Leur interprétation est donc orientée par le contexte, sans même que ceux-ci en aient nécessairement conscience : le contexte permet notamment de lever de nombreuses ambiguïtés potentielles, qui passent dans la plupart des cas inaperçues ; c'est seulement lorsqu'elles subsistent et que l'interlocuteur s'en rend compte que celui-ci va prêter attention au contexte et éventuellement conclure à un manque d'information lui permettant de trancher en faveur d'une interprétation particulière [2]. Il n'en va d'ailleurs pas différemment pour les spécifications techniques et pour ceux qui sont chargés de les appliquer (et donc, au préalable, de les interpréter) : si le co-texte (en tant que trace écrite) est le plus directement accessible et qu'il permet à lui seul de lever nombre d'ambiguïtés, le contexte extralinguistique joue également un rôle fondamental. En effet, contrairement à ce que l'on pourrait penser, ces textes (comme tous les autres textes) ne s'interprètent pas isolément et hors contexte, mais bien en fonction de la situation dans laquelle ils ont été produits et de ce qu'elle suppose de connaissances (censément) partagées par ceux qui y sont régulièrement confrontés [3].

En réalité, la prise en compte du contexte est indispensable à une communication fructueuse ; sans cela, il serait par exemple nécessaire de redéfinir en perma-

2. ↑Bien entendu, dans certains cas, la possibilité d'interprétations multiples est un effet de style recherché : double sens, jeux de mots, etc.

3. ↑Toute la difficulté étant, pour le linguiste qui s'intéresse à ces usages, de parvenir à « accéder » à cette situation qui ne lui est pas familière.

nence tous les concepts mobilisés, posant ainsi bien vite l'écueil de la circularité (un concept se définissant nécessairement par rapport d'autres). Cela suppose également que les interlocuteurs partagent une connaissance commune (sur la langue et sur le monde[4]) dans laquelle chacun ira puiser :

> It has been pointed out numerous times that even sentences such as 'The cat is on the mat' require a minimal (non-empty) set of *background* conditions, e.g., that the earth's gravitational field is in effect, that the mat is lying flat, that the mat is a standard (normal size) mat, and so on. (Akman, 2000 : 746)

Cela est d'autant plus vrai dans les situations de communication professionnelle, où les membres de la communauté professionnelle (et discursive) partagent des expériences et des objectifs similaires. Enfin, il va sans dire que pour que la communication réussisse, il est indispensable que les interlocuteurs y soient disposés, et donc qu'ils coopèrent en ce sens ; ils doivent pour cela respecter certains principes, tels que ceux édictés par les maximes de Grice et les théories qui s'en inspirent, notamment la théorie de la pertinence, dans laquelle le contexte est central (Wilson et Sperber, 1979 ; Sperber et Wilson, 1989).

Comme on l'a dit, le manque d'information linguistique ou situationnelle (donc, de contexte) peut conduire soit à l'impossibilité totale d'interpréter un énoncé, soit à l'impossibilité de décider quelle interprétation lui donner, lorsque plus d'une semble possible, soit encore à en privilégier une qui n'était pas celle qu'envisageait l'émetteur. Si ces risques restent le plus souvent négligeables dans la vie quotidienne, il n'en est évidemment pas de même lorsqu'une interprétation erronée peut avoir des conséquences graves. On peut par ailleurs penser que de ces trois cas, celui de l'interprétation divergente est le plus dangereux, puisque le risque n'est justement pas détecté :

> Le récepteur, quant à lui, peut être piégé par l'ambiguïté de deux façons. Soit il perçoit tout à la fois qu'il doit choisir entre deux sens et qu'il n'en a pas les moyens : le blocage ne pourra être levé que par des indications supplémentaires, venues du contexte, de la situation, ou du co-énonciateur (sollicité de s'expliquer). Soit – et c'est le cas le plus courant – il ne perçoit pas l'ambiguïté : à la manière des figures ambiguës [...], l'ambiguïté linguistique ne donne généralement lieu qu'à une seule lecture (la plus immédiate, la plus plausible pour le sujet). Quitte à faire un contre-sens, le récepteur sera incapable d'envisager le sens alternatif – du moins tant qu'un indice contraire ne viendra pas remettre en question son choix inconscient. (Fuchs, 2008 : 5)

Dans les faits, et bien que l'on n'en prenne rarement conscience, les énoncés regorgent le plus souvent d'ambiguïtés potentielles. Cela tient notamment à la nature très souvent polysémique des mots de la langue (ainsi qu'aux cas d'homonymie, plus marginaux d'un point de vue quantitatif) : Victorri et Fuchs (1996)

4. †Pour Harris (1988), qui identifie sept dimensions du contexte, les deux premières et principales sont la connaissance de la réalité (« knowledge of reality ») et la connaissance de la langue (« knowledge of language »).

ont calculé que plus de 40 % des mots recensés dans l'édition 1990 du *Petit Robert* sont polysémiques. Surtout, ces mots polysémiques font partie du vocabulaire le plus usuel, les termes strictement monosémiques se cantonnant pour la plupart aux domaines techniques. Ils en concluent dès lors que « ce phénomène est tout à fait central dans la langue : il représente la règle plutôt que l'exception » (p. 5).

Les facteurs d'incertitude dans l'interprétation d'un énoncé sont nombreux et ne se limitent pas à l'ambiguïté au sens strict, tant s'en faut. Fuchs (2008) choisit de les classer en trois grandes catégories : l'incertitude interprétative par défaut ou excès de sens, par concurrence de sens et par glissement de sens.

Fuchs parle de sous-détermination lorsque l'incertitude interprétative est due à un défaut de sens. Ce défaut de sens peut s'apparenter à de l'incomplétude, « lorsque l'interprétation se heurte à l'impossibilité de reconstruire certains éléments dont l'absence est signalée en creux dans l'énoncé » (ex. : « il a donné ses livres » : à qui les a-t-il donnés ?). Il peut également s'apparenter à de l'imprécision, « lorsque l'interprétation débouche sur un sens trop 'général' » (ex. : « il s'est passé quelque chose »), qui est affaire de degré. Il peut tenir au caractère « vague » de certains termes, « le flou [étant] lié au caractère gradable du prédicat » (ex. : « Pierre est grand » : grand comment ?). L'incertitude interprétative peut enfin être due à la difficulté d'identifier les référents, ce qui est le cas « quand une expression linguistique est en emploi 'approximatif' (c'est-à-dire utilisée pour renvoyer à un référent auquel elle ne s'applique que de façon limite) », avec les indexicaux (« aujourd'hui », « je », etc.) et avec les expressions appelées « descriptions définies » (ex. : « le président de la République »), pour lesquelles « il n'est pas toujours clair de savoir si l'on a affaire à un emploi 'référentiel' ou 'attributif' ». Comme nous le verrons plus loin, certains de ces problèmes (incomplétude due à l'absence d'un argument du verbe, imprécision et flou liés à un terme trop vague) sont bien connus en ingénierie des exigences.

Elle parle, au contraire, de surdétermination lorsque l'incertitude interprétative tient à un excès de sens, où « [d]errière le sens explicite peut [...] se cacher un second sens, implicite, qui est à reconstruire ». On distingue entre présupposé (« sens implicite dont la reconstruction est entraînée par la forme même de certaines expressions » ; ex. : si « Jean a cessé de fumer », il est présupposé que Jean fumait avant) et sous-entendu (« sens implicite qui ne peut être inféré à partir de sa forme ou de sa signification que dans certaines conditions » ; ex. : « la poubelle est pleine » ... s'agit-il d'un acte de langage indirect adressé à l'interlocuteur ?).

La deuxième catégorie établie par Fuchs est celle de l'incertitude interprétative par concurrence de sens, c'est-à-dire celle de l'ambiguïté à proprement parler :

> L'ambiguïté constitue un cas de figure très particulier : celui où l'interprétation d'un énoncé se trouve confrontée à un choix nécessaire entre plusieurs (généralement deux) sens mutuellement exclusifs, sans que soient données les clés pour opérer ce choix (faute d'indices provenant d'un contexte lin-

> guistique plus large et/ou d'une meilleure connaissance de la situation). Il s'agit donc toujours d'un moment de l'interprétation, appelé à être dépassé ultérieurement. (FUCHS, 2008 : 4)

L'ambiguïté peut procéder de l'homonymie (« ambiguïté accidentelle ») ou de la polysémie (« ambiguïté conjoncturelle »). Elle est traditionnellement reliée par les linguistes au niveau de l'analyse linguistique duquel elle dépend : ambiguïté phonétique (homophonie), ambiguïté morphologique, ambiguïté lexicale, ambiguïté syntaxique de constituants, ambiguïté syntaxique actancielle, ambiguïté sémantico-logique et ambiguïté pragmatique (FUCHS, 1995).

La dernière catégorie est celle de l'incertitude par glissement de sens, que l'auteure nomme la « plurivocité constitutive » : « l'interprétation de l'énoncé est constitutivement instable : elle semble osciller et couvrir toute une zone de sens variable, sans pouvoir se fixer en un point déterminé. L'incertitude tient alors, non pas à l'impossibilité d'un choix entre plusieurs sens mutuellement exclusifs, mais à la difficulté à pointer un sens précis » (ex. : « Quelques averses se produiront encore » : « de nouvelles averses » et/ou « viendront prolonger la série des averses »).

On réalise par conséquent, lorsque l'on recense les sources potentielles d'incertitude interprétative, que celle-ci est en réalité omniprésente et qu'elle est difficilement évitable (et ce, même dans les discours qui cherchent à tout prix à s'en prémunir) : impossible, en effet, d'obtenir une précision parfaite, qui nécessiterait d'expliciter indéfiniment chacun des mots employés. Du reste, les interlocuteurs (tant qu'ils sont prêts à fournir l'effort cognitif requis par l'activité d'interprétation) s'en accommodent généralement très bien, la langue restant, pour l'espèce humaine, le moyen de communication par excellence.

On aurait cependant tort de limiter le langage à un simple moyen de transmission de l'information. Bien que cela soit sans doute sa fonction la plus évidente, c'est en effet loin d'être la seule.

Ainsi, dans son célèbre schéma de la communication, JAKOBSON (1960) associe à chacun des six « facteurs » impliqués dans la communication verbale (le *contexte*, l'*émetteur*, le *récepteur*, le *message*, le *contact* et le *code*) une fonction différente (respectivement *référentielle*, *émotive*, *conative*, *phatique*, *métalinguistique* et *poétique*). Et s'il reconnaît un rôle particulier à la fonction référentielle (ou *dénotative*), orientée vers le contexte, il ne manque pas de souligner l'importance relative des cinq autres fonctions :

> Although we distinguish six basic aspects of language, we could, however, hardly find verbal messages that would fulfill only one function. The diversity lies not in a monopoly of some one of these several functions but in a different hierarchical order of functions. The verbal structure of a message depends primarily on the predominant function. But even though a set [...] toward the referent, an orientation toward the context — briefly, the so-called

> referential, "denotative", "cognitive" function — is the leading task of numerous messages, the accessory participation of the other functions in such messages must be taken into account by the observant linguist. (Jakobson, 1960 : 3-4)

Selon le type de discours ou le genre de texte auquel on a affaire, certaines de ces fonctions peuvent être très effacées ou, à l'inverse, prédominantes. Aussi ne s'attend-on pas, dans un texte technique, à trouver des traces évidentes des fonctions émotive et phatique, par exemple. En revanche, il n'est pas non plus concevable que de tels textes aient été rédigés sans tenir compte de leur destinataire supposé ; il n'est d'ailleurs pas rare qu'ils soient rédigés sur le mode impératif, qui est une des manifestations possibles de la fonction conative citées par Jakobson.

Pour Halliday aussi, les messages servent presque toujours plusieurs fonctions sociales[5] ; la fonction qu'il appelle (en traduction française) *informationnelle* (ou précédemment *représentationnelle*) ne se développe d'ailleurs qu'après les autres[6] (les fonctions *instrumentale*, *régulatrice*, *interactionnelle*, *personnelle*, *heuristique* et *imaginaire* chez l'enfant), mais prend une importance croissante dans la vie de l'individu :

> [...] les énoncés du langage adulte sont fonctionnellement complexes. Chaque acte linguistique adulte, mis à part quelques exceptions que l'on pourrait sans doute recenser, sert plus d'une fonction à la fois. (Halliday, 1974 : 64)

> Après un certain temps apparaît chez l'enfant l'utilisation du langage pour transmettre de nouveaux éléments d'information, c'est-à-dire communiquer un contenu (que le locuteur considère comme) inconnu de l'allocutaire. [...] Au cours de la maturation de l'individu cette fonction prend de plus en plus d'importance, et finit par dominer, sinon le langage de l'adulte, du moins la conception qu'il se fait de son utilisation. (Halliday, 1974 : 64-65)

> Quoiqu'il [sic] fasse avec le langage, l'adulte doit en exploiter les ressources idéationnelles, le potentiel permettant d'exprimer un contenu à partir de l'expérience du locuteur et de la communauté linguistique. Il y a des exceptions, des types d'énoncé comme *comment ça va ?* ou *c'est pas étonnant !* où le contenu idéationnel est non existant ; mais autrement il y a toujours un élément idéationnel dans les activités linguistiques de l'adulte. C'est sans doute la raison pour laquelle l'adulte tend à concevoir le langage en tant qu'instrument de communication. (Halliday, 1974 : 66)

Les fonctions du langage enfantin évoluent ensuite pour devenir chez l'adulte ce que Halliday nomme des « macro-fonctions » ; la macro-fonction consistant à utili-

5. ↑Toujours selon lui, les messages « monofonctionnels » seraient le propre du petit enfant, à un stade où les fonctions du langage sont toujours en développement.

6. ↑Ce qu'avait également mis en évidence Jakobson (1960 : 5), parlant de la fonction phatique : « [infants] are prone to communicate before being able to send or receive informative communication. »

ser le langage pour communiquer de l'information, omniprésente (ou presque), se nomme alors *idéationnelle* (« the content function of language, language as 'about something' » (HALLIDAY, 1978 : 113)). Il est toutefois intéressant de remarquer que, selon l'auteur, l'élément idéationnel trouve sans doute son origine, chez le très jeune enfant, dans la combinaison des fonctions personnelle et heuristique, bien plus que dans la fonction informationnelle, elle-même secondaire. Quoi qu'il en soit, l'auteur insiste sur le fait qu'il y a « une immense diversité fonctionnelle dans l'utilisation adulte du langage. [...] Ces 'macro-fonctions' reflètent à un niveau linguistique fort abstrait la multipicité [sic] des utilisations sociales du langage. » (HALLIDAY, 1974 : 65)

Ce que nous disent Jakobson et Halliday du caractère multifonctionnel de la communication en général reste également valable – certes souvent dans une moindre mesure – en situation de communication professionnelle. En effet, bien que dans ce cadre, la transmission efficace de l'information soit l'objectif premier des interlocuteurs et que, dès lors, les autres fonctions du langage soient généralement atténuées, ces dernières ne disparaissent cependant jamais complètement. Elles continuent d'ailleurs à jouer un rôle (surtout à l'oral non différé), les situations de communication professionnelle restant par nature des interactions sociales.

Du reste, la distinction entre une utilisation purement professionnelle et « utilitaire » de la langue et une communication que l'on pourrait qualifier de plus « banale » n'est sans doute pas aussi évidente à établir qu'il n'y paraît. Ainsi, par exemple, même dans le cadre hautement codifié des échanges entre pilotes et contrôleurs aériens (avec les enjeux matériels et humains que l'on sait), on constate l'emploi (facultatif et pourtant récurrent) de formules de politesse telles que des salutations (« hello », « bye ») ou des remerciements (« thank you », « thanks ») (LOPEZ, 2013 : 209). Ces salutations ne transmettent aucune information utile (la maxime de quantité de Grice ne s'applique donc pas à elles), mais permettent d'entamer ou de clôturer un échange, ce qui est à rapprocher des rôles de la fonction phatique telle que l'envisage JAKOBSON (1960 : 5) : « messages primarily serving to establish, to prolong, or to discontinue communication [...] ». Plus généralement, ces marques de courtoisie (qui, dans la communication quotidienne, sont pour la plupart des routines parsemant les échanges) relèvent de ce que Kerbrat-Orecchioni, aménageant le modèle de Brown et Levinson, appelle les *Face Flattering Acts* : ces « actes susceptibles d'avoir sur les faces des effets positifs » (KERBRAT-ORECCHIONI, 2000 : 24) contribuent en fait à maintenir l'harmonie de la conversation (ce qui est évidemment dans l'intérêt de tous les interlocuteurs), voire, dans certains cas, participent d'une stratégie plus ou moins consciente d'un des locuteurs pour obtenir plus facilement de l'autre ce qu'il désire qu'il fasse ou lui accorde.

Si la langue naturelle (en vertu de son expressivité et de sa simplicité apparente pour ceux qui la maîtrisent) s'impose comme premier moyen de transmettre de l'information, il faut néanmoins garder à l'esprit les risques qui y sont associés.

Ces risques sont l'une des raisons qui peuvent expliquer l'émergence de normes visant à les limiter.

1.3 LES NORMES LANGAGIÈRES

À première vue, l'utilisation individuelle de la langue semble être placée sous le signe de la liberté et de la créativité, comme en atteste la possibilité de produire des énoncés inédits. Mais, dès lors qu'on la considère comme un objet social, il devient évident que la langue n'est pas exempte de contraintes, auxquelles les individus doivent se soumettre s'ils veulent être admis dans la communauté linguistique[7].

De telles contraintes pesant sur l'utilisation de la langue sont bien connues des (socio)linguistes et sont traditionnellement appelées *normes*. Il faut en fait distinguer (au moins) deux sens relatifs au mot *norme* tel qu'il est utilisé en linguistique[8], qui est donc lui aussi polysémique.

1.3.1 *Normes subjectives et normes objectives*

L'acception la plus ancienne (de laquelle est dérivée l'adjectif *normatif*) renvoie à un ensemble de règles (c'est-à-dire des interdictions et des prescriptions) sur les façons dont la langue doit (ou ne doit pas) être utilisée pour se conformer à un idéal (théorique) ; autrement dit, à « une pratique de prescription des comportements langagiers » (SIOUFFI et STEUCKARDT, 2007 : VII). Ce type de norme – qualifiée en cela de *prescriptive* – trouve bien entendu son origine dans la longue tradition grammaticale, dont on recense des attestations très anciennes[9]. Elle vise avant tout à limiter la variation linguistique (c'est-à-dire l'existence simultanée de formes concurrentes, de plusieurs signifiants pour un même signifié) en imposant une variété unique (la *langue standard*, ou plus simplement « la norme »), choisie sur la base de critères externes à la langue (très souvent politiques ou économiques) au détriment des autres et ressentie par les membres de la communauté linguistique concernée comme étant la plus prestigieuse. Elle s'accompagne généralement de considérations axiologiques, que celles-ci soient d'ordre esthétique (certaines formes seraient plus élégantes que d'autres[10]) ou moral (comme le laisse à penser, par exemple, le mot « faute »). Pour ces raisons, les normes de ce type sont également appelées *normes subjectives*. Le poids de la norme subjec-

7. ↑« [...] speech community: any human aggregate characterized by regular and frequent interaction by means of a shared body of verbal signs and set off from similar aggregates by significant differences in language usage » (GUMPERZ, 2009 : 66).

8. ↑Le concept de « norme » ne se limite évidemment pas à la (socio)linguistique, puisque l'on parle aussi de « normes sociales » (FEHR et FISCHBACHER, 2004), de « normes techniques » (BENEZECH, 1996), etc.

9. ↑Notamment la grammaire de Panini pour le sanskrit, que l'on situe vers le V^e siècle avant notre ère (EMENEAU, 1955), ou les nombreux manuels de grammaire rédigés durant l'antiquité gréco-latine.

10. ↑Un exemple fréquemment cité est la préposition « à » dans l'expression « aller *au* coiffeur », jugée populaire et à laquelle le « bon usage » préfère « aller *chez* le coiffeur ».

tive peut être plus ou moins important selon les communautés linguistiques (les langues écrites qui sont enseignées à l'école y sont, pour des raisons évidentes, particulièrement sensibles) ; il est indéniable s'agissant du français (pour lequel l'Académie française et son dictionnaire font autorité depuis le XVII^e siècle), comme en témoigne entre autres le nombre important de travaux publiés sur cette question (AUROUX, 1998 ; CALVET et MOREAU, 1998 ; GADET, 2007, parmi d'autres). Pour Labov, la norme subjective est même indissociable de la notion de communauté linguistique :

> Il serait faux de concevoir la communauté linguistique comme un ensemble de locuteurs employant les mêmes formes. On la décrit mieux comme étant un groupe qui partage les mêmes normes quant à la langue. (LABOV, 1976 : 228)

L'un des effets pervers de la norme subjective est le sentiment d'insécurité linguistique qui peut naître chez les locuteurs qui ont conscience de ne pas la maîtriser, et plus largement la stigmatisation de ceux qui ne la maîtrisent pas. Son principal intérêt pour les travaux en linguistique est qu'elle permet, par commodité, d'offrir une vision relativement homogène (mais certainement réductrice) de la langue et des règles qui la régissent.

La seconde acception (à laquelle se rapporte cette fois l'adjectif *normal*) apparaît plus tardivement, sous la forme d'un emprunt à l'anglais *norm*, employé pour la première fois dans ce sens en sociologie dans la seconde moitié du XIX^e siècle (SIOUFFI et STEUCKARDT, 2007). Parce qu'elle se base non plus sur un jugement, mais sur l'observation (autrement dit, sur les fréquences d'utilisation), cette norme est dite *objective* ou *descriptive* (et parfois *observable*) ; est dès lors considéré « normal » ce qui est le plus habituel, le plus régulier.

L'ambiguïté qui peut exister autour du terme norme est bien connue des sociolinguistes [11] :

> Avant toute tentative pour définir la « norme », la moindre considération lexicologique découvre derrière le terme deux concepts, l'un relevant de l'*observation*, l'autre de l'élaboration d'un système de *valeurs*, l'un correspondant à une situation *objective* et *statistique*, l'autre à un faisceau d'intentions *subjectives*. Le même mot, utilisé sans précaution, correspond à la fois à l'idée de moyenne, de fréquence, de tendance généralement et habituellement réalisée, et à celle de conformité à une règle, de jugement de valeur, de finalité assignée. (REY, 1972 : 5 ; c'est nous qui soulignons)

La linguistique et la philosophie du langage s'intéressent depuis longtemps à la distinction entre la langue vue comme un système abstrait et homogène, d'une

11. [†]Certains vont jusqu'à distinguer entre cinq types de normes (plus ou moins fortement liées entre elles) : objectives, descriptives, prescriptives, subjectives et fantasmées (BULOT et BLANCHET, 2013).

part, et les réalisations variables et potentiellement infinies qu'elle permet (et desquelles relève ce que l'on nomme « usage »), d'autre part. Cette interrogation passe le plus souvent par une réflexion sur les concepts de règle et de norme (qu'ils soient ou non désignés explicitement comme tels). L'opposition saussurienne langue *vs* parole (Saussure, 1995) sera ainsi élargie par Hjelmslev puis Coseriu, qui utilisent respectivement les termes *schéma*, *usage* et *norme* (cette dernière étant définie comme une « fiction », « le concept derrière les faits rencontrés dans l'usage » (Hjelmslev, 1942 : 80, cité par Rey, 1972 : 9)) pour l'un, et *système*, *norme* et *parole* (*habla*) (Coseriu, 1952), pour l'autre. Elle est également redéfinie peu ou prou par Chomsky (1965), qui oppose pour sa part *compétence* et *performance* (en accordant toutefois un rôle beaucoup plus important au « sujet individuel », là où Saussure voyait avant tout la langue comme collective (Chevrot, 2012)).

Les sociolinguistes, quant à eux, se sont intéressés aux liens qui unissent la norme objective et la norme subjective (qui, dans les faits, peuvent ou non coïncider) ; il est probable que les deux s'influencent réciproquement, la norme subjective s'appuyant partiellement sur la norme objective (les formes les plus fréquentes sont les plus susceptibles d'être promues par le bon usage) (Gadet, 2007) et tendant également à la définir (les formes valorisées deviennent de fait plus fréquentes).

Quoi qu'il en soit, épistémologiquement, il semble évidemment plus acceptable (ou tout du moins plus facile), pour les linguistes, de tenir un discours qui s'appuie sur la norme objective, l'objectivité étant revendiquée par les disciplines scientifiques – ce qui permet d'ailleurs parfois d'évacuer un peu rapidement la question de la variation linguistique. C'est pourquoi la linguistique se présente volontiers comme descriptive, là où la grammaire traditionnelle est le plus souvent prescriptive. Ce principe semble d'autant plus évident que la langue est l'objet d'étude de la linguistique et que par conséquent, celle-ci doit décrire celle-là, et non la modifier ou l'influencer – en théorie, du moins [12].

Nous pensons néanmoins que la norme linguistique (au sens premier) n'est pas à bannir aussi catégoriquement et qu'au contraire, une certaine forme de norme (c'est-à-dire de contrainte) peut se révéler utile si elle est consciencieusement définie et appliquée. Il ne s'agit naturellement pas ici de verser dans le discours subjectif et de prétendre que certaines formes seraient intrinsèquement meilleures que d'autres, mais plutôt d'identifier et de déconseiller celles qui sont susceptibles de desservir les interlocuteurs par rapport à l'objectif qu'ils se sont fixé ou, à l'inverse, de recommander celles qui sont considérées comme les plus efficaces à cette fin.

En effet, comme on l'a vu, la langue naturelle est loin d'obéir au principe d'univocité ; au contraire, non seulement il existe une infinité théorique de façons de

12. [↑] Auroux (1998) soutient pour sa part l'idée que la linguistique, quoiqu'elle s'en défende, est nécessairement normative. Il faut admettre, en effet, que la langue que décrivent les linguistes est (en partie) une construction, fondée sur des usages (idéalement) attestés.

communiquer une même information, mais surtout un même énoncé peut généralement être interprété de plusieurs façons différentes, ce qui n'est pas sans entraîner des risques pour la compréhension. Or, il est des cas (particulièrement dans des contextes professionnels, où d'importantes sommes d'argent, voire des vies humaines peuvent être en jeu) où une communication efficace et sûre est primordiale (et où ces critères doivent primer sur tous les autres), tant une mauvaise compréhension pourrait avoir des conséquences fâcheuses.

Compte tenu des risques encourus, normaliser la langue apparaît donc parfois souhaitable, voire nécessaire. Nous pensons à ce titre que ce sont les linguistes, de par leurs connaissances métalinguistiques, qui sont les plus à même de déterminer où une telle norme serait nécessaire et quelle forme elle devrait prendre pour être aussi adaptée et efficace que possible par rapport à l'objectif poursuivi. Naturellement, et comme nous le suggérions en introduction de cette première partie, cette démarche doit se faire en concertation avec les porteurs du besoin, qui sont aussi les utilisateurs de cette langue spécialisée et qui en ont donc une connaissance (au minimum) intuitive.

Plus concrètement, pour en revenir à notre demande initiale, nous voudrions faire en sorte que les exigences techniques rédigées au CNES soient les plus claires possible (autrement dit, que l'interprétation qui en est faite par les ingénieurs en charge de les lire et de les respecter concorde avec ce que le rédacteur souhaitait exprimer). Pour atteindre un tel objectif, il semble par exemple nécessaire de proscrire les tournures les plus susceptibles de générer de l'ambiguïté et, idéalement, de recommander les structures que nous considérons comme les plus adéquates dans ce type d'écrits techniques.

Nous cherchons donc à définir une norme qui permette de limiter (sinon d'éliminer) les risques inhérents à l'utilisation de la langue naturelle dans un contexte spécifique, contrôlé et bien délimité. Car il serait évidemment illusoire de vouloir créer et appliquer une norme valable en toutes circonstances ; vouloir contraindre la communication quotidienne serait sans aucun doute voué à l'échec, l'usage s'écartant invariablement de la norme. Nous pensons logiquement que pour pouvoir fonctionner, la norme ne doit pas être perçue comme trop contraignante par ceux qui sont tenus de la respecter et qu'elle ne doit pas trop s'éloigner de leurs usages, au risque d'être rapidement délaissée, comme c'est très souvent le cas, quel que soit le domaine, lorsque le poids de la norme (dont relèvent notamment les langues contrôlées, que nous envisagerons plus loin) se fait trop important.

Dans tous les cas, la norme que nous visons devrait rester suffisamment souple, parce qu'il est impossible de prévoir de façon certaine tous les usages qui seront faits de la langue, qui, on le sait, évolue sans cesse. Le but n'est pas de brider le rédacteur ou d'en dévaloriser la production, mais de l'assister dans sa tâche et d'attirer son attention sur certains cas potentiellement problématiques qu'il serait

préférable d'éviter [13]. Idéalement, donc, une norme de ce type ne devrait pas être vue comme un carcan ou être source d'insécurité linguistique chez le rédacteur ; on peut même accessoirement espérer, si ce dernier parvient à se l'approprier, qu'elle développe sa conscience métalinguistique.

Nous pensons qu'il faut être conscient tant des avantages que des risques réels et des limites qu'impose la langue naturelle lorsqu'il s'agit de transmettre de l'information importante. Son expressivité illimitée lui permet une grande flexibilité (et la rend même dans certains cas indispensable), mais en fait par là même un outil d'une extrême complexité. De même, en raison de son fonctionnement particulier, reposant en grande partie sur le contexte et l'implicite, elle permet des économies cognitives bienvenues, mais elle peut aussi être source de confusions, parfois même à l'insu de ses locuteurs. À l'opposé, certains langages plus formels (tels que la logique des propositions, la logique des prédicats ou des langages basés sur la modélisation graphique, par exemple) sont à la fois beaucoup plus simples à définir et beaucoup plus précis, ne laissant que peu, voire pas de place à l'interprétation. En revanche, ils ne permettent pas toutes les nuances et modalités qu'autorise la langue et qui sont souvent nécessaires en contexte réel [14]. De ce fait, nous croyons qu'une langue normalisée représente un compromis intéressant, parce qu'elle ne demande qu'un apprentissage moindre à ceux qui maîtrisent déjà la langue naturelle et surtout parce qu'elle conserve l'expressivité de cette dernière, tout en réduisant les risques qu'elle peut occasionner. Les normes et leur utilité (pour l'ergonomie ou les facteurs humains, notamment (Karwowski, 2005)) étant d'ailleurs déjà largement reconnues en milieu professionnel, il semble tout à fait concevable que la langue y soit elle aussi soumise (Condamines, Lopez et Warnier, 2017).

1.3.2 *Normes spontanées et normes imposées*

Nous avons déjà rappelé l'opposition entre normes prescriptives et normes descriptives. Une autre façon présenter cette opposition est de distinguer les normes émergeant de façon spontanée au sein d'une communauté de locuteurs de celles qui sont imposées par une autorité ; différence qui est bien illustrée par les concepts de *normaison* et *normalisation* proposés par la socioterminologie, une approche de la terminologie que l'on doit aux sociolinguistes de l'école de Rouen et qui, à partir des années 1980, a remis en cause plusieurs des principes de la théorie générale de la terminologie.

13. [↑]Condamines, dans un article à paraître, montre en quoi le concept d'utilisabilité (déjà central en ergonomie) est adapté à ce qu'elle nomme la « linguistique ergonomique ». Notre démarche s'inscrit dans ce sens, puisque, dans un premier temps, nous nous intéresserons aux productions réelles des ingénieurs du CNES pour proposer notre norme langagière et que, dans un second temps, nous évaluerons nos propositions par le biais d'une enquête d'acceptabilité auprès de ces mêmes utilisateurs, dans le but d'offrir à cette norme une utilisabilité maximale.

14. [↑]Remarquons par ailleurs que dans les trois cas que nous citons, il est nécessaire de recourir à la langue naturelle pour définir les propositions, les prédicats ou certains éléments du modèle.

La normalisation terminologique est donc pour ce courant le fait de choisir et de favoriser de façon consciente et planifiée certaines désignations au détriment de dénominations concurrentes (ce qui relève de ce fait de l'aménagement linguistique, que les socioterminologues préfèrent pour leur part nommer *glottopolitique*) – de la même façon que la normalisation technique vise à harmoniser les pratiques dans un secteur par l'établissement d'un référentiel commun (une « norme » ou un « standard » [15]). Une politique linguistique, pour être respectée, a toutefois tout intérêt à être en harmonie avec les pratiques des locuteurs concernés : c'est avec cet objectif que des sociolinguistes de terrain ont été sollicités, notamment sous l'impulsion des autorités québécoises et catalanes (GAUDIN, 2005), et que le concept de normaison prend tout son sens. Comme son nom le laisse entendre, la normaison se définit elle aussi comme un processus résultant en la constitution d'une norme, mais, à la différence de la normalisation, celui-ci est spontané et non délibéré. Loin d'être marginal, ce phénomène est au contraire essentiel à l'établissement d'un système langagier permettant l'intercompréhension en discours :

> Elles [= les terminologies] connaissent des réalisations, notamment à l'oral, qui sont le lieu de créations ; ce sont les usages professionnels du laboratoire, de l'atelier, de l'usine qui les suscitent (*cf.* Candel, 1993). La genèse de ces formes a lieu dans les pratiques langagières et leur stabilisation leur confère le statut de normes de discours permettant l'intercompréhension. Et c'est ici que l'on peut distinguer deux types de procès aboutissant à la construction d'une norme, la *normalisation* et la *normaison* : « En résumé, on peut dire que la normalisation, c'est le processus qui vise à la construction consciente d'une norme unifiée, et la normaison, le processus responsable de la logique même de tout système linguistique » (Guespin, 1993 : 218). Qu'elle soit nationale ou internationale, politique ou technique, la normalisation émane toujours d'une institution qui fixe les termes recommandés ou obligatoires. En revanche, la normaison relève de ce que Teresa Cabré décrit comme « un processus au moyen duquel un système terminologique déterminé s'autorégule en accord avec ses utilisateurs » (1998 : 244). (GAUDIN, 2003 : 179)

Pour HUMBLEY (2001 : 112), « cette distinction permet de mieux appréhender la façon de régler sens et dénomination dans des contextes sociaux moins formalisés que ceux de la normalisation de type ISO ou des CMT [Commissions ministérielles de terminologie]. » Il n'empêche que la normalisation ne devrait idéalement pas s'opposer à la normaison, mais bien en tenir compte :

> La normaison est le processus qui conduit à ce qu'une langue et, pour ce qui nous occupe, les vocabulaires, sont en état d'équilibre et de renouvellement permanent du fait de la multitude des usages qui traversent la langue. Il

15. [↑] Lorsqu'il s'agit en particulier pour les autorités linguistiques de promouvoir, voire d'imposer sur un territoire donné une variété de langue considérée comme plus prestigieuse (la fameuse « langue standard »), on parle souvent de « standardisation ». Certains auteurs ont par ailleurs proposé de distinguer la normalisation au sens large (« toute procédure qui vise à fixer des termes ») de la standardisation à proprement parler (qui serait, elle, une « imposition *de fait* ») (DEPECKER, 1996), mais l'usage ne s'y est pas conformé et il est sans doute plus simple d'y voir des quasi-synonymes.

> s'agit d'un processus spontané et collectif. Spontané (ce que n'est pas la normalisation) et collectif, ce que la normalisation n'est pas non plus. Et à mon sens, [...] toute normalisation qui ne s'appuie pas sur la normaison, est une entreprise brutale et presque entièrement vouée à l'échec. Il convient en effet de s'appuyer sur l'existant pour décider, mais encore faut-il avoir un tableau de l'existant. (Depecker, 1996 : 87)

En effet, pour que les locuteurs (à qui revient le choix final d'utiliser ou non un terme) acceptent librement de suivre les recommandations qui leur sont faites, il est nécessaire que celles-ci soient réalistes et, donc, que les usages réels soient pris en considération pour les établir :

> La réalisation d'une politique linguistique nécessite des choix d'aménagement linguistique qui ne peuvent rencontrer le succès qu'à la condition d'être compatibles avec les opinions des locuteurs et donc avec les sentiments et les pratiques linguistiques. C'est pourquoi la négociation terminologique, l'information et la consultation des acteurs concernés sont des facteurs favorables pour que les décisions soient suivies d'effets. Il faut pour cela que les conditions d'une adhésion sociale soient réunies et, donc, que les décisions prises reposent sur une description fine des pratiques et une consultation préalable des usagers. [...] En fait, on peut affirmer que la **normaison** devrait être la priorité des organismes de politique linguistique. (Gaudin, 2005 : 85-86)

Nous adhérons pleinement à ce point de vue et pensons que ce qui est dit à propos des terminologies et des politiques linguistiques est tout aussi vrai dans l'optique de proposer des recommandations langagières (qui peuvent inclure, mais ne se limitent pas à un vocabulaire contrôlé) dans un contexte professionnel. En fait, notre démarche correspond exactement à celle qui devrait, selon Gaudin, être mise en œuvre dans le cadre d'une glottopolitique :

> Le linguiste, le terminologue ou le « langagier » engagé dans une démarche glottopolitique doit décrire une situation langagière particulière, établir un diagnostic et proposer des solutions. [...] Ces tâches de description relèvent de la recherche ; elles sont rendues nécessaires pour comprendre les changements linguistiques en cours, que ceux-ci soient spontanés ou planifiés, et pour agir. (Gaudin, 2005 : 85)

Nous pouvons donc désormais préciser que ce que nous cherchons à faire en tant que linguiste devant répondre à un besoin de contrôle des productions langagières, c'est proposer une normalisation de la rédaction des exigences au CNES basée sur la normaison constatée ; autrement dit, dans un premier temps, observer les *usages* réels et recueillir le sentiment des rédacteurs, puis, dans un second temps, proposer des recommandations fondées sur ces usages. Il s'agit, pour nous, du meilleur moyen de nous assurer que ces recommandations aient de meilleures chances d'être réellement suivies par celles et ceux à qui elles se destinent. Nous nous servirons d'un corpus d'exigences authentiques comme moyen privilégié

afin d'observer la norme spontanée résultant de la normaison [16].

Dans les sections qui vont suivre, nous présenterons rapidement les concepts de genre textuel, sous-langage, langue de spécialité et communauté de discours, que nous rapprochons des normes spontanées (ou normaisons), puis celui de langue contrôlée, que nous assimilons aux normes imposées (ou normalisations).

1.4 GENRE TEXTUEL ET SOUS-LANGAGE

Nous venons de le voir, la définition de normaison met en exergue que lorsque des locuteurs sont amenés à communiquer entre eux de façon régulière, ils régulent inconsciemment le système langagier utilisé et génèrent des régularités spontanées – ce qui permet de marquer l'appartenance à une communauté linguistique par une forme de connivence tout en facilitant les échanges. Cette idée sous-tend aussi plusieurs autres concepts issus de traditions différentes, en particulier celui de *genre (textuel)*.

Des études sur corpus toujours plus nombreuses s'appuient (explicitement ou non) sur cette notion de genre textuel, que ce soit pour caractériser finement un genre en particulier – GLEDHILL (2000) pour les articles scientifiques, en particulier dans le domaine pharmaceutique ; BEAUVISAGE (2001) pour le roman policier ; POUDAT et RINCK (2006) pour l'article scientifique de linguistique – ou classiquement pour comparer des usages selon les genres – BIBER (1988 ; 1993) ; MALRIEU et RASTIER (2000), qui utilisent plus de 200 variables morphosyntaxiques ; et ROLAND, DICK et ELMAN (2007). Les spécificités de certains genres sont aussi exploitées dans des études de TAL : ANDERSSON et al. (2014) s'intéressent par exemple à l'extraction automatique d'hyponymes dans les brevets. En raison à la fois de leur importance et de leur nature, les exigences nous semblent particulièrement bien se prêter à de telles études. En outre, l'identification de textes appartenant à des genres distincts fut une étape fondamentale pour la constitution des principaux corpus de référence, qui, par définition, se devaient d'être suffisamment larges et variés pour pouvoir être considérés comme représentatifs (essentiellement de la langue écrite pour les plus anciens, pour des raisons pratiques évidentes) [17].

Pourtant, en dépit de son importance (soulignée notamment par MALRIEU et RASTIER, 2000), le terme « genre » est régulièrement employé sans que le concept ne soit toujours rigoureusement défini (par exemple dans ROLAND, DICK et ELMAN, 2007), sans doute en raison de la multiplication de ces études (qui semble légitimer ce type d'approche) ou car la distinction entre genres apparaît (trompeusement)

16. ↑Il n'y a rien d'étonnant à ce que GAUDIN (2005 : 89) parle de « convergence entre socioterminologie et linguistique de corpus ».

17. ↑Le manuel du *Brown Corpus*, célèbre corpus d'anglais américain contemporain compilé dans les années 1960 qui fut à l'origine de plusieurs autres corpus anglophones de référence, parle ainsi de « catégories » de textes, dont les quinze principales sont nommées « genres » (« Genres [...] are the fifteen mayor categories of samples », http://www.hit.uib.no/icame/brown/bcm.html). Le *British National Corpus* (BNC, années 1980-1990) a quant lui été constitué pour représenter plusieurs « domaines » et « médiums ».

comme allant de soi. Si la notion de genre de texte semble en effet assez simple à percevoir intuitivement (ce qui pourrait expliquer pourquoi, à l'instar des grands genres littéraires, ils ne sont que rarement remis en question), elle est en réalité bien plus complexe à définir et délimiter avec précision.

Une des difficultés tient bien sûr à ce que les frontières des genres sont floues : il peut y avoir des chevauchements, des genres hybrides, des catégories plus larges que d'autres. Mais, du point de vue terminologique aussi, la notion de genre en recoupe parfois d'autres, notamment celles de *registres* (ainsi Biber (1993) classe des registres plus ou moins vastes (conversation, fiction, prose académique, mais aussi éditoriaux de journaux, lettres professionnelles, etc.) selon cinq dimensions) ou encore de *types de textes*.

Lee (2002) s'est efforcé de clarifier la distinction entre les concepts de genre, registre et type de texte (auxquels il ajoute domaine et style) tels qu'ils sont mobilisés dans la littérature. Selon lui, le genre se distingue essentiellement du type de texte en ce qu'il est défini sur la base de critères externes (non linguistiques), contrairement à ce dernier [18], identifié grâce à ses caractéristiques internes (c'est-à-dire strictement linguistiques) :

> A *genre*, in this view, is defined as a category assigned on the basis of **external** criteria such as intended audience, purpose, and activity type, that is, it refers to a conventional, culturally recognised grouping of texts based on properties other than lexical or grammatical (co-)occurrence features, which are, instead, the **internal** (linguistic) criteria forming the basis of *text type* categories.

Cette conception semble en phase avec celle de Biber, qui indique :

> Genre categories are determined on the basis of external criteria relating to the speaker's purpose and topic; they are assigned on the basis of use rather than on the basis of form. (Biber, 1988 : 170)

Elle rejoint également celle de Swales (2016) et de Bhatia à sa suite, pour qui, on le verra, le genre (textuel) est indissociable d'une communauté (professionnelle ou académique) aux objectifs bien définis :

> A textual genre is a recognizable communicative event characterized by a set of communicative purpose(s) identified and mutually understood by the members of the professional or academic community in which it regularly occurs. (Bhatia, 1993 : 13)

Selon cette définition, il ne fait aucun doute que les exigences rédigées en langue naturelle relèvent d'un genre (qui est défini *a priori* sans avoir à se soucier des

18. ↑D'autres auteurs encore (Paltridge, 1996) distinguent le type de texte selon des motifs rhétoriques : description, procédure, etc.

formes linguistiques), puisqu'il s'agit bien d'un événement communicatif récurrent (répété pour chaque projet) au sein d'une communauté professionnelle clairement délimitée (les ingénieurs en charge de les écrire et ceux en charge de les respecter) et dont la nature et le but sont connus (spécifier un système conformément aux attentes des parties prenantes). Il s'agit d'une communication spécialisée, d'experts à experts :

> It is assumed that author and reader share a common language and that when certain words or phrases are used, each understands what is meant [...]. Writer and reader, or speaker and hearer, are assumed to have the same level of expertise [...]. This expert to expert communication context is likely to be the one with the highest density of terms. (Pearson, 1998, citée par KÜBLER et FRÉROT (2003 : 430))

L'idée de l'existence de ce genre particulier (différent par exemple d'autres textes techniques) est en outre confortée par l'existence même d'une désignation idoine, « exigences ». Lors de l'analyse du corpus, nous chercherons à vérifier cette hypothèse et à en identifier les traits caractéristiques (sa grammaire).

La distinction entre genre et registre, en revanche, semble moins aisée. Ainsi, les définitions données par Ferguson, citées par LEE (2002 : 43), sont étonnamment proches : « communicative situation that recurs regularly in a society » (registre) et « message type that recurs regularly in a community » (genre) ; Ferguson donne par ailleurs les recettes de cuisine comme exemple à la fois de registre et de genre. Chez certains auteurs, la notion de registre englobe celle de genre, tandis que chez d'autres, c'est l'inverse. Lee conclut de ces observations :

> [...] there is a superficial terminological difference in the way genre is used by some theorists, but no real, substantive disagreement because they both situate it within the broader context of situational and social structure. (LEE, 2002 : 45)

Nous opterons pour notre part pour le terme « genre », qui nous semble bien indiqué et a l'avantage de ne pas créer de confusion avec la notion de « registres de langue », fréquemment utilisée en français, mais sans lien direct avec le phénomène qui nous occupe ici. Il faut toutefois garder à l'esprit que chaque auteur peut avoir une définition légèrement différente du genre, et que certains nommeront « registre » ce que d'autres considéreront être un genre, ou inversement. Retenons surtout de ces définitions du genre qu'une situation de communication récurrente génère des régularités langagières (lexicales et/ou morphosyntaxiques) et qu'en retour, ces régularités peuvent permettre d'identifier la situation communicative concernée.

Quoi qu'il en soit, il est évident qu'au-delà des contraintes qui peuvent peser plus ou moins fortement sur un genre (ou du moins des habitudes qui sont rattachées à ce dernier), il y a nécessairement place – comme dans toute production

langagière – pour la variation. Ainsi que le précisent justement Poudat et Rinck dans leur étude sur l'article scientifique de linguistique :

> En tant que principe de structuration et instance de normalisation des pratiques langagières, le genre se caractérise par des propriétés linguistiques régulières et conventionnelles, qui ne doivent cependant pas occulter la diversité des textes. Aussi, si l'article de recherche est généralement envisagé comme un genre contraint, fortement normé voire routinisé, il se définit comme tout genre du double point de vue de sa *stabilité* et de sa *variation intrinsèque* (Poudat et Rinck, 2006 : 781 ; c'est nous qui soulignons)

Dans la lignée de Rastier, elles rattachent ces différences individuelles à ce qu'elles nomment *style* (propre à un auteur). Le style est donc, de fait, un « usage singulier du genre », plus ou moins contraint (voire conditionné) par ce dernier. Guilbert, s'appuyant principalement sur les travaux de Bakhtine et de Charaudeau [19], rappelle d'ailleurs que

> [...] si Bakhtine ne nie pas la notion de style individuel, il précise que le style est fonction du genre du discours dans lequel il est tenu. Le genre a une "valeur normative" et impose les "formes prescriptives de l'énoncé" [20] (Bakhtine, 1984). Par ailleurs, le style est plus ou moins individuel et plus ou moins effectif selon les genres ; on admettra sans peine que tel genre littéraire est plus propice au style individuel que le genre "lettre officielle". (Guilbert, 2014 : 5)

Notre travail porte logiquement sur les régularités – et donc sur le genre –, sans vraiment tenir compte des styles. D'une part parce que ceux-ci sont justement trop particuliers (nos conclusions, pour être utiles, doivent idéalement être aussi peu dépendantes que possible des rédacteurs individuels), et, d'autre part, si même nous souhaitions nous y essayer, parce qu'il ne nous est pas possible de rattacher chaque exigence à un rédacteur en particulier (à supposer qu'ils ne soient pas plusieurs à avoir participé à sa rédaction ou à sa révision, ce qui n'est pas certain), puisque nous ne disposons tout simplement pas des informations nécessaires pour cela (les contributeurs étant listés par spécification, et non par exigence). Si nous ne nions évidemment pas le rôle joué par chaque rédacteur individuellement (même dans un genre que l'on peut considérer comme encore beaucoup plus contraint formellement que d'autres), il faut souligner que le nombre de rédacteurs étant intervenus sur les exigences composant notre corpus est relativement élevé (une trentaine au total). Par ailleurs, on peut s'attendre à ce que les différences idiosyncrasiques les plus marquées et les plus flagrantes aient été partiellement « lissées » lors des phases de révision des spécifications.

19. ↑Pour qui « les genres [du discours] dépendent de la "nature communicationnelle" de l'échange verbal » (Charaudeau, 2002 : 279, cité par Guilbert, 2014 : 3).

20. ↑Ce qui légitime d'autant plus l'idée de se servir du genre comme d'une base pour une langue contrôlée (même si les prescriptions ne sont bien sûr pas forcément identiques).

Après avoir envisagé la notion de genre (et son intérêt pour notre travail), il nous faut maintenant évoquer celle de *sous-langage*, introduite par Harris :

> In *Mathematical Structures of Language* (p. 152) he [= Harris] states: "certain proper subsets of the sentences of a language may be closed under some or all of the operations defined in the language, and thus constitute a sublanguage of it." (Kittredge et Lehrberger, 1982 : 1)

Ainsi, cette notion nous vient originellement d'une perspective mathématique et distributionnaliste, là où le genre trouve son origine dans une approche plutôt littéraire et historique (Bakhtine) ou sociolinguistique (Swales, Bhatia), ou encore dans la linguistique de corpus (Biber). Ces origines diverses n'empêchent pas des similarités parfois frappantes entre genre et sous-langage, par exemple lorsque l'on examine cette définition de *sous-langage* :

> Because there is sometimes some dispute about the use of the term 'sublanguage', let us clarify from the start that for our purposes a sublanguage is an identifiable *genre* or *text-type* in a given subject field, with a relatively or even absolutely closed set of syntactic structures and vocabulary. (Somers, 1998 : 131 ; c'est nous qui soulignons)

Kittredge donne deux conditions nécessaires à l'émergence d'un sous-langage, qui se voient également vérifiées dans notre cas et ne sont évidemment pas sans rappeler celles du genre textuel :

> — a community of speakers (i.e. 'experts') shares some specialized knowledge about a restricted semantic domain;
> — the experts communicate about the restricted domain in a recurrent situation, or set of highly similar situations. (Kittredge, 2005)

Selon lui, la systématicité du sous-langage permet notamment une économie d'expression (l'anaphore zéro ou l'omission des articles définis seraient communes à des sous-langages très différents). Les propriétés connues des sous-langages incluent un lexique plus limité (mais contenant éventuellement des mots qui ne sont utilisés nulle part ailleurs dans la langue), une syntaxe elle aussi plus restreinte et « déviante », et des fréquences d'occurrences des mots et des motifs syntaxiques différentes par rapport à la langue en général : chaque sous-langage aurait ainsi son propre profil statistique.

Tout comme nous, Kittredge pense qu'une meilleure connaissance du fonctionnement des sous-langages est nécessaire pour concevoir des langues contrôlées plus adaptées aux domaines ou aux genres auxquels elles se destinent :

> [...] a good appreciation of how sublanguages work is a prerequisite for any attempt to engineer a standard controlled language for one or more domains. [...] [C]ontrolled languages (CLs) do not always have the tight single-domain restrictions one finds in true sublanguages. However, a CL

> standard is usually based on a specific type of text, which is used for one or more distinct, but similar, sublanguages. In other words, the motivation for setting up a controlled language usually comes from the need to introduce writing standards for an important family of preexisting sublanguages. [...] (KITTREDGE, 2005)

Nous partageons également son avis sur l'impossibilité de proposer une langue contrôlée « universelle » et sur la nécessité d'adapter les normes à ceux à qui elles sont destinées :

> One thing seems already clear – that the 'one-size-fits-all' mentality inherent in some CL standards is likely to be inadequate for the future. [...]
>
> One long-term solution [...] is to use a better understanding of sublanguage rationale in the design of controlled languages, so that good writing practice can be enforced, while keeping desirable sublanguage features that are needed for economy of expression and for maintaining nuances of meaning. (KITTREDGE, 2005)

Ces remarques rejoignent totalement notre objectif : prendre en compte les régularités spontanées (qu'on les nomme normaison, genre ou sous-langage) pour établir une norme langagière (normalisation ou langue contrôlée) en adéquation avec les usages (et les attentes) de ceux à qui elle s'adresse, à la fois efficace et suffisamment souple.

Enfin, il nous reste à évoquer deux autres concepts que l'on peut également vouloir rapprocher des normes spontanées associées à une connaissance particulière : celui de langue de spécialité (ou langue spécialisée) et celui de communauté de discours. Si toutes ces notions peuvent être comparées, c'est parce qu'elles partagent des points communs importants, même si certaines peuvent paraître plus adéquates en fonction de l'objectif que l'on se fixe :

> Dans notre quête de détermination du sous-ensemble à observer, nous avons préféré, à l'instar de Williams (1999), remplacer les notions de sous-langage, de langues de spécialité et de plurisystèmes, telles qu'elles sont définies par Lerat (1995, 11, 28) et Habert & *al.* (1997, 148-152), par la notion de communauté de discours, introduite par Swales (1990). (PECMAN, 2007 : 81)

Concernant la *langue spécialisée*, tout d'abord, Lerat dit d'elle qu'elle « est une langue naturelle considérée en tant que vecteur de connaissances spécialisées » (LERAT, 1995 : 20). Pour lui, « les langues spécialisées ne sont rien d'autre que des usages spécialisés des langues naturelles » [21] (LERAT, 1997 : 1) et « [l]eur spécificité est essentiellement lexico-syntaxique, du côté des formes, et lexico-conceptuelle, du côté des contenus » ; « [l]es enjeux sont principalement l'aide à la traduction,

21. †PECMAN (2007 : 80) précise : « Généralement, une langue de spécialité désigne un sous-système linguistique correspondant à l'emploi de la langue dans un domaine de connaissances unique ».

à la rédaction, à la documentation et à l'apprentissage. » Elles ont en effet rencontré un succès certain en didactique des langues (non sans susciter quelques débats sur la meilleure façon de les enseigner (CHARNOCK, 1999)). Toujours selon LERAT (1997 : 2), « Aucune théorie linguistique, quelle qu'elle soit, n'a jamais isolé le fonctionnement des langues spécialisées de celui des langues naturelles en général. » CHARNOCK (1999 : 2) précise en outre : « Les langues de spécialité semblent fonctionner non pas comme des langues autonomes, ayant chacune ses caractéristiques spécifiques, mais comme des fragments ou des sous-ensembles de la langue naturelle. » Ces « sous-ensembles » rappellent forcément les « subsets » des sous-langages harrissiens.

Quant au terme « communauté de discours »[22], dont la paternité est attribuée par SWALES (2016) à Nystrand (1982), il réfère d'après lui à

> a largely heterogeneous, socio-rhetorical assemblage of people who broadly share occupational or recreational experiences, goals, and interests. Thus, members of a DC may have different first languages, different religions, and belong to diverse ethnicities. Examples might include members of GERAS or of TESOL, all those who work in an animal clinic, or those who are members of a local choir. (SWALES, 2016 : 10)

On le voit, cette définition est assez large et englobe autant les communautés de discours locales (par exemple des collègues travaillant au même endroit, ou des professionnels exerçant la même activité dans la même ville) que celles que Swales nomme « focales » (par exemple les membres d'associations qui peuvent être internationales, partageant un même intérêt professionnel ou récréationnel). Lui-même (avec d'autres) admet volontiers qu'il est difficile de proposer une définition consensuelle de ce concept, qui reste malgré tout très commode, mais propose une série de critères définitoires (dont le nombre a pu varier légèrement avec le temps). Parmi ceux-ci, on retrouve la nécessité, pour la communauté de discours, de disposer d'objectifs partagés et reconnus, de mécanismes permettant la communication entre membres et d'une terminologie spécifique et, surtout, d'utiliser et de « posséder » un ou plusieurs genres. Ces genres apparaissent en réponse à des besoins socio-rhétoriques récurrents (liés à ses objectifs). Ils peuvent être partagés plus ou moins largement avec d'autres communautés de discours, mais une communauté de discours, pour exister en tant que telle, doit nécessairement se les approprier (c'est-à-dire les particulariser) à un certain degré, notamment en développant ses propres conventions. Le concept de communauté de discours est donc intimement lié à celui de genre, le second étant un attribut du premier.

22. [↑]En anglais : « discourse community » (DC), à ne pas confondre avec « speech community » (« communauté linguistique ») : « a community sharing rules for the conduct and interpretation of speech, and rules for the interpretation of at least one linguistic variety » (HYMES, 2005 : 6). Les définitions ne sont pas très éloignées, mais la communauté linguistique est réputée plus homogène que la communauté de discours.

1.5 LANGUES CONTRÔLÉES

Comme exposé en introduction, notre objectif, pour répondre à la demande du CNES, est de proposer une méthode de constitution d'une langue contrôlée pour la rédaction des exigences en français. Nous présenterons les grands principes de la rédaction des exigences dans le prochain chapitre et analyserons certaines prescriptions récurrentes des langues contrôlées, dont nous pouvons nous inspirer, dans le chapitre 3. Dans cette section, nous nous contenterons donc de dresser un bref historique des langues contrôlées, de leurs origines jusqu'à nos jours, et de mettre en évidence les principes communs aux langues contrôlées, qui peuvent être assez diverses. Rappelons que celles-ci constituent le versant prescriptif de notre travail.

1.5.1 *Introduction*

L'une des solutions visant à éliminer ou limiter les risques inhérents à la langue naturelle comme moyen de transmission de l'information (cf. section 1.2) les plus communément employées consiste à déconseiller ou interdire les éléments (mots et structures syntaxiques) qui sont les plus susceptibles de générer des problèmes de compréhension. De telles recommandations, lorsqu'elles sont regroupées en un ensemble cohérent de règles précisant quels énoncés sont autorisés et lesquels ne le sont pas (le plus souvent à l'écrit, plus rarement à l'oral), peuvent prendre le nom de *langue contrôlée* (Ryan, 2011), *langage contrôlé* (Da Silva, 2014) ou moins souvent de *langue naturelle contrôlée* (Cheminot, 1999) ou *langage naturel contrôlé*[23] (Sadoun, 2014) (soit *controlled language* (O'Brien, 2003) ou *controlled natural language* (Wyner et al., 2009) en anglais, termes qui ont eu une influence certaine sur leurs équivalents français). Par *langue contrôlée*, on entend donc un système de règles établi sur la base d'une langue qui lui préexiste et qu'il vise à contraindre ; règles qui sont le plus souvent compilées et expliquées dans un document qui s'apparente à un guide et qui doit permettre l'apprentissage de ladite langue contrôlée par ceux qui souhaitent l'utiliser. La maîtrise de cette langue contrôlée leur permet ensuite de produire ce que l'on pourrait appeler des « discours contrôlés ».

1.5.2 *Historique*

Les tentatives visant à simplifier la langue naturelle et à la rendre plus claire existent depuis très longtemps et vont toujours de pair avec la volonté d'être compris de ses interlocuteurs ; au cours de l'histoire, penseurs et auteurs se sont interrogés sur la meilleure façon de rendre leurs discours compréhensibles par ceux à qui ils sont adressés.

23. ↑Si l'association des adjectifs *naturel* et *contrôlé* peut sembler paradoxale (voire oxymorique), le premier est là pour rappeler qu'une telle langue tend malgré tout à rester proche d'une langue naturelle – précision d'autant plus importante en anglais que le terme *language* est polysémique –, tandis que le second permet d'insister sur les restrictions posées sur cette langue.

Ainsi, s'inspirant peut-être du philosophe grec Théophraste (IVe-IIIe siècles ACN) (KENNEDY, 1957), le célèbre homme d'état et auteur romain Cicéron (IIe-Ier siècles) établissait, dans son traité *De oratore*, trois « niveaux de style » (simple, moyen, élevé [24]) que tout orateur se doit de maîtriser et d'appliquer selon le sujet discuté et le but poursuivi. Le style dit « simple » – qui n'a en réalité rien d'aisé à mettre en œuvre – est employé afin d'informer ou d'enseigner et, à cet effet, il se caractérise avant tout par la recherche de la clarté. C'est pourquoi Cicéron le juge approprié notamment aux affaires civiles, ainsi que l'illustre en particulier son plaidoyer *Pro Caecina* : les phrases y sont courtes, les répétitions de mots plus fréquentes, l'ordre des mots varie peu et les néologismes sont proscrits (ALBRECHT, 2003 ; voir aussi DAVIES, 1970). Cette théorie des trois styles aura une influence certaine sur la tradition rhétorique dans l'Antiquité (et même au-delà). Plus tardivement, le rhétoricien grec Hermogène de Tarse (IIe-IIIe siècles PCN) recensera dans un traité en deux livres (*Du Style*) ce qu'il considère comme les sept formes idéales de style ; parmi celles-ci, la plus importante est la clarté [25], qui devrait selon lui être recherchée dans tous les discours :

> The thought of a clear sentence must be one that is familiar to most people. It must be expressed in a conversational manner, as a direct statement of fact, using clauses that are short and complete in themselves and simple structures consisting of a subject in the nominative case and a main verb. (WOOTEN, 1987 : xiii)

La question du style rédactionnel – évidemment centrale en littérature – se pose aussi avec acuité aux chercheurs et penseurs à l'époque moderne, alors qu'ils s'interrogent sur la meilleure façon de diffuser le savoir scientifique. Le philosophe et scientifique anglais Francis Bacon (XVIe-XVIIe siècles) est considéré par beaucoup comme le premier et principal promoteur d'un style « clair » [26] pour la rédaction scientifique, ce qui rejoint par ailleurs sa conception de la philosophie naturelle :

> R. F. Jones, in particular, has argued that Bacon was the early inspiration for the writing and thinking of later seventeenth-century scientists who advocated a style that should avoid metaphor, aim at purity, and provide an exact equation between world and word. The main features of the 'plain style' are a lack of ornate metaphors; simple syntactic patterns; a logical progression of sentence and meaning; and a preference for well-tried words rather than neologisms (AUGHTERSON, 2000 : 97-98)

Sans que l'on puisse parler de filiation, on constate déjà certaines récurrences dans ces recommandations – il ne s'agit pas encore vraiment de « règles » –, par exemple une mise en garde contre les néologismes, qui concernent les niveaux lexical, syntaxique et discursif.

24. ↑*Genus tenue*, *genus medium* et *genus sublime*, respectivement.
25. ↑*Sapheneia*.
26. ↑*Plain style* : simple, direct, dépourvu d'ornements ou de fioritures, facile à comprendre.

Au-delà du style, les réflexions portent sur la nature de la langue elle-même ainsi que sur la possibilité d'une langue dite universelle – à une époque où le latin perd de plus en plus son statut de langue de communication internationale en Europe –, qui pourrait être une langue construite de façon à être plus logique, rationnelle et régulière – idée déjà envisagée par Bacon (Knecht, 1981). C'est notamment ce dont rêve Descartes, puis ce que désire réaliser Leibniz à la fin du XVII^e siècle à travers ce que ce dernier nomme la « caractéristique universelle », à savoir une langue universelle et univoque qui permettrait d'exprimer efficacement tous les concepts philosophiques, mathématiques et scientifiques en général :

> Le projet, repris à Descartes, d'une langue utile à la philosophie, langue universelle allégée de tous les génies particuliers des langues naturelles, va dans le même sens. Il y a chez Leibniz une gradation qui va des langues particulières au latin pris comme langue universelle, et à cette langue à inventer qui en aurait épuré les risques d'équivoque. L'exigence est ici conforme au principe du meilleur : du plus économique, plus fécond, ainsi qu'au principe de continuité : la langue universelle serait une langue, tout en différant des autres par un degré de perfection. (Guéry, 1995 : 243-244)

Mais si l'idée d'une langue universelle, c'est-à-dire permettant une intercompréhension optimale, va finalement rencontrer un certain succès, c'est davantage grâce aux langues artificielles telles que le volapük de l'Allemand Schleyer et surtout l'espéranto du Polonais Zamenhof, toutes deux conçues à la fin du XIX^e siècle. S'inspirant pour leur part bien davantage du fonctionnement des langues naturelles (dont elles entendent néanmoins réduire fortement la complexité, par exemple en réduisant les risques d'homonymie et en proposant une grammaire plus régulière), et donc bien loin des langues formelles auxquelles s'intéressaient entre autres Leibniz ou Frege à sa suite (Bartlett, 1964), elles visent à acquérir le statut de langues auxiliaires (véhiculaires) internationales : ainsi, quoiqu'il ne parvienne pas à concurrencer directement l'anglais (incontournable dans les domaines politiques et économiques, notamment), l'espéranto compterait actuellement plusieurs centaines de milliers (voire plusieurs millions) de locuteurs à travers le monde (Wandel, 2015). Élaboré dans ce souci d'universalité et donc de neutralité, l'espéranto n'est basé sur aucune langue naturelle en particulier, bien que son lexique provienne en très grande partie de racines indo-européennes (Marlaud, 2013).

Pour voir apparaître ce que l'on nomme aujourd'hui langue contrôlée, il faut néanmoins attendre le début des années 1930 et la publication par le linguiste britannique Charles Kay Ogden de son ouvrage *Basic English : A General Introduction with Rules and Grammar*. Ainsi que l'indique son nom, le Basic[27] English est basé sur l'anglais, qu'il cherche à simplifier fortement : le noyau du vocabulaire se limite à 850 mots (dont une dizaine de verbes seulement), sélectionnés arbitrairement par Ogden parmi les plus courants (*come*, *yes*, *tomorrow*, *plane*, etc.) et auxquels s'ajoutent quelques autres centaines de mots pour les domaines plus spé-

27. ↑BASIC pour British – American – Scientific – International – Commercial.

cialisés, et la grammaire est contrainte par des règles simples qui permettent de former de nouveaux mots (par ajouts d'affixes : *un-* pour les adjectifs avec un sens négatif ou *-ly* pour les adverbes, par exemple, ou par composition : *milkman*), ou de former des phrases (sur ce point, le Basic English reste finalement assez proche de l'anglais standard). Le Basic English se veut donc être une langue facile à apprendre [28], et doit par la même occasion faciliter l'apprentissage de la langue anglaise (et ainsi en faire une langue de communication internationale, tout comme l'espéranto avant lui) :

> Basic English is an attempt to give to everyone a second, or international, language which will take as little of the learner's time as possible. (OGDEN : http://ogden.basic-english.org/be21.html)

Pour Ogden, le Basic English et ses 850 mots de base (qui doivent pouvoir être appris en un ou deux mois, estime-t-il, et qui, avec leurs formes fléchies ou dérivées, permettent selon lui d'exprimer autant de concepts que 20 000 mots de la langue anglaise (OGDEN, 1944)) ont pour vocation de faciliter la communication internationale (sans pour autant faire disparaître les autres langues), puisque chacun, au prix d'un effort réduit, disposerait ainsi d'une langue auxiliaire commune.

Bien que, pour ces raisons, le Basic English ait pu remporter un certain succès et trouver un écho notamment auprès de Winston Churchill, alors Premier Ministre du Royaume-Uni, et que l'idée d'utiliser un vocabulaire limité dans l'apprentissage de l'anglais langue seconde soit toujours d'actualité, il n'a pas été épargné par les critiques. FLESCH (1944 : 343), entre autres, lui reprochait de n'être « ni basique, ni de l'anglais » (« neither Basic nor English »), faisant en particulier remarquer sarcastiquement que des mots aussi courants que *tomato*, *chair* ou encore *lamp* sont absents de son lexique ; ces lacunes obligent dès lors le locuteur à user de périphrases qui peuvent rendre particulièrement longues et complexes des expressions pourtant simples en anglais standard (« my cousin » devenant « my father's sister's daughter »). Lui aussi convaincu de l'intérêt d'un anglais simplifié, il regrettait qu'Ogden n'ait pas suivi une approche scientifique pour élaborer son Basic English (et particulièrement pour en définir le vocabulaire de base, qu'il juge beaucoup trop restrictif [29]) :

28. ↑Étant basé sur l'anglais, le Basic English est forcément confronté à ses irrégularités. Ainsi, par exemple, les comparatifs et superlatifs doivent en théorie se former en plaçant « more » ou « most » devant l'adjectif, mais il reste des exceptions (« good/better/best » et « bad/worse/worst » sont les deux seules séries qui sont données par Ogden) et l'apprenant doit « être préparé » à rencontrer des formes en *–er* et *–est*.

29. ↑En comparaison, l'initiative du *français fondamental* (lancée en 1954, initialement sous le nom *français élémentaire*), apparaît comme bien plus rigoureuse et fiable, puisque les quelque 1 400 mots qui composent son vocabulaire dit de premier degré ont été sélectionnés avant tout sur la base d'un critère objectif et statistique : leur fréquence d'apparition dans un corpus censément représentatif de la langue française orale (163 conversations ont été enregistrées et transcrites, pour un total d'environ 312 000 mots) (GOUGENHEIM, 1962). Ce vocabulaire « de base » a servi (et sert encore) pour l'enseignement du français langue seconde ou étrangère.

> Ogden deliberately avoided the scientific approach, and he was not lucky enough to find the key to simplicity by accident. (Flesch, 1949 : 341)

> But [...] let us not forget that Basic is still a thousand times better than the academic or bureaucratic jargon we have to wade through every day; that it is the first attempt to create a simplified language within a language; that its analysis of word meanings is an indispensable technique that will have to be taken over by whatever system of simplified English we are going to adopt. For simplified English is bound to come. (Flesch, 1949 : 343)

Les récriminations de Flesch à l'encontre des langages administratifs, politiques et juridiques sont d'ailleurs partagées par certains de ses contemporains. George Orwell[30] publie en 1946 un essai intitulé *Politics and the English Language*, dans lequel il critique l'usage qui est fait de l'anglais moderne, particulièrement dans les écrits politiques (remplis, selon lui, de « perversions ») :

> The writer either has a meaning and cannot express it, or he inadvertently says something else, or he is almost indifferent as to whether his words mean anything or not. This mixture of vagueness and sheer incompetence is the most marked characteristic of modern English prose, and especially of any kind of political writing. (Orwell, 1946)

Vingt ans plus tard, un employé du Bureau of Land Management des États-Unis rédige un document intitulé *Gobbledygook*[31] *has gotta go*, dans lequel il met en garde contre le style verbeux qui est utilisé dans les documents officiels et enjoint ses collègues à écrire de façon plus claire et directe :

> Our communications have sometimes failed because of a fascination with the traditions of officialese, an in-grown compulsion to be impressively ornate rather than simply direct, to be "proper" rather than personal. We've had costly false starts because of false notions about written communications, because of our failure to read our own writing through the other fellow's eyes.

> If we are to succeed in these times of new technologies, new demands and new attitudes, we must improve our communications radically. We must abandon soggy formality and incoherence in favor of modern personal communications. (O'Hayre, 1966 : 2)

Ces réflexions sont à l'origine du *plain language* (ou *plain English*), apparu en réaction contre les écrits (particulièrement juridiques et administratifs) dont le style, inutilement complexe (de par de longues énumérations, un vocabulaire technique

30. ↑Orwell avait initialement été séduit par l'idée d'une langue facilement accessible à tous, principe prôné par le Basic English, qu'il utilisa entre 1942 et 1944 ; mais la simplification excessive qu'implique le Basic English l'aurait finalement rendu extrêmement critique vis-à-vis de celui-ci, qui aurait d'ailleurs servi de modèle pour la « novlangue » (*Newspeak*) de son roman *1984* (publié en 1949).

31. ↑*gobbledegook* : « language, especially used in official letters, forms, and statements, that seems difficult or to mean nothing because you do not understand it » (*Cambridge Dictionary*, en ligne).

ou archaïsant, etc.[32]), rend la compréhension difficile, en particulier aux non initiés (ce qui apparente cette forme de langage à un jargon). Il est difficile de définir précisément ce qu'est (ou n'est pas) le *plain language* (et beaucoup préfèrent s'en abstenir), parce que la « clarté » d'un texte dépend de nombreux facteurs et reste évidemment subjective (et ce même si son degré de « lisibilité » peut être estimé de façon plus objective par des formules mises au point à cet effet, voir ci-après), mais les membres de l'association des professionnels du *plain language* en ont récemment proposé la définition suivante :

> A communication is in plain language if its wording, structure, and design are so clear that the intended audience can easily find what they need, understand what they find, and use that information. (citée par CUTTS, 2013)

Ainsi, le *plain language* ne se limite pas à des caractéristiques proprement langagières (il peut aussi s'agir de mise en page et de structuration de l'information) et, surtout, il ne peut pas être défini en dehors du public (supposé) auquel le document rédigé s'adresse[33]. Bien souvent, lorsqu'il n'est pas identifié précisément, celui-ci est assimilé au « citoyen moyen » – dont le niveau de littératie est limité. L'intérêt de la rédaction en *plain language* est évident : permettre à chacun d'accéder plus facilement et correctement à l'information qui lui est destinée, qu'il s'agisse d'instructions pour un jury, d'un formulaire administratif ou encore d'un contrat d'assurance, par exemple. Il est tel qu'aux États-Unis, le *Plain Writing Act* en vigueur depuis 2010 impose que les moyens nécessaires soient mis en œuvre pour que tous les documents émis par des agences fédérales à destination du grand public soient suffisamment clairs ; il avait déjà été précédé de lois similaires dans certains états et au niveau fédéral. On ne s'étonne donc pas que depuis plusieurs années, une foison de guides de rédaction en *plain language*, dont les recommandations sont souvent très proches de celles des langues contrôlées (préférer la voix active à la voix passive, faire des phrases courtes, etc.), ait été publiée.

C'est dans une optique comparable que sont apparues, il y a près d'un siècle, les premières formules visant à évaluer la *lisibilité*[34] des textes écrits (à l'origine pour des raisons pédagogiques : sélectionner des ouvrages adaptés à une certaine classe

32. ↑HALLIDAY (2004 : 131) avance que le langage bureaucratique est complexe à dessein, parce que cette complexité permet d'exclure ceux qui ne sont pas en mesure de le saisir. En cela, il s'inspirerait selon lui du langage scientifique et technique, qui, lui, est exigeant parce que les sujets abordés sont réellement pointus : « [...] the language of science is difficult ; and this has two related consequences. On the one hand, you have to be an expert in order to understand it ; it is not enough merely to be educated and inquiring. So technical discourse soon becomes technocratic discourse, written by experts with the message 'this is too hard for you to understand ; better leave the decisions to us' (see Lemke 1995, Chapter 4). On the other hand, it is stamped with authority, derived from the familiar equation of 'knowledge equals power'. So technical discourse is parodied by bureaucratic discourse [...] ».

33. ↑Il faut comprendre par là que le langage utilisé ne doit pas être trop complexe, mais aussi (même si ce problème se pose probablement moins souvent) qu'il ne doit pas être trop simple : il ne faut évidemment pas s'adresser à un public universitaire comme à de jeunes enfants.

34. ↑La *lisibilité* (souvent assimilée à la *compréhensibilité*) est un concept assez général qui peut renvoyer à des propriétés langagières (vocabulaire, longueur des phrases, etc.), mais aussi typographiques et de mise en forme (taille des caractères, présence d'images en appui du texte, etc.) ; la

d'âge). Exactement comme le *plain language*, dont elle est très proche et qu'elle a influencé, la lisibilité est une notion à la fois simple à appréhender (« what makes some texts easier to read than others » (DuBay, 2004 : 3)) et difficile à définir, ce qui explique pourquoi il y a pratiquement autant de définitions que d'auteurs qui s'y sont intéressés. Pour DuBay, celle de Dale et Chall est l'une des plus exhaustives :

> The sum total (including all the interactions) of all those elements within a given piece of printed material that affect the success a group of readers have with it. The success is the extent to which they understand it, read it at an optimal speed, and find it interesting. (Dale et Chall, 1949, cités par DuBay, 2004 : 3)

Mais comment calculer et surtout estimer à quel point un écrit est facile à comprendre et à lire et intéressant [35] pour un groupe de lecteurs ? C'est à cette question que les formules de lisibilité tentent d'apporter une réponse. Il en existe de nombreuses, plus ou moins similaires, des plus simples aux plus complexes : DuBay (2004), citant Klare, signale qu'au début des années 1980, leur nombre dépassait déjà deux cents. Leur succès est indéniable depuis le milieu du siècle passé, à tel point que certaines sont implémentées dans des logiciels destinés au grand public ; dans certains cas (documents administratifs, ouvrages scolaires), un niveau déterminé de lisibilité est même exigé de la part des rédacteurs, ce qui rejoint les exigences du *plain language*.

Sans entrer dans le détail des variables qui ont pu être prises en compte par ces formules pour attribuer un score de lisibilité à un (extrait de) texte, très diverses (Klare, 1988), ni de leur pondération, soulignons que les facteurs qui ont été les premiers à être suspectés d'influencer la lisibilité, à savoir la difficulté et la richesse du vocabulaire ainsi que la longueur des phrases [36] ou des mots, sont aussi ceux que l'on retrouve dans la plupart des formules ultérieures les plus populaires. Celle de Flesch est basée sur la longueur moyenne des mots (en syllabes) et des phrases (en mots), celle de Dale-Chall sur la longueur moyenne des phrases et le pourcentage de mots difficiles (c'est-à-dire absents d'une liste de trois mille mots fréquents et simples établie par leurs soins), celle de Gunning (nommée Fog Index) sur la longueur moyenne des phrases et sur le nombre de mots de plus de deux syllabes, tandis que celle de Fry se présente sous la forme d'un graphique dont l'axe horizontal représente la longueur moyenne des mots (en syllabes) et l'axe vertical la longueur moyenne des phrases. La différence peut alors se situer dans la façon de combiner ces variables, dans leur coefficient, dans la façon d'interpréter les résultats, etc.

---

langue anglaise distinguant pour cela entre *readability* et *legibility*, respectivement. Nous n'évoquerons ici que les premières, les secondes étant en dehors de notre champ disciplinaire.

35. ↑L'intérêt d'un texte est forcément subjectif, donc difficile à prédire de façon fiable, mais il est entendu qu'il participe de la facilité à lire le texte et à en mémoriser le contenu.

36. ↑En 1880, un professeur de littérature anglaise mettait en évidence qu'au fil des époques, le nombre moyen de mots par phrase diminuait progressivement dans les œuvres étudiées, rapprochant ainsi la langue écrite de la langue orale, qu'il jugeait plus compréhensible (DuBay, 2004 : 10).

Il en ressort que ces formules restent très simples : compter le nombre de syllabes ou de mots et vérifier la présence d'un mot dans une liste sont des tâches relativement aisées[37] et qui peuvent être automatisées sans peine. C'est leur grande force (et la raison de leur succès) en même temps que leur principale limitation, qui leur aura valu nombre de critiques, puisqu'elles laissent de côté d'autres facteurs sans doute au moins aussi importants : la complexité syntaxique, la cohérence du texte (il est très simple de produire un écrit dépourvu de sens et de logique qui obtienne malgré tout un excellent score de lisibilité), son intérêt (dont parlaient pourtant Dale et Chall dans leur définition), et bien d'autres.

Elles ont logiquement leurs partisans et leurs détracteurs[38], et leur fiabilité a fait l'objet de très nombreuses études (plus d'un millier, à en croire DuBay (2004)). En particulier, les écarts constatés entre les résultats obtenus, pour un même texte, avec des formules différentes ont de quoi laisser perplexe (Zamanian et Heydari, 2012). D'un autre côté, elles permettent un gain de temps et d'objectivité par rapport au jugement humain et peuvent guider le rédacteur dans son travail d'écriture, à condition d'être utilisées à bon escient. Certains chercheurs ont en outre montré que leurs résultats étaient assez bien corrélés à ceux d'autres mesures de compréhension à la lecture, et des formules beaucoup plus complexes n'apporteraient probablement pas un gain significatif dans cette évaluation.

En conclusion, ces formules de lisibilité, pour critiquables qu'elles soient, ont le mérite de fournir très simplement et très rapidement un indicateur de la complexité du texte. Celui-ci n'est toutefois pas d'une fiabilité absolue et il convient d'en relativiser la portée ; de plus, elles ne fournissent pas d'information sur la façon d'améliorer le texte. Mais qu'en est-il de leur intérêt pour la rédaction des exigences ? Comme nous le verrons au chapitre 3, la longueur des phrases est une mesure dont ont hérité bon nombre de langues contrôlées et qui reste en vogue aujourd'hui encore. En revanche, diminuer la complexité du texte en diminuant la complexité de son vocabulaire paraît plus difficile lorsque l'emploi de nombreux termes techniques est inévitable, comme c'est le cas dans les domaines spécialisés. À ce titre, il faut rappeler que peu de formules de lisibilité ont été élaborées spécifiquement pour des adultes, ce qui rend leur pertinence douteuse lorsqu'il s'agit de les appliquer à la rédaction technique au sens large (Redish et Selzer, 1985)[39]. Enfin, et c'est selon nous le point le plus crucial, elles passent totalement sous silence les phénomènes qui nous intéressent le plus (l'ambiguïté, le flou et l'incomplétude), qu'elles n'ont tout simplement pas vocation à détecter. Comme le précise Connatser :

37. ↑Si l'on fait abstraction, notamment, des problèmes d'homonymie.

38. ↑Parmi ces derniers, un certain Ogden : « "Word-counters and text-graders" Ogden calls, [...] laughing at their "parrot-mentality". » (Flesch, 1944 : 339).

39. ↑Une expérimentation menée par Cadwell (2008), qu'il admet imparfaite et dont rend compte O'Brien (2010), a montré qu'appliquer certaines des règles des langues contrôlées sur des textes techniques (de difficulté variée) n'offrait qu'une amélioration marginale de leur score de lisibilité. En revanche, le jugement des évaluateurs humains et leur taux de mémorisation du texte parlent clairement en faveur des versions modifiées selon les principes des langues contrôlées.

> Most audiences of technical documents *read to do*. Therefore, usability testing of a document seems much more appropriate for measuring how effectively a text conveys technical information than a formula. (Connatser, 1999 : 284, cité par O'Brien, 2010 : 26)

Le Basic English, le *plain language* et les formules de lisibilité ont donc en commun d'être centrés sur la « compréhensibilité » (quelle que soit la définition exacte que l'on donne à ce terme), qu'il s'agisse de l'évaluer ou de l'améliorer, et d'être très généraux : ils concernent la langue (anglaise essentiellement) en général et n'ont pas été conçus pour être appliqués à un domaine particulier. Le Basic English et les formules de lisibilité sont d'ailleurs mal adaptés aux textes techniques.

C'est au début des années 1970, soit quarante ans après le Basic English (dont l'influence est avérée), qu'apparaît la première langue contrôlée connue pour être issue d'une entreprise, le Caterpillar Fundamental English (CFE) (Kuhn, 2014). Comme son nom l'indique, le CFE est lui aussi basé sur l'anglais et a été développé par le groupe américain Caterpillar. Son objectif était de faciliter la communication avec les travailleurs non anglophones et plus spécifiquement d'éliminer les besoins de traduction des documents. L'idée, semblable à celle qui a donné naissance au Basic English, était qu'un employé devait être capable de comprendre les textes écrits en CFE après avoir suivi un nombre limité de leçons. Ce dernier consiste en un vocabulaire limité[40] (environ mille mots) et en une série de règles grammaticales qui, contrairement à celles du Basic English, l'éloignent notablement de l'anglais non contrôlé et constituent des contraintes fortes pour les rédacteurs : utiliser des phrases positives, éviter le conditionnel, utiliser des structures de phrase uniformes, etc. Il sera abandonné en 1982, les règles ne pouvant être appliquées systématiquement dans les documents produits au sein de l'entreprise, et remplacé près de dix ans plus tard par le Caterpillar Technical English – dont l'objectif n'est plus de se passer totalement de la traduction de l'anglais vers d'autres langues, mais de la faciliter et ainsi d'en réduire le coût (une motivation partagée par beaucoup d'autres langues contrôlées, en particulier à l'ère de la traduction automatique).

Le CFE initiera ainsi une série de langues contrôlées destinées à la rédaction technique (nous en présenterons deux autres dans le chapitre 3) dont le but avoué est de faciliter la communication humaine (et la traduction automatique) en restreignant plus ou moins fortement le lexique et la syntaxe de la langue sur laquelle elles sont basées (l'anglais le plus souvent, mais on trouve des langues contrôlées conçues à partir de nombreuses autres langues naturelles). Malgré tout, les énoncés produits en respectant ces langues contrôlées restent relativement similaires à ceux produits dans la langue naturelle correspondante ; le rédacteur doit fournir un effort supplémentaire pour se conformer aux recommandations, mais le texte ainsi rédigé est censément plus simple pour le lecteur et demeure (relativement)

40. ↑Qui se justifie plus aisément que pour le Basic English, puisqu'il est limité aux activités de Caterpillar.

naturel.

Certaines langues contrôlées, apparues au milieu des années 1990 (KUHN, 2014), vont plus loin et ambitionnent de permettre des traitements et des raisonnements automatiques sur les textes. La plus connue d'entre elles est sans doute Attempto Controlled English (ACE) (FUCHS, SCHWERTEL et SCHWITTER, 1998), qui permet de traduire automatiquement les énoncés en logique du premier ordre. Un tel objectif nécessite impérativement que ces derniers soient non ambigus, ce qui les rend souvent un peu moins « naturels » et limite quelque peu l'expressivité. Depuis lors, de nouvelles applications ne cessent d'être proposées pour les langues contrôlées, qu'il s'agisse de communication d'humain à humain ou d'humain à machine, notamment avec l'essor des objets connectés (O'LEARY et al., 2017).

Les langues contrôlées sont donc apparues dans des milieux différents (industriels, académiques, gouvernementaux) et avec des objectifs parfois différents, mais non exclusifs (augmenter la compréhensibilité, faciliter la traduction ou la représentation formelle). Certaines s'appliquent uniquement à l'écrit, d'autres à l'oral ; certaines sont générales, d'autres liées à un domaine spécifique. En conséquence, leur nombre n'a cessé de croître : en 2014, Kuhn en recensait cent, rien que pour l'anglais. Malheureusement, parmi celles qui ont été développées par des entreprises privées, beaucoup sont soumises à une politique de confidentialité contraignante et sont donc méconnues en dehors.

1.5.3 *Définition*

On l'a vu, les langues contrôlées ont pu revêtir de nombreuses formes et recevoir de nombreuses appellations depuis la création du Basic English en 1930. KUHN (2014 : 121) cite quelques-uns des noms qui ont été donnés, en anglais, à ces langues construites : *controlled* (d'où les termes *controlled natural language*, ou simplement *controlled language*, qui semblent aujourd'hui s'imposer), *processable*, *simplified*, *technical*, *structured* et, donc, *basic*. On trouve également un français *rationalisé* (LUX, 1998). Ces différentes étiquettes sont pour certaines le reflet des objectifs variés qui ont présidé à leur création, parmi lesquels la simplification de la langue (*simplified*) et de son traitement (*processable*), particulièrement en vue de la traduction automatique.

Kuhn fait donc remarquer que le concept même de *langue contrôlée* possède des contours très flous et qu'il ne semble pas y avoir de consensus sur les propriétés essentielles de telles langues. Il y voit deux raisons. Premièrement, les langues contrôlées sont apparues dans différents environnements (l'industrie, le milieu académique et les gouvernements), dans différentes disciplines (l'informatique, la philosophie, la linguistique et l'ingénierie) et durant plusieurs décennies, ce qui explique l'emploi de termes différents pour des concepts parfois très proches. Deuxièmement, ces langues contrôlées se présentent sous des formes et avec des objectifs parfois divers, ce qui rend difficile l'identification de leurs caractéristiques

communes.

En comparant, ainsi que le fait Kuhn, deux définitions de *langue contrôlée* données par des auteurs différents, ces divergences apparaissent clairement :

> Controlled natural language is a subset of natural language that can be accurately and efficiently processed by a computer, but is expressive enough to allow natural usage by non-specialists. (Fuchs et Schwitter, 1995 : 1, cités par KUHN, 2014 : 123)

> A controlled language (CL) is a restricted version of a natural language which has been engineered to meet a special purpose, most often that of writing technical documentation for non-native speakers of the document language. A typical CL uses a well-defined subset of a language's grammar and lexicon, but adds the terminology needed in a technical domain. (Kittredge, 2003 : 441, cité par KUHN, 2014 : 122)

La première définition n'envisage que les langages destinés à être traités par des ordinateurs, « formalistes » ; la seconde se concentre sur les langages techniques visant à améliorer la compréhensibilité (en particulier pour les locuteurs non natifs), « naturalistes » (CLARK et al., 2009).

Elles font néanmoins toutes deux remarquer qu'une langue contrôlée est nécessairement basée sur une langue naturelle, tout en étant plus restrictive que cette dernière (« subset of natural language », « restricted version of a natural language », « [...] well-defined subset of a language [...] »). Le terme *subset* (*sous-ensemble*) qui y est employé ne peut toutefois pas être pris littéralement dans son acception mathématique, dans la mesure où certaines langues contrôlées s'écartent sensiblement de la langue naturelle dont elles s'inspirent, voire lui ajoutent de nouveaux éléments. En conséquence, il ne semble pas possible d'identifier un noyau commun à toutes les langues contrôlées, ni même à celles qui sont basées sur une même langue naturelle (voir O'BRIEN (2006) pour une analyse de huit langues contrôlées basées sur l'anglais : une seule règle est partagée par toutes). Il faut ajouter à cela qu'une langue contrôlée est normalement un langage construit – il n'est pas apparu spontanément. En cela, elle tient davantage de la normalisation que de la normaison.

Partant de ces constats, Kuhn propose la définition suivante (qui semble s'être imposée comme référence, tant elle a été reprise et citée depuis [41]) :

> A language is called a *controlled natural language* if and only if it has all of the following four properties:

41. ↑Y compris dans la norme ISO 24620 (2015 ; CHOI et ISAHARA, 2014), consacrée aux langues contrôlées, où elle est paraphrasée et dans laquelle une autre définition est également donnée : « subtype of human language whose structure is restricted and whose ambiguity is resolved for a particular purpose ».

1. It is based on exactly one natural language (its "base language").
2. The most important difference between it and its base language (but not necessarily the only one) is that it is more restrictive concerning lexicon, syntax, and/or semantics.
3. It preserves most of the natural properties of its base language, so that speakers of the base language can intuitively and correctly understand texts in the controlled natural language, at least to a substantial degree.
4. It is a constructed language, which means that it is explicitly and consciously defined, and *is not* the product of an implicit and natural process (even though it is based on a natural language that is the product of an implicit and natural process). (KUHN, 2014 : 123)

Ou, plus concisément :

> A *controlled natural language* is a constructed language that is based on a certain natural language, being more restrictive concerning lexicon, syntax and/or semantics, while preserving most of its natural properties. (*ibid.*)

Il faut distinguer des langues contrôlées proprement dites des concepts proches, mais qui s'en distinguent car ils ne respectent pas toutes les conditions énoncées plus haut (KUHN, 2014 : 124). Un *sous-langage* (voir section 1.4) est basé sur une langue naturelle (condition 1), dont il reste proche (condition 3), tout en étant plus restrictif (condition 2), mais, contrairement à un langage contrôlé, il n'est pas construit (condition 4) ; au contraire, il apparaît naturellement. Les *guides de style* ne décrivent pas un nouveau langage, mais se contentent de donner des conseils sur la manière de rédiger dans la langue naturelle considérée. Les *phraséologies* sont des ensembles d'expressions (en anglais, *phrases*) utilisées par un groupe particulier ; si elles sont construites, elles peuvent alors être considérées comme des langues contrôlées. Les *vocabulaires contrôlés* sont des collections de noms et d'expressions et ne proposent pas de règles grammaticales permettant de combiner ces termes entre eux pour former des phrases. Les langues contrôlées destinées à un domaine spécifique incluent souvent un vocabulaire contrôlé. Enfin, le terme *langage construit* (ou *langue construite* ou *artificielle*) est plus générique, puisqu'il regroupe tous les langages qui ont été consciemment définis (au contraire des langues naturelles), qu'il s'agisse de langues contrôlées, de langues telles que l'espéranto, ou encore de langages de programmation.

La distinction entre ces concepts voisins n'est pas toujours aisée et il n'est pas certain que tout ce qui est appelé « langue contrôlée » dans la littérature respecte strictement toutes les conditions ; inversement, certaines langues contrôlées peuvent y répondre sans être présentées comme telles par leurs créateurs (SUCHOWOLEC, 2014). De plus, d'autres catégories de langues contrôlées pourraient encore être identifiées : on peut par exemple distinguer entre celles qui optent pour une approche ascendante (en listant tout ce qui est autorisé ; ce qui n'y figure pas étant interdit) et celles dont l'approche est plutôt descendante (en listant tout ce qui est

interdit par rapport à la langue de base ; ce qui n'y figure pas étant autorisé). Parallèlement, certaines règles pourraient être qualifiées de prescriptives (celles qui imposent une ou plusieurs structures dans certains cas bien définis) et d'autres de « proscriptives » ou « prohibitives » (celles qui, au contraire, en interdisent l'utilisation) – et l'on conçoit bien que les deux types de règles puissent se retrouver au sein d'une même langue contrôlée.

Qu'importe : l'essentiel, pour notre travail, est de montrer qu'il est possible, en contraignant la langue naturelle, de pallier (en partie) ses inconvénients tout en conservant une solution plutôt simple à mettre en place et tout aussi expressive, contrairement aux langages purement formels. Nous entendons donc construire (condition 4) une langue contrôlée basée sur le français (condition 1), qui le restreigne aux niveaux lexical, morphosyntaxique et sémantique (condition 2) tout en conservant au maximum son caractère naturel (condition 3) – tout l'enjeu étant qu'elle soit *suffisamment*, mais pas *trop* contraignante.

1.5.4 *Évaluation*

Pour terminer ce premier chapitre, il nous reste à évoquer le manque (surprenant) d'évaluation systématique des langues contrôlées. Comme on l'a dit, les langues contrôlées se sont multipliées au cours des dernières décennies – ce qui indique que leur intérêt potentiel est bien perçu – mais, contrairement aux formules de lisibilité (auxquelles une large littérature est consacrée), assez peu d'études ont été menées pour en vérifier l'efficacité réelle. Autrement dit, certaines langues contrôlées ont pu être conçues et même utilisées sans que leurs concepteurs et utilisateurs ne se soient au préalable assurés du bénéfice apporté. Pourtant, on s'en doute, mettre au point une telle norme est loin d'être chose aisée – l'abandon du Basic English et du CTE en sont des preuves concrètes – et une norme mal conçue court plusieurs risques (par degré croissant de préjudiciabilité) : celui d'être inutilisable (non respectée), celui d'être inutile (sans effet) et celui d'être contreproductive (avec des effets indésirables[42]), trois situations qui doivent évidemment être évitées. Selon nous, l'évaluation des règles devrait faire partie intégrante de la conception de toute langue contrôlée[43] ; or, de façon générale, les langues contrôlées précisent pourquoi les règles devraient être suivies (réduire l'ambiguïté, rendre le texte plus court, etc.), mais pas comment ces dernières ont été élaborées, ni si (et, le cas échéant, comment) elles ont été évaluées.

Bien entendu, nous ne pouvons pas exclure catégoriquement que ces évaluations aient eu lieu, mais, en l'absence d'études publiées pour en rendre compte,

42. †Une étude effectuée à la NASA a montré que les messages présentés sous forme de phrases complètes (ex. : « The fuel pressure is low ») étaient mieux compris, mais aussi plus rapidement traités que ceux qui se réduisaient à des mots-clés (« fuel low »), pourtant plus courts (Simpson, 1976, citée par Jahchan, Condamines et Cannesson, 2016).

43. †De deux points de vue, malheureusement généralement aussi négligés l'un que l'autre : leur utilisabilité pour les rédacteurs et leur intérêt pour les lecteurs. Le bénéfice obtenu par ceux-ci devrait surpasser le coût imposé à ceux-là.

tout porte à croire que ce n'est pas le cas. O'Brien, qui s'intéresse en particulier à l'intérêt des langues contrôlées pour faciliter la traduction, attribue cet état de fait à l'aspect « propriétaire » et confidentiel d'une majorité de langues contrôlées :

> As stated previously, few empirical studies on CL have been published. This can be attributed to the fact that the implementation of CL is most often executed in a proprietary environment. Thus, there is little published evidence that using a CL reduces information production costs. (O'Brien, 2006 : 14)

Le cas de l'anglais simplifié (que nous présenterons en détail au chapitre 3), l'une des langues contrôlées les plus populaires encore aujourd'hui, est symptomatique : connu depuis les années 1980, il faudra pourtant attendre le milieu des années 1990 et certaines critiques pour que son efficacité soit réellement éprouvée. Les résultats obtenus sont globalement positifs, mais doivent être relativisés : certains ne sont tout simplement pas significatifs (Stewart, 1998) et d'autres ne le sont que dans des circonstances déterminées (avec des locuteurs non natifs, sur des textes plus difficiles) (Shubert et al., 1995 ; Chervak, Drury et Ouellette, 1996).

Ils sont donc plutôt encourageants, mais il est tout à fait regrettable qu'ils ne soient arrivés qu'aussi tardivement, alors qu'ils auraient dû être une condition nécessaire à l'adoption de l'anglais simplifié et des langues contrôlées qui s'en inspirent. Les pratiques tendent heureusement à évoluer positivement, que ce soit par la mise en place de méthodologies d'évaluation des langues contrôlées (Kuhn, 2009) ou grâce à l'utilisation de techniques de mesure plus avancées, telles que l'oculométrie. O'Brien (2010) est ainsi parvenue à une conclusion similaire à celle que nous évoquions juste avant : l'effet des langues contrôlées sur la lisibilité existe, mais il est plus limité qu'on ne le croit et s'applique surtout dans certaines situations (fragments de textes complexes) ; il n'est du reste pas exclu que cet effet puisse être négatif dans d'autres (par exemple, un texte trop simple pourrait être jugé trop peu stimulant par des experts).

1.6 CONCLUSION

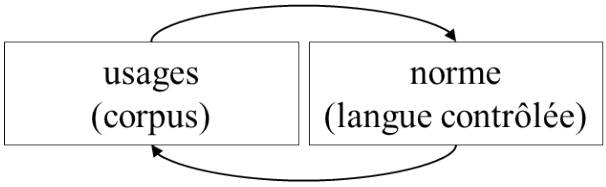
Les risques liés à l'utilisation de la langue naturelle pour transmettre de l'information sont bien connus : flou, généralité, ambiguïté (Zhang, 1998), incomplétude des mots, des énoncés et du discours. De ce fait, et en particulier dans des contextes bien définis (comme celui d'une communauté de discours professionnelle), proposer ou imposer une norme langagière nous semble être une solution acceptable (à défaut de mieux) pour améliorer la communication et prévenir les conséquences indésirables liées à ces risques.

Il n'est un secret pour personne que les normes prescriptives sont parfois mal considérées, en particulier par la linguistique, dont le point de vue, en tant que science, doit être descriptif (et explicatif), et non normatif. Cela tient aussi à ce que ces normes peuvent être stigmatisantes pour ceux qui n'en maîtrisent pas les

codes (comme c'est le cas pour toute autre norme sociale) et ne sont pas toujours bien adaptées (et donc utilisables efficacement) à ceux à qui elles sont destinées (ce qui devrait être l'objectif de toute norme technique) ; plus généralement, on peut admettre que « trop de norme tue la norme » – la norme ne doit être qu'un moyen et pas une fin en soi.

Notre opinion, en tant que linguiste confronté à un besoin réel, est que toute norme qui entend contraindre les usages doit nécessairement tenir compte de ces usages (ce qui implique de les décrire et de les comprendre au préalable) (FIGURE 1). Ainsi, pour reprendre les termes adoptés par la socioterminologie, la normalisation devrait être basée sur la normaison et ne l'influencer que là où c'est nécessaire (et profitable). Nous sommes convaincu que faire fi de cette dernière expose à l'échec, que celui-ci se manifeste par la non-utilisation de la norme ou par une perte de temps due au respect de règles inadéquates.

FIGURE 1 – Influence réciproque de la norme et des usages



De la même façon, les langues contrôlées (dans leur sens le plus large) ont déjà eu, depuis de nombreuses années maintenant, l'occasion de montrer leur intérêt dans des contextes et des applications très divers. Il n'en reste pas moins que la façon dont elles sont conçues et évaluées peut encore être sensiblement améliorée : c'est précisément cet apport méthodologique que nous visons dans ce travail, appliqué à la rédaction des exigences (que nous présenterons dans le chapitre suivant). Pour cela, nous essayerons (au chapitre 3) de mettre en évidence l'existence d'un genre (ou sous-langage) propre aux exigences.

2

APPORT DE LA LINGUISTIQUE DE CORPUS POUR LA RÉDACTION DES EXIGENCES

2.1 LA RÉDACTION DES EXIGENCES : BESOINS ET ENJEUX

2.1.1 *Besoin, exigences et spécifications*

Supposer l'existence d'un genre textuel (voir section 1.4) – en l'occurrence, celui des exigences – implique d'identifier les facteurs extralinguistiques qui le définissent. Une *exigence* (en anglais : *requirement*) est définie conjointement par l'Organisation internationale de normalisation (ISO), la Commission électrotechnique internationale (IEC) et l'*Institute of Electrical and Electronics Engineers* (IEEE), qui font autorité internationalement, comme

> [a] statement which translates or expresses a need and its associated constraints and conditions (ISO/IEC/IEEE29148, 2011 : 5)

Cette définition plutôt générale met l'emphase sur la notion de *besoin* (« need »), dont l'exigence est l'expression et qui s'accompagne éventuellement de contraintes et de conditions. Contrairement à ce que peut peut-être laisser penser le mot « statement », les exigences ne se présentent pas toujours sous forme textuelle ; il peut aussi s'agir d'un diagramme, d'une formule, d'un modèle simulé informatiquement ou de tout autre moyen apte à transcrire clairement le besoin dont il est question, voire d'une combinaison de plusieurs systèmes sémiotiques. La définition proposée en français par la Société pour la Promotion Et la Certification de l'Ingénierie des Exigences en langue Française (SPECIEF) sur son site est assez semblable à la première :

> Une exigence est l'expression d'une condition ou d'une fonctionnalité à laquelle doit répondre un système ou un logiciel. (http://www.specief.org/index.php/ingenierie-des-exigences/)

Cette définition précise que le besoin correspond à une « condition » ou une « fonctionnalité » [1], ce qui peut se traduire par deux types d'exigences [2] : les exigences dites *fonctionnelles* (par exemple, « Divers outils de sélection [...] doivent être à la disposition de l'utilisateur [...] ») et *non fonctionnelles* (qui sont souvent

1. ↑Elle est très proche d'une ancienne définition proposée par l'IEEE : « [A documented representation of] a condition or capability that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed document » (IEEE, 1990 : 62).

2. ↑Évidemment, il existe d'autres typologies des exigences, plus complexes. Il faut aussi tenir compte d'autres contraintes telles que la législation, des réglementations propres au domaine, etc. La définition même de « fonction » n'est de toute façon pas triviale (Tommila et Pakonen, 2014 : 8).

des exigences de performance, par exemple : « [...] le guidage devra assurer une mesure stellaire permanente [...] »). Elle nous indique également que les exigences relèvent de l'ingénierie système et du génie logiciel. Enfin, la présence du verbe « devoir » est significative : l'exigence crée en effet une obligation, ce qui est une de ses caractéristiques majeures (nous y reviendrons au chapitre 3)[3]. Sur ce point, la définition de l'Association Française d'Ingénierie Système (AFIS) est encore plus explicite :

> Une exigence prescrit ce qu'un Système doit faire, avec quelles performances et sous quelles conditions, pour atteindre un but donné. (AFIS, 2012 : 20)

Le besoin (qui se transforme dans cette dernière définition en objectif à atteindre, ce qui revient sensiblement au même) est effectivement une notion centrale : il est exprimé par une des parties prenantes[4] (le maître d'ouvrage ou client) et traduit en *exigence de besoin* à l'issue d'une « négociation » (ou, pour le dire plus positivement, d'un échange aboutissant à la construction de connaissances communes, le plus souvent de façon incrémentale) avec le maître d'œuvre (ou fournisseur). Ces exigences de besoin, généralement exprimées dans le langage éventuellement moins technique du maître d'ouvrage, constituent le cahier des charges et servent de base à la définition d'exigences techniques (ou *exigences système*) sur lesquelles s'engage le maître d'œuvre.

Il est important de noter qu'une exigence de besoin ne décrit théoriquement pas une solution ; autrement dit, elle dit ce qui est attendu du système (son comportement), mais pas comment celui-ci doit être réalisé (son implémentation)[5] – ce choix revenant au fournisseur. Les systèmes complexes se décomposent généralement en plusieurs niveaux ; à chacun de ces niveaux sont rattachées des exigences dont le degré d'abstraction varie (elles sont logiquement plus générales au niveau du système et plus détaillées dans les niveaux inférieurs). Les exigences sont rassemblées (par exemple, par niveau) dans des *spécifications*, qui peuvent être définies comme des collections structurées d'exigences[6].

Exigences et spécifications relèvent donc de la rédaction technique (sur un sujet spécialisé), d'expert à expert, mais se distinguent des textes procéduraux, parmi

3. ↑L'utilisation du présent de l'indicatif est trompeuse, le système ou le logiciel n'existant pas (encore) au moment où l'exigence est définie. C'est là l'autre caractéristique majeure, sur laquelle nous reviendrons également.

4. ↑Par souci de simplicité, nous résumons les parties prenantes à la maîtrise d'ouvrage et à la maîtrise d'œuvre, mais la réalité est beaucoup plus complexe, puisque « partie prenante » (*stakeholder*) peut désigner « [any] individual or organization having a right, share, claim, or interest in a system or in its possession of characteristics that meet their needs and expectations » (ISO/IEC/IEEE29148, 2011 : 6) : il peut donc aussi s'agir d'utilisateurs, de partenaires, d'actionnaires, etc.

5. ↑Bien entendu, dans la pratique, la frontière entre les deux n'est pas toujours aussi nette. Il peut aussi arriver qu'une solution soit imposée au fournisseur par le client.

6. ↑Ainsi, au CNES, les différents niveaux d'un projet sont les suivants : Mission, Système, Segment, Module, Équipement. On trouve des spécifications pour chaque produit (par exemple, la charge utile, qui est un segment) ainsi que des spécifications d'*interface* entre produits d'un même niveau (par exemple, l'interface entre plateforme et charge utile).

lesquels on classe les modes d'emploi et autres manuels de maintenance. En effet, sont appelés *procéduraux*

> [...] les textes qui ont en commun de dire de faire et de dire comment faire en prédisant un résultat et en incitant très directement à l'action (Adam, 2001 : 19, cité par COUTHERUT, 2017 : 70)

Or, si les exigences incitent bien à l'action – il ne s'agit pas de textes narratifs ou argumentatifs, par exemple –, cette incitation n'est qu'implicite (et certainement pas « très directe »). Surtout, elles ne disent pas *comment* faire : il ne s'agit jamais d'instructions. Selon Pitar, citée par COUTHERUT (2017 : 71), les textes procéduraux suivent tous le même format : « objectif – état initial – mode d'action – état final ». Les exigences se limitent à l'objectif, l'état initial n'existant tout simplement pas (puisque ce qui est décrit n'existe pas encore). Elles partagent toutefois avec eux le fait de laisser au lecteur la marge d'interprétation la plus réduite possible (GANIER, 2006), du moins dans l'idéal.

2.1.2 *Enjeux*

Les exigences revêtent une importance cruciale, car elles ont valeur contractuelle (et sont donc des actes juridiques) : le maître d'œuvre doit être en mesure de prouver (par des méthodes de vérification et de validation : démonstration, test, etc.) qu'il a correctement répondu aux exigences sur lesquelles il s'est engagé. Même pour les projets réalisés en interne, la bonne gestion des exigences, de leur élucidation[7] à leur validation, est un facteur majeur de succès ou d'échec. Le Standish Group[8] publie annuellement depuis plus de vingt ans le « Chaos Report », un rapport (dont les chiffres, abondamment cités, sont néanmoins à prendre avec précaution (EVELEENS et VERHOEF, 2010)) rendant compte du taux de performance d'un grand nombre de projets informatiques, qui sont classés en trois catégories (pleinement réussis ; terminés avec du retard et/ou un dépassement de budget et/ou des fonctionnalités manquantes ; abandonnés), et précisant les principales causes d'échec identifiées : les exigences de mauvaise qualité y trônent régulièrement en tête (DICK, HULL et JACKSON, 2005). D'autres études aboutissent à des conclusions similaires, à savoir que la phase de définition des exigences est déterminante pour la bonne réalisation des objectifs ; cela est évidemment d'autant plus vrai s'agissant des projets de grande ampleur, qui s'inscrivent dans une durée longue et font intervenir de nombreux besoins et de nombreuses parties prenantes – ce qui est le cas des systèmes orbitaux.

Les conséquences potentielles d'une mauvaise gestion ou d'une mauvaise interprétation des exigences sont en effet particulièrement dommageables : retards et

7. ↑Nom de l'activité consistant à les définir (en anglais *elicitation*).
8. ↑https://www.standishgroup.com/

augmentation des coûts, litiges[9] ou, dans le pire des cas, accidents[10]. Dès lors, il est logique que les exigences fassent l'objet d'une attention toute particulière depuis déjà plusieurs dizaines d'années[11], à tel point qu'une discipline leur soit dédiée : l'ingénierie des exigences. Celle-ci vise notamment à définir les meilleures pratiques à adopter ainsi qu'à définir les caractéristiques d'une « bonne » exigence (ISO/IEC/IEEE29148, 2011 : 11) :

— nécessaire
— libre d'implémentation
— cohérente
— réalisable
— traçable
— vérifiable
— singulière ou atomique (elle ne doit exprimer qu'un seul besoin)
— complète
— non ambiguë.

En pratique, ces propriétés des exigences individuelles – il en existe d'autres pour les spécifications – sont plus ou moins simples à vérifier ; il faudrait sûrement davantage y voir des idéaux. La non-ambiguïté nous intéresse tout particulièrement :

> The requirement is stated in such a way so that it can be interpreted in only one way. The requirement is stated simply and is easy to understand. (ISO/IEC/IEEE29148, 2011 : 11)

La « facilité de compréhension » et la simplicité renvoient directement aux problématiques évoquées dans la section consacrée aux langues contrôlées (section 1.5). Cette contrainte prévoit aussi que l'exigence soit strictement univoque, une propriété qui découle directement de sa formulation. Ceci (et certaines autres propriétés, dans une moindre mesure) explique le besoin du CNES, auquel nous nous efforçons d'apporter une solution : améliorer la rédaction des exigences. En effet, les ingénieurs du CNES disposent déjà de solutions dédiées à la gestion des exigences (et en particulier à leur traçabilité), telles que les logiciels IBM Rational DOORS[12] (pour lequel un modèle de données spécifique est utilisé) et Reqtify[13]. Ces outils permettent, entre autres, de gérer les versions successives des exigences

9. ↑Un juriste du CNES nous a rapporté qu'un différend les avait opposés à l'un de leurs fournisseurs sur l'interprétation à donner à un terme (« systèmes sensibles ») utilisé dans une spécification et auquel aucune définition n'était donnée.

10. ↑L'exemple de la sonde américaine *Mars Climate Orbiter* (lancée en 1998 par la NASA et perdue avant qu'elle n'atteigne son objectif en raison d'une mauvaise spécification des unités de mesure utilisées) est resté tristement célèbre.

11. ↑À titre informatif, la première conférence internationale consacrée à l'ingénierie des exigences s'est tenue en 1993.

12. ↑DOORS pour Dynamic Object-Oriented Requirements System (https://www.ibm.com/support/knowledgecenter/fr/SSYQBZ_9.6.1/com.ibm.doors.requirements.doc/topics/c_welcome.html). Disponible dans de nombreuses langues et disposant de son propre langage de script (DXL) pour en étendre les fonctionnalités, il s'agit sans conteste de la référence actuelle dans le domaine.

13. ↑https://www.3ds.com/products-services/catia/products/reqtify/

et leurs attributs, de créer des matrices pour s'assurer de leur traçabilité ou de faciliter le travail collaboratif. En revanche, aucune aide ou recommandation ne leur est apportée lors de la phase – pourtant essentielle – de rédaction des exigences. Autrement dit, des moyens sophistiqués sont mis en place pour s'assurer que les exigences soient bien mises en œuvre, vérifiées et validées, mais aucun outil ne permet de garantir leur qualité ou de les améliorer [14]. Or, une « mauvaise » exigence, même bien gérée, reste problématique, en particulier parce que les erreurs commises en amont se répercutent sur les phases qui suivent. Le CNES ayant fait le choix d'utiliser la langue naturelle (français ou anglais, selon les cas) pour exprimer la plupart des exigences, les risques langagiers sont une réalité qu'il ne faut pas négliger.

Des solutions existent pourtant (certaines spécifiques à la rédaction des exigences, d'autres dont le champ d'application est plus général) ; les langues contrôlées en font partie. Nous nous proposons maintenant d'en dresser un panorama.

2.1.3 *Solutions existantes*

La rédaction des exigences étant une tâche à laquelle les ingénieurs sont confrontés de longue date, de nombreuses solutions, aux approches parfois radicalement différentes, ont été proposées dans les milieux professionnel et académique.

Les risques encourus lorsque la langue naturelle est utilisée à cette fin sont les mêmes que ceux que nous avons présentés au début du chapitre précédent : flou, ambiguïté [15], incomplétude. Le contexte spécifique (communication d'experts à experts, partageant des connaissances sur un domaine donné) doit cependant permettre de restreindre l'interprétation dans la plupart des cas : les termes employés dans les spécifications, sans être toujours monosémiques, sont tout de même beaucoup moins polysémiques que ceux de la conversation courante. De même, la connaissance du domaine (et du fonctionnement attendu du système décrit) permet généralement aux experts de restituer l'information laissée implicite par les rédacteurs. Cependant, les conséquences étant potentiellement beaucoup plus graves, la tolérance à l'incertitude devrait être aussi basse que possible.

Les solutions peuvent être plus ou moins contraignantes et plus ou moins efficaces par rapport à l'objectif poursuivi (éliminer l'ambiguïté, détecter l'incomplé-

14. ↑À l'exception du correcteur orthographique intégré à DOORS (qui ne vérifie pas en temps réel et n'est pas systématiquement utilisé).

15. ↑Kamsties et Paech (2000 : 3) parlent toutefois d'ambiguïté spécifique à l'ingénierie des exigences : « We call the translational ambiguities that occur due to the translation from the requirement in the customer's language to a requirement in a requirements engineer's language *RE-specific ambiguities*. RE-specific ambiguities require a background in RE and the particular project to be detected. » Le problème survient donc lorsqu'une même exigence de besoin peut donner lieu, en raison d'interprétations divergentes, à plusieurs exigences techniques.

tude[16], etc.) et il faut se résigner à admettre que la panacée – une solution totalement efficace et sans la moindre contrainte – n'existe pas. Il est donc nécessaire d'opter pour un compromis entre, d'un côté, la langue naturelle non contrôlée et, de l'autre, la formalisation pure.

Certaines des solutions utilisées donnent l'illusion de se passer totalement de la langue naturelle : c'est le cas, notamment, des diagrammes UML, qui ont fait leurs preuves pour offrir une représentation visuelle des systèmes. Force est de constater, cependant, que la langue naturelle reste indispensable pour pouvoir les interpréter correctement : même en connaissant déjà le formalisme, il reste nécessaire de nommer (textuellement) les éléments qui sont décrits. Surtout, leur expressivité est loin d'égaler celle de la langue naturelle (Glinz, 2000) et leur utilisation se cantonne à des situations récurrentes dans la pratique de spécification.

Meyer (1993) faisait très tôt remarquer les défauts de la langue naturelle et les avantages de la formalisation pour les spécifications. Il propose donc un formalisme reposant exclusivement sur les concepts des mathématiques, de la logique et de l'algorithmique et en illustre l'intérêt. Il précise toutefois d'emblée que celui-ci n'a pas vocation à remplacer les spécifications en langage naturel, mais à leur venir en complément. Bien que le bénéfice de la formalisation soit indéniable, on est en droit de se demander si cette solution peut être systématisée en contexte réel, dans la mesure où elle double la documentation à produire et à maintenir. D'autres solutions formelles ont été mises au point et se sont révélées plus adaptées, comme la notation Z (O'Regan, 2006), un langage de spécification basé sur le calcul des prédicats.

Nous avons déjà présenté les langues contrôlées et souligné leur intérêt pour la rédaction technique, qui en a favorisé l'essor ; certaines visent spécifiquement à améliorer la rédaction des spécifications et des exigences. Là aussi, le niveau de formalisme peut fortement varier : le *Guide for Writing Requirements*, que nous présenterons au chapitre 3, permet de produire des énoncés proches de ceux de la langue naturelle, tandis qu'ACE (que nous avons déjà mentionné au chapitre précédent) adopte des conventions de notation plus strictes, mais qui doivent permettre de raisonner automatiquement sur les spécifications.

Puisqu'il apparaît que les solutions formelles et les solutions plus naturelles ont toutes deux leurs avantages et leurs inconvénients, certaines approches proposent de les combiner (Lamsweerde, 2001 ; Naumchev et Meyer, 2017). Il peut aussi s'agir de modéliser automatiquement des spécifications ou des cas d'utilisation afin de pouvoir raisonner dessus, et ainsi offrir une aide à la rédaction en facilitant la détection d'anomalies ou d'erreurs (Charnois, 1999 ; Nair et Somé, 2006).

16. ↑Étant entendu que la complétude absolue n'est pas atteignable, puisque tout devrait être défini et spécifié, il s'agit surtout de détecter l'incomplétude là où elle peut l'être, grâce à la modélisation du domaine notamment (Zowghi et Gervasi, 2002).

Certaines solutions sont dites semi-formelles. Ainsi, lorsque le même type d'exigence doit être exprimé de façon répétée, il devient de plus en plus courant de recourir aux *boilerplates* [17], c'est-à-dire à des modèles d'exigences ou de morceaux d'exigences prédéfinis, composés d'éléments fixes et d'éléments variables (les attributs) qui seront remplacés par des valeurs textuelles dans les exigences réelles. Par exemple, le *boilerplate* élémentaire *Le <système> doit <action>* pourra être instancié en *Le véhicule doit respecter les limitations de vitesse*. Ils prennent tout leur intérêt lorsqu'ils sont associés à une ontologie du domaine, qui permet de déterminer quelle valeur peut être affectée à un attribut (Farfeleder et al., 2011b). Nous reviendrons sur les boilerplates au chapitre 4.

Enfin, des outils et des méthodes automatisées toujours plus nombreux ont été mis au point pour assister les rédacteurs d'exigences dans leur tâche. Leur objectif est de détecter et signaler les incohérences (Saint-Dizier, 2018) ou les risques langagiers (Lami et al., 2004 ; Carlson et Laplante, 2014 ; Femmer et al., 2014) dans les exigences rédigées en langue naturelle grâce au traitement automatique du langage. Ils interviennent donc *a posteriori*, pour vérifier la qualité rédactionnelle de ces exigences. Certains se focalisent sur quelques phénomènes bien précis (Landhausser et al., 2015), tandis que d'autres se veulent aussi complets que possible (Sateli et al., 2012). Ils sont particulièrement utiles pour s'assurer du respect des règles de la langue contrôlée utilisée par les rédacteurs. À cet effet, certains peuvent être adaptés au référentiel choisi (Barcellini et al., 2012 ; Génova et al., 2013), mais quelques-uns sont paramétrés en fonction d'une langue contrôlée en particulier. C'est le cas pour l'anglais simplifié : plusieurs outils de vérification automatique ont été proposés (dont un par l'avionneur Boeing, le *Simplified English Checker* [18]), mais – curieusement – le groupe de maintenance de l'anglais simplifié, sans les désapprouver complètement, met sérieusement en garde contre leur utilisation [19].

2.1.4 *Conclusion*

L'ingénierie des exigences a pour objectif d'améliorer l'élucidation et la gestion des exigences, deux étapes cruciales pour la réussite des projets de grande envergure. La qualité de ces exigences, c'est-à-dire leur capacité à exprimer clairement le besoin qu'elles représentent est également fondamentale, comme en témoigne le nombre élevé d'études récentes consacrées à la question.

Les solutions proposées pour améliorer les exigences se répartissent entre celles qui entendent contraindre ou modifier directement l'expression du besoin et celles

17. ↑Apparu il y a une quinzaine d'années en ingénierie des exigences, le terme est aussi utilisé en informatique (*boilerplate code*) et en droit des contrats (*boilerplate language*) avec un sens similaire (Tommila et Pakonen, 2014). L'expression *template* est également fréquemment utilisée : plus transparente, elle a néanmoins l'inconvénient de revêtir un sens légèrement différent selon l'application (par exemple en extraction d'information).

18. ↑https://www.boeing.com/company/key-orgs/licensing/simplified-english-checker.page

19. ↑« Software will not think in your place. » (http://www.asd-ste100.org/software.html).

qui interviennent dans un second temps pour signaler automatiquement aux rédacteurs les problèmes potentiels (et en particulier les manquements aux règles énoncées par les premières), deux approches qui sont évidemment complémentaires. Parmi le premier type de solutions, il faut encore distinguer entre les plus « naturalistes » et les plus formalistes – au milieu desquelles on trouve d'autres propositions, intermédiaires ou hybrides.

Notre travail se situe plutôt en amont, notre objectif étant de définir des règles de rédaction qui permettent d'améliorer la qualité des exigences au CNES (et qui, dans cette même optique, pourront par la suite être intégrées dans un outil de vérification automatique, existant ou à concevoir). Rappelons qu'à l'heure actuelle, aucune solution de ce type n'est utilisée de façon systématique par les ingénieurs du CNES, ce qui signifie qu'ils rédigent librement en langue naturelle non contrôlée. Par conséquent, et parce que nous sommes convaincu que pour avoir une chance d'être effectivement adoptée, notre proposition se doit de rester suffisamment proche de la langue naturelle et des habitudes de rédaction des ingénieurs, une langue contrôlée nous paraît être la solution la plus adaptée pour réduire les risques langagiers.

Bien que nous soyons également tout à fait favorable à l'utilisation de méthodes formelles ou semi-formelles, qui peuvent aider à détecter des propriétés indésirables telles que l'incohérence ou l'incomplétude, éliminer totalement la langue naturelle ne nous semble ni possible, ni souhaitable. En effet, ainsi que le font remarquer Kamsties et Paech (2000 : 2), « In practice, [. . .] stakeholders refrain from reading requirements written in artificial languages. » La langue naturelle reste incontestablement la solution la plus facile à utiliser par toutes les parties prenantes, au moins pour certaines étapes du processus de spécification. Nous pensons du reste qu'une solution n'exclut pas l'autre, et que notre langue contrôlée pourrait inclure une série de *boilerplates*, qui présentent le double avantage d'être simples à utiliser et de garder un aspect assez naturel.

2.2 VERS LA CONSTITUTION D'UNE LANGUE CONTRÔLÉE BASÉE SUR CORPUS

Dans cette section, nous essayerons de montrer comment la linguistique de corpus peut contribuer à améliorer la rédaction des exigences au CNES en prenant en compte les usages réels, dans l'optique de proposer une langue contrôlée adaptée aux modes de rédaction des ingénieurs. Nous exposerons pour cela la méthodologie que nous avons mise au point et qui sera appliquée à un corpus d'exigences authentiques (Warnier, 2015 ; Warnier et Condamines, 2016). Les premiers résultats obtenus seront quant à eux présentés dans la deuxième partie de cette thèse.

Néanmoins, en guise d'introduction à la question de la rédaction des exigences au CNES et pour montrer l'intérêt d'analyser des productions réelles, nous présenterons tout d'abord une (modeste) étude qui adopte un point de vue différent de celui qui sera le nôtre par la suite. En effet, notre méthodologie repose sur l'analyse

de productions langagières en synchronie, puisque les exigences qui constituent notre corpus sont issues des versions finales des spécifications (cf. section 2.2.3). Dans les faits, pourtant, ces exigences sont amenées à évoluer dans le temps, au gré des révisions, parfois nombreuses, qui sont effectuées sur les spécifications. C'est cette dimension temporelle qui sera prise en compte dans cette brève étude préliminaire.

2.2.1 *Premier aperçu de la rédaction des exigences au CNES*

Nous nous sommes intéressé à l'évolution des spécifications rédigées au CNES au fil des révisions dont elles font régulièrement l'objet. Nous pensons qu'une comparaison manuelle des versions successives d'une même spécification peut offrir un bon premier aperçu de la façon dont les exigences sont écrites, puis révisées par l'équipe qui en est responsable. En effet, si une formulation a été modifiée par rapport à celle que l'on trouvait dans la version précédente, nous avons toutes les raisons de penser [20] que cette dernière a, entretemps, été identifiée comme problématique ou porteuse d'un risque, et que la nouvelle version est considérée, de ce point de vue, comme « meilleure ». Ces changements de formes langagières nous permettent donc (indirectement) d'en savoir un peu plus sur les pratiques des ingénieurs du CNES en la matière.

Pour ce faire, nous avons obtenu du CNES les versions successives de trois spécifications différentes, toutes liées à un même projet (Pléiades). Ces spécifications comptent respectivement douze, quatorze et vingt-trois versions différentes, ce qui correspond à autant de documents (que nous avons comparés deux par deux). Pour l'une d'entre elles, les deux premières versions étaient rédigées en anglais [21], avant qu'elle ne soit intégralement traduite en français à partir de la troisième version [22].

Le principal constat que l'on peut poser est que les modifications concernant directement les formes langagières sont proportionnellement très peu nombreuses (en dehors bien sûr du changement majeur qu'a constitué la traduction en français d'une des versions), la plupart portant sur d'autres éléments : mise en forme, ajout ou suppression d'exigences, etc. Il ne nous est bien sûr pas possible de savoir si ce faible nombre de révisions linguistiques s'explique par un manque d'attention porté aux formulations, ou simplement parce que ces dernières sont jugées satisfaisantes en l'état. On constate néanmoins plusieurs corrections portant sur des

20. ↑Malheureusement, le motif du changement n'étant jamais précisé dans les documents dont nous disposons, nous ne pouvons faire à ce propos que des suppositions.

21. ↑Dans une langue parfois approximative ; par exemple : « A message shall be informed [sic] the operator if the connection is impossible because the system doesn't recognize the operator (login), or the password is not valid », traduit par la suite en « Un message doit informer l'opérateur si la connexion est impossible parce que le système ne reconnaît pas l'opérateur (login) ou parce que le mot de passe n'est pas valide », ce qui confirme que la voix passive n'avait pas lieu d'être utilisée. Remarquons que le modal propre à l'anglais « shall » a été traduit ici par « doit ».

22. ↑« CAUSE D'EVOLUTION : [...] Traduction en français [...] et correction du style de certaines exigences. [...] »

erreurs de langue (en particulier des accords) – ce qui n'empêche pas plusieurs autres erreurs, plus ou moins évidentes, de subsister.

Une observation intéressante est que les balises « T£ » et « £T » (marquant, au sein de la spécification, le début et la fin d'une exigence) peuvent être amenées à être déplacées, ajoutées ou supprimées lors d'une révision. Ceci montre bien que le statut d'« exigence » (par opposition aux simples commentaires, utiles mais théoriquement non contraignants), bien que fondamental, ne va pas totalement de soi. Ainsi, par exemple, dans le fragment de texte suivant [23] :

> Effet de la TC dans le logiciel :
> N£ PHR-IF-1/6-11-CNES_112310 £N
> T£ Sur réception de cette TC, le logiciel programme la commande demandée sur la trame 1553. Cette commande sera envoyée à l'équipement sur le cycle 8 Hz suivant.
> Le nombre de paramètres est lié au type de la commande. Il n'excède pas 64 bytes. £T
> Pour la commande « UPLOAD MEMORY », la taille, en octet, des données fournies en paramètre doit être paire.

On constate que, par la suite, la balise fermante « £T » a été déplacée à la fin du paragraphe (soit après « doit être paire »), incluant de fait la dernière phrase dans l'exigence et modifiant ainsi son statut. Le cas inverse est également possible :

> Contraintes CCC :
> N£ PHR-IF-1/6-11-CNES_16020 £N
> T£ Cette TC sera générée uniquement au DUPC, donc non applicable au CCC. £T

Ici, les balises « N£ » et « £N » et l'identifiant qu'elles encadrent ont été supprimés, de même que les balises « T£ » et « £T ». Quant au texte restant (« Cette TC… »), il a été conservé, mais s'est vu relégué au rang de simple remarque.

Pour le reste, on peut classer les modifications linguistiques que nous avons repérées dans trois catégories principales. La première de ces catégories concerne la façon d'insister sur le caractère « injonctif » des exigences (sur lequel nous reviendrons plus longuement au chapitre 3). Les modifications vont très clairement dans le sens d'un renforcement de l'obligation. Par exemple, l'exigence :

> Il est souhaitable que toute l'informatique bord (calculateur bord et systèmes décentralisés exécutant des logiciels) soit présente dans le BOSS sous forme de maquettes fonctionnelles.

23. [↑]Pour chacun des extraits retranscrits, nous avons fait le choix de recopier le texte à l'identique. Les éventuelles erreurs ne sont donc pas corrigées. Un glossaire des spécifications CNES est donné à l'annexe A.

qui laissait penser que cette caractéristique n'était pas obligatoire (mais simplement « souhaitable ») a été remplacée par une formulation sans équivoque :

> Toute l'informatique bord doit (calculateur bord et systèmes décentralisés exécutant des logiciels) être présente dans le BOSS sous forme de maquettes fonctionnelles.

où « Il est souhaitable que ... soit ... » a laissé place à « ... doit être ... ». Idem avec « pouvoir » :

> Les opérations de téléchargement des coefficients d'égalisation à bord (LVC) peuvent se faire sans entraîner d'indisponibilité des fonctions de prise de vue de la charge utile.

là aussi remplacé par « devoir » :

> Les opérations de téléchargement des coefficients d'égalisation à bord (LVC) doivent se faire sans entraîner d'indisponibilité des fonctions de prise de vue de la charge utile.

« Pouvoir » et « devoir » peuvent coexister, par exemple quand il s'agit de lister des possibilités (ici, hors exigence) :

> Les surveillances standards peuvent être : [...]

modifié ensuite en :

> Les surveillances standards doivent pouvoir être : [...]

La deuxième catégorie concerne le remplacement d'une forme par une autre, jugée moins ambiguë. Les deux exemples que nous avons trouvés se rapportent à des déterminants :

> Ce document présente la définition des télécommandes à destination du satellite Pléiades ainsi que de ces contraintes d'émission.

a été reformulé en :

> Ce document présente la définition des télécommandes à destination du satellite Pléiades ainsi que de leurs contraintes d'émission.

Le démonstratif « ces », assez flou (à quoi renvoyait-il ?), a été remplacé par le possessif « leurs ». Les « contraintes d'émission » dont il est question sont donc, plus clairement, celles des « télécommandes à destination du satellite Pléiades ». De la même façon, dans l'exigence :

> Dans le cas ou le LVC a détecté des secteurs défaillants sur une (ou plusieurs tranches), le sol doit dumper cette table pour l'analyser et en extraire les numéros de secteurs défaillants.

révisée en :

> Dans le cas ou le LVC a détecté des secteurs défaillants sur une (ou plusieurs tranches), le sol doit dumper la (ou les) tables pour les analyser et en extraire les numéros de secteurs défaillants.

C'est à nouveau un démonstratif (« cette ») qui a été remplacé, cette fois par les définis « la (ou les) » (selon qu'il y a une ou plusieurs tables). Une raison pouvant expliquer ce changement est que « table » n'apparaissait pas auparavant – ce qui n'empêche pas le syntagme « la (ou les) tables » de rester anaphorique. Par conséquent, le sens exact de la phrase nécessite sans doute des connaissances du domaine (les tables contiennent les numéros de secteurs et sont peut-être aussi liées aux tranches).

Enfin, la dernière catégorie regroupe tous les changements visant une plus grande précision ou clarté. Ils peuvent prendre différentes formes, notamment le choix d'un terme plus spécifique :

> Les logiciels décentralisés nécessitant une datation bord devront entretenir le temps bord qu'ils reçoivent du LVC [...].

reformulée en :

> Les logiciels décentralisés nécessitant une datation en Heure Bord du LVC devront entretenir l'heure bord qu'ils reçoivent du LVC [...].

Ici, la structure elliptique « datation bord » est remplacée par « datation en heure bord », et « temps bord » par « heure bord » (ce qui n'empêche pas que « temps bord » apparaisse encore ailleurs dans le document). La précision « du LVC » est également ajoutée, quitte à créer une redondance avec ce qui suit.

> Les valeurs des paramètres MIN et MAX seront renseignées en valeurs physiques.

Dans cette exigence, « renseignées » est remplacé par « saisies ».

Parfois, une information complémentaire est ajoutée :

> Les format des commandes 1553 à destination du chiffreur/déchiffreur bord (DCU) ne seront pas toutes détaillées.

où « en BDS » est ajouté à la fin, rendant l'exigence plus complète.

> Aucun contrôle n'est demandé sur les champs de cette TC, ce sera sous responsabilité opérationnelle.

où « (dans le générateur de TCH) » est ajouté après « contrôle ».

> Si une «TC liste surveillances standards» est reçue alors qu'une autre TC de ce type est en cours d'exécution, le précédent liste est annulée.

où « le précédent liste » [sic] est remplacé par « la TC 'liste' précédente », précisant ainsi la nature de la liste en question (une TC).

> Le type immédiat/différé avec HB [...]

où le sigle HB a été développé : « Heure Bord (HB) ».

Ces quelques exemples de révision portant sur la langue (certes moins nombreux que nous l'avions espéré) attestent que les rédacteurs ont conscience qu'il est possible – et parfois même nécessaire – de reformuler les exigences pour les rendre plus claires. Certains risques langagiers sont perçus (et éliminés) : c'est le cas, par exemple, de l'ambiguïté liée aux déterminants anaphoriques.

2.2.2 *Propositions méthodologiques*

En section 1.5.4, nous avions signalé le manque d'évaluation systématique des langues contrôlées et dit en quoi il est problématique. En effet, des langues contrôlées sont définies parfois sans assurance que leurs règles soient réellement efficaces : peut-être le sont-elles, peut-être ne le sont-elles pas, peut-être pourraient-elles l'être encore plus si elles étaient mieux définies. Cette situation est d'autant plus préjudiciable que les enjeux sont importants. Notre postulat est que s'intéresser aux usages existants ainsi qu'aux utilisateurs à qui elles s'adressent doit permettre de mieux les adapter à leur situation. Nous souhaiterions également proposer des recommandations langagières fondées sur une vraie réflexion linguistique (qui, en retour, enrichirait notre connaissance de la langue en contexte spécialisé).

Aucune langue contrôlée n'étant actuellement imposée au CNES, nous sommes libre de définir la nôtre comme nous le souhaitons. Il nous semble néanmoins indispensable de tenir compte de l'état de l'art en la matière, c'est-à-dire des recommandations récurrentes parmi les langues contrôlées pour la rédaction technique. Si celles-ci sont populaires, c'est sans doute parce qu'elles sont perçues par les experts comme une solution à des problèmes concrets, tels que l'ambiguïté (lexicale, syntaxique, référentielle), l'imprécision ou encore le risque d'incomplétude, qui peuvent ou non avoir été identifiés précisément.

Le meilleur moyen de comprendre les usages est de les observer directement. Pour ce faire, nous avons la chance de disposer, à travers les spécifications du

CNES, de textes écrits authentiques (c'est-à-dire rédigés et utilisés en contexte réel) au format numérique. Nous proposons de partir de ces spécifications pour constituer un corpus d'exigences sur lequel effectuer des analyses préalables à la définition de règles langagières. Les corpus (qui doivent être pris pour ce qu'ils sont : leur représentativité ne peut jamais être garantie) permettent en effet d'objectiver des intuitions sur le fonctionnement de la langue, ou d'une variété de la langue – ce qui est exactement ce que nous cherchons à faire. Travailler sur corpus, c'est s'intéresser aux usages attestés et à leur fréquence, sans devoir s'en remettre au jugement – subjectif – des locuteurs de cette langue. En d'autres termes, le corpus permet d'accéder à la norme objective.

Nous disposons aujourd'hui des outils nécessaires pour constituer facilement et exploiter pleinement les corpus langagiers. Les corpus s'appréhendent en effet avant tout en fonction de l'objectif (linguistique) auquel ils sont destinés et par les critères (extralinguistiques) qui ont déterminé leur constitution. En raison sans doute de leur succès, de nombreuses définitions ont été proposées [24]. Pour Tognini-Bonelli, un *corpus* est

> [...] a computerised collection of authentic texts, amenable to automatic or semi-automatic processing or analysis. The texts are selected according to explicit criteria in order to capture the regularities of a language, a language variety or a sub-language (Tognini-Bonelli, 2001 : 55).

Cette définition justifie pleinement de fonder notre méthodologie de constitution d'une langue contrôlée sur un corpus, puisque nous souhaitons précisément identifier les régularités qui nous permettront, après les avoir analysées, d'établir des règles langagières. Or, la linguistique de corpus a permis de montrer que ces régularités variaient d'un genre à l'autre :

> One major contribution of the corpus-based approach is that it establishes the centrality of register for descriptions of language use. That is, corpus-based research has shown that almost any linguistic feature or variant is distributed and used in dramatically different ways across different registers. Taken together, corpus-based studies challenge the utility of general linguistic descriptions of a language; rather, these studies have shown that any linguistic description that disregards register is incomplete or sometimes even misleading. (Biber, 2010 : 163)

Les propriétés linguistiques d'un genre ne sont donc pas (nécessairement) partagées par d'autres genres ; ce qui se constate le plus facilement au niveau lexical (encore plus lorsque l'on dispose d'outils statistiques), mais qui est également valable aux autres niveaux linguistiques. Par corollaire, une règle adaptée à un genre risque de ne pas être adaptée à un autre. Ceci accepté, l'idée d'une langue contrôlée « universelle », efficace en toutes situations, perd grandement de sa force. Bien

24. [↑]Ces définitions sont généralement plutôt concordantes entre elles, bien qu'il puisse y avoir débat pour savoir si telle ou telle collection de textes est ou non un corpus.

entendu, cela ne signifie pas qu'il ne peut exister de recommandations suffisamment générales pour être valables dans la plupart des contextes. Mais il y aura toujours un intérêt à tenir compte des spécificités du genre duquel relèvent les usages qui nous occupent. Dans l'optique de proposer une langue contrôlée, il peut s'agir de proposer de nouvelles règles, d'en éliminer d'autres ou tout simplement de les raffiner.

Au-delà des résultats obtenus en analysant notre corpus d'exigences, que nous présenterons dans la seconde partie de cette thèse et que nous voulons voir comme une preuve de faisabilité, l'apport de notre travail est donc surtout méthodologique. Nous pensons que la linguistique de corpus est en mesure de fournir aux concepteurs de langues contrôlées des arguments objectifs pour proposer des règles plus finement adaptées, et qui ont donc plus de chances d'être adoptées. La linguistique de corpus elle-même peut être vue comme une méthodologie :

> Corpus linguistics is not in itself a model of language. In fact, at one level it can be regarded as primarily a methodological approach:
> - it is empirical, analyzing the actual patterns of use in natural texts;
> - it utilizes a large and principled collection of natural texts, known as a "corpus", as the basis for analysis;
> - it makes extensive use of computers for analysis, using both automatic and interactive techniques;
> - it depends on both quantitative and qualitative analytical techniques (Biber et al. 1998: 4). (Biber, 2010 : 159-162).

Notre idée peut donc se résumer simplement comme suit : constituer un corpus à partir d'usages réels, puis l'exploiter pour déterminer les caractéristiques linguistiques (les régularités) du genre représenté par le corpus (s'il existe) pour, enfin, proposer des règles adaptées à ces caractéristiques (qu'il s'agisse de les favoriser ou de les contraindre). La difficulté est donc de savoir comment exploiter le corpus, c'est-à-dire comment découvrir des propriétés qui ne sont pas évidentes.

Pour analyser le corpus, la tradition distingue deux approches – qui, en pratique, se complètent sans doute plus qu'elles ne s'opposent. L'une, plutôt descendante et déductive, est nommée *corpus-based* ; l'autre, ascendante et inductive, est nommée *corpus-driven* :

> Corpus-based research assumes the validity of linguistic forms and structures derived from linguistic theory; the primary goal of research is to analyze the systematic patterns of variation and use for those pre-defined linguistic features. [...]
>
> In contrast, "corpus-driven" research is more inductive, so that the linguistic constructs themselves emerge from analysis of a corpus. [...] The corpus-driven approach differs from the standard practice of linguistics in that it makes minimal a priori assumptions regarding the linguistic features that should be employed for the corpus analysis. (Biber, 2010 : 162)

Pour simplifier, on peut donc dire que l'approche *corpus-based* (sans doute la plus classique en linguistique) consiste à formuler des théories (hypothèses), puis à les vérifier sur le corpus, tandis que l'approche *corpus-driven* consiste à faire émerger directement les régularités du corpus, sans apriori. La première permet ainsi au chercheur d'apporter directement une réponse aux questions qu'il se pose, alors que la seconde lui permet de découvrir des régularités qu'il n'aurait peut-être pas suspectées. La première nécessite des hypothèses ou des intuitions, la seconde une quantité plus grande de données langagières.

Toutes deux ont leur intérêt pour la définition d'une langue contrôlée. Pour l'approche *corpus-driven*, il s'agit surtout de disposer des outils adaptés et de déterminer comment exploiter les résultats qu'ils fournissent. Ceux-ci peuvent par exemple identifier les traits ou les structures fréquents, le chercheur se chargeant ensuite de déterminer s'ils doivent être encouragés ou au contraire proscrits (parce que porteurs d'un risque langagier) à l'avenir.

Ces structures (au sens large) peuvent typiquement relever de la phraséologie, une notion aussi vaste qu'intéressante. Les nombreuses études qui ne cessent de paraître sur ce sujet, loin de l'épuiser, ne font qu'enrichir une terminologie déjà foisonnante : Bendinelli (2017) signale que l'étiquette « fait phraséologique » renvoie à plus de soixante catégories opératoires, qui se définissent selon des critères différents. Issues de courants théoriques très variés, celles-ci « ont en commun de postuler l'existence de structures pré-formées combinant deux (ou plus) unités. » (Bendinelli, 2017 : 2) Les éléments constitutifs des faits phraséologiques sont souvent unis par une relation orientée : le locuteur progresse de la base vers le collocatif (ainsi avec « célibataire » et « endurci », par exemple) (Bendinelli, 2017 : 2).

Si ces unités suscitent autant d'intérêt, c'est aussi parce qu'elles représentent de sérieux défis, notamment en traitement automatique du langage (où elles sont fréquemment nommées *multiword-expressions* (MWEs)) (Heid, 2008), mais plus que tout en traduction et pour les apprenants étrangers. Pour l'activité de traduction, Heid et Freibott (1991 : 78) avancent que la présence de collocations [25] est « l'un des éléments principaux de la "lisibilité" d'un texte, en langue générale ou technique », mais aussi une difficulté majeure (surtout pour la traduction *vers* une langue étrangère) puisque ces associations préférentielles doivent être apprises, tout comme les termes. Malheureusement, les ressources consacrées aux collocations (ou à d'autres unités phraséologiques) ont longtemps été peu nombreuses. Plusieurs initiatives sont depuis lors venues combler ce manque (Todiraşcu et al., 2008), dont ARTES (Kübler et Pecman, 2012) qui propose [26] un dictionnaire multilingue pour les langues de spécialité incluant des collocations et d'autres unités phraséologiques (par exemple, « to gain considerable attention »). Les phra-

25. [↑]Sans doute les unités phraséologiques les plus étudiées, elles se limitent normalement à deux lexèmes.

26. [↑]En plus d'une base de données terminologiques et phraséologiques (à accès restreint) qui tente d'intégrer des phénomènes complexes à représenter, tels que la prosodie sémantique et la préférence sémantique.

séologismes posent évidemment aussi des problèmes aux apprenants de langue seconde (d'où l'apparition de la « phraséodidactique »). Et pourtant :

> [...] mastering phraseology is one of the proofs of belonging to a discourse community; in particular, it is one of the proofs of belonging to a scientific community (Gledhill 2000). (VOLANSCHI et KÜBLER, 2010 : 345)

Leur non-maîtrise est donc fortement préjudiciable pour les chercheurs non anglophones. Les dictionnaires et autres ressources électroniques prenant en compte la phraséologie sont dès lors d'une aide précieuse pour la rédaction de publications en anglais (VOLANSCHI et KÜBLER, 2010).

Les unités phraséologiques sont omniprésentes : « [...] corpus analysis has demonstrated time and time again that regularities of expression can always be found in any type of text (as long as a representative corpus can be analysed) » (GLEDHILL et KÜBLER, 2016 : 11). À côté des séquences idiomatiques pragmatiquement marquées et facilement identifiables (telles que les proverbes) qui relèvent de ce qu'ils nomment l'approche *phrase-oriented*, GLEDHILL et KÜBLER (2015) soulignent l'importance, au sein des genres et des types de texte, d'unités d'apparence plus discrète et auxquelles s'intéresse l'approche *pattern-oriented*, telles que les *collocational frameworks*, les *clusters* et les motifs lexicaux. Pour Gledhill, en effet, certaines collocations (les *collocations génériques*), de par leur régularité même, sont intimement liées à un genre de texte et à une communauté de discours :

> Collocations appear to confim the existence of a discourse community. Their very consistent nature suggests that collocations have a central role to play in discourse [...] The regularity and pervasive nature of collocation appears to be incompatible with the intuition that an individual's use of language is inherently unique and creative. [...] it is likely that the collocational coherence of a text is an acquired characteristic derived by the writer from wide reading and sub-conscious attempts to conform to the norm of speech [...] (GLEDHILL, 2000 : 205)

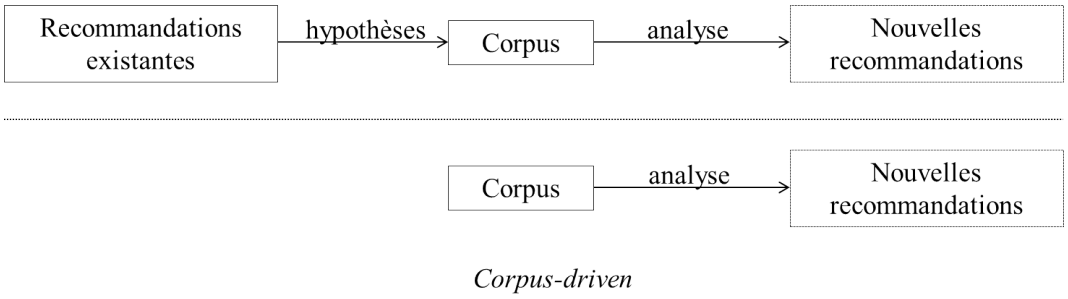
La profusion de catégories liées aux phraséologismes montre bien qu'il s'agit d'une réalité complexe, qui peut être appréhendée sous plusieurs angles, et que le statut de ces unités dépend notamment de l'application que l'on souhaite leur donner. On ne peut en tout cas que se réjouir du développement d'outils facilitant leur extraction automatique (HEID, 2001), qu'ils se basent sur des critères syntaxiques et/ou statistiques (avec des mesures telles que l'information mutuelle).

Pour l'approche *corpus-based*, en revanche, c'est d'abord au chercheur de décider quelles structures il souhaite étudier. Il nous semble alors évident que les structures à étudier en premier sont celles qui relèvent de phénomènes envisagés par les langues contrôlées déjà existantes. Ainsi, comme nous le disions au début de cette section, il devient possible de tenir compte à la fois de l'« état de l'art », c'est-à-dire des règles récurrentes, et des usages réels, auxquels on les confronte pour en vérifier la pertinence. La démarche peut ensuite aller plus loin : fort de sa

connaissance du fonctionnement de la langue, le linguiste peut proposer d'autres hypothèses et les vérifier à leur tour.

Dans un cas comme dans l'autre, le corpus est au centre de l'analyse (FIGURE 2), qui profite aux recommandations (et *in fine* à leurs utilisateurs) ; ce qui diffère, c'est la façon de procéder et les ressources nécessaires. Ce constat souligne que les deux approches sont bel et bien complémentaires, les résultats de l'une enrichissant ceux de l'autre. Précisons que cette analyse est rendue possible et facilitée par les outils qui sont à sa disposition, mais qu'au final, c'est bien au chercheur qu'il appartient d'en tirer les conclusions qui permettront d'améliorer les recommandations.

FIGURE 2 – Approches *corpus-based* et *corpus-driven*



Nous utiliserons les approches *corpus-based* et *corpus-driven* pour analyser notre corpus d'exigences aux chapitres 3 et 4, respectivement. Par la même occasion, cela nous permettra de vérifier notre hypothèse de l'existence d'un genre des exigences.

Enfin, une fois de nouvelles propositions de règles formulées, ou pour évaluer les règles existantes, il devrait être possible d'en vérifier l'intérêt et l'efficacité auprès d'un échantillon des utilisateurs finaux. Cette évaluation peut prendre plusieurs formes, qui peuvent être plus ou moins complexes à mettre en place ; nous en donnerons un exemple au chapitre 5, où nous rendons compte d'un dispositif permettant de recueillir le jugement d'experts du CNES sur des exigences formulées selon des recommandations trouvées dans des langues contrôlées actuelles.

Cette méthodologie est suffisamment générale pour pouvoir être reproduite dans de nombreux autres contextes : elle peut en effet être déclinée différemment selon les objectifs poursuivis et les ressources dont on dispose.

2.2.3 *Corpus d'étude*

Notre méthodologie reposant sur l'analyse d'un corpus d'exigences, nous avons constitué le nôtre à partir des spécifications qui nous ont aimablement été transmises par le CNES après concertation. Ce corpus se compose lui-même de deux

sous-corpus. Le premier est constitué d'exigences extraites de spécifications du projet Pléiades et le second d'exigences rédigées pour le projet Microscope [27].

Par choix, nous nous limitons dans cette étude aux exigences rédigées en français – certaines exigences du CNES étant rédigées en anglais, par volonté d'internationalisation, notamment lorsque les parties prenantes ne sont pas toutes francophones. Si notre corpus ne se compose que d'exigences (excluant ainsi le texte présent dans les spécifications en dehors de ces exigences, qui comporte entre autres choses des titres et des commentaires), c'est parce que nous pensons qu'elles sont beaucoup plus homogènes et plus susceptibles de représenter un genre textuel. Les exigences étant de loin les éléments les plus importants de la spécification (puisque ce sont elles qui constituent un engagement et qui devront être vérifiées), c'est sur celles-ci que se concentrent logiquement tous les efforts de rédaction (le reste étant considéré comme moins critique – même si, bien sûr, la cohérence de la spécification est importante) et ce sont donc spécifiquement les exigences qui sont visées par les recommandations de la langue contrôlée que nous entendons concevoir (ce qui n'interdit évidemment pas d'appliquer ces dernières aux spécifications dans leur ensemble).

Pléiades est un projet de grande ampleur qui consiste en deux satellites (Pléiades 1A et Pléiades 1B, lancés respectivement en 2011 et 2012) destinés à fournir chaque jour des centaines d'images de la Terre en très haute résolution. Microscope est un projet comparativement plus modeste (un microsatellite, lancé en 2016) dont l'objectif est avant tout scientifique, puisqu'il vise à vérifier le principe d'équivalence depuis l'espace avec une précision encore jamais atteinte [28]. Il s'agit donc dans les deux cas de projets de systèmes orbitaux, qui diffèrent cependant par leur taille, leur objectif et leur domaine secondaire (l'optique pour l'un, la physique pour l'autre).

Les spécifications sur lesquelles nous avons travaillé (qui sont des versions finales) se présentent sous la forme de documents Microsoft Word, généralement longs de plusieurs centaines de pages et contenant de nombreuses informations liminaires (titre du document, autorisations de diffusion, auteurs, résumé, historique des modifications, glossaire, etc.), suivies des exigences, sur lesquelles nous avons décidé de nous concentrer. Les documents ne suivent pas forcément tous strictement le même format (des variations de mise en page assez significatives peuvent être constatées), mais contiennent toujours ces informations principales. Les fonctionnalités de traitement de texte (feuilles de style, numérotation, en-têtes et pieds de page, etc.) sont mises à profit pour faciliter la consultation et la modification du document, qui est dès lors particulièrement riche en informations et

27. ↑Nous avions initialement envisagé d'inclure des exigences de la mission Taranis (un satellite pour l'observation des orages depuis l'espace, dont le lancement est prévu en 2019), mais le trop faible nombre d'exigences en français (environ deux cents, soit à peine plus de cinq pour cent de la taille actuelle du corpus) nous en a finalement dissuadé. La plupart des exigences de ce (petit) projet sont en anglais.

28. ↑Les premiers résultats, qui confirment le principe d'équivalence, ont été publiés fin 2017.

s'éloigne donc considérablement du texte brut.

Les exigences y sont hiérarchisées à l'aide de titres (ex. : « Télémesures du mode normal ») et de sous-titres (ex. : « TM de calcul des manœuvres ») et toutes sont précédées d'un identifiant unique (ex. : « PHR-IF-1/6-11-CNES_122900 »), lequel est entouré des caractères *N£* et *£N*, qui permettent de le repérer. Le contenu de l'exigence est quant à lui délimité par les caractères *T£* (au début) et *£T* (à la fin) ; il s'agit essentiellement de texte (mis en forme si nécessaire), de listes à puces, de tableaux, d'images (schémas) et plus rarement d'algorithmes – ces différents formats pouvant très bien se retrouver au sein d'une même exigence. L'extrait suivant est tiré de l'une des spécifications Pléiades :

> Facilités d'exploitation et interfaces
>
> *N£ PHR-ST-0-12-CNES_17200 £N*
>
> T£ Le simulateur sera interfacé au centre de contrôle du satellite suivant un protocole TCP/IP. Il reçoit du centre de contrôle des paquets de télécommande et émet des paquets de télémesure. £T
>
> *N£ PHR-ST-0-12-CNES_17300 £N*
>
> T£ Il peut être également utilisé en autonome, les TC étant émises en local. £T

Pour pouvoir constituer un corpus exploitable automatiquement, nous avons dû convertir les spécifications en texte brut, ce qui signifie que nous avons perdu tous les éléments de mise en forme (couleurs, grasses, italique, etc.) et les images. Nous avons également choisi d'en retirer automatiquement tous les tableaux à l'aide d'une macro, parce que les outils de traitement automatique du langage que nous utiliserons ne sont pas conçus pour traiter des informations tabulaires, qui ne s'interprètent pas comme du texte linéaire. Leur présence dans le corpus aurait constitué du bruit qui aurait partiellement faussé les analyses.

Grâce aux balises *T£* et *£T* [29] et à l'aide d'un script *ad hoc*, il est aisé de récupérer automatiquement le contenu des exigences et d'éliminer le reste ; ainsi, dans l'extrait précédent, nous n'avons conservé que « Le simulateur [...] de télémesure » et « Il peut [...] en local ». Par conséquent, le corpus est constitué de la concaténation des exigences présentes dans les spécifications dont nous disposons (lesquelles peuvent être considérées comme de courts textes).

Pour des raisons de confidentialité évidentes, il ne nous est pas permis de le reproduire *in extenso*. Nous en citerons donc des extraits chaque fois que cela est pertinent. Tous les extraits sont repris à l'identique : nous ne corrigeons pas les

29. ↑Il arrive – très rarement – que la balise ouvrante ou la balise fermante soit manquante. Nous avons détecté les balises orphelines automatiquement pour corriger ce problème.

erreurs d'orthographe qui peuvent s'y trouver (et qui peuvent perturber les outils de TAL les moins robustes). Toute troncation au sein d'une exigence sera indiquée comme il se doit par [...].

Le TABLEAU 1 indique, pour chacun des deux sous-corpus, le nombre d'exigences, le nombre d'occurrences et le nombre de types différents. Pléiades étant un plus gros projet que Microscope, nous avons récupéré beaucoup plus d'exigences (malgré un nombre identique de spécifications, quinze), ce qui signifie également plus du double de mots.

TABLEAU 1 – Composition du corpus

Corpus	**Nombre d'exigences**	**Nombre d'occurrences**	**Nombre de types**
Pléiades	2 528	118 036	5 944
Microscope	1 067	43 367	3 020
Total	**3 595**	**161 403**	**7 211**

Parmi les exigences, certaines présentent de fortes ressemblances (par exemple, « Cette TC permet de mettre ON la carte HKOBS de l'OBMU. [...] » et « Cette TC permet de mettre OFF la carte HKOBS de l'OBMU. [...] ») et il est probable qu'elles aient été copiées-collées, puis adaptées par les rédacteurs. Nous les avons conservées : nous estimons que si un même type de besoin est exprimé fréquemment, il est normal que celui-ci soit représenté plusieurs fois au sein du corpus.

Pour que les observations que nous ferons sur le corpus puissent être généralisées (sans quoi elles n'auraient que très peu d'intérêt), il faut s'assurer que celui-ci soit représentatif. La représentativité d'un corpus dépend avant tout de l'authenticité des données (qui est garantie pour notre corpus : il s'agit de véritables exigences), de l'échantillonnage (TOGNINI-BONELLI, 2001) et de la taille du corpus (qui doit être fonction de l'échantillonnage) :

> A central concern for corpus-based studies is the representativeness of the corpus (see Biber 1993; Biber et al. 1998: 246–50; McEnery et al. 2006: 13–21, 125–30). Two considerations are crucial for corpus design: size and composition. First, corpora need to be large enough to accurately represent the distribution of linguistic features. Second, the texts in a corpus must be deliberately sampled to represent the registers in the target domain of use. (BIBER, 2010 : 162)

Concernant la taille de notre corpus, nous avons inclus toutes les exigences qui ont été mises à notre disposition ; il ne pourrait donc pas être plus grand. Cette taille reste évidemment très modeste par rapport aux standards actuels, puisque des corpus supérieurs au nôtre de plusieurs ordres de magnitude peuvent facilement être constitués. Il faut cependant garder à l'esprit qu'il s'agit d'un corpus spécialisé : cela signifie, d'une part, que les données sont évidemment plus difficiles à collecter (d'autant plus que les exigences sont des textes généralement

courts) et, d'autre part, qu'une taille moindre peut être suffisante ; il ne s'agit pas d'être représentatif de la langue générale. De ce point de vue, le nombre d'exigences est sans doute au moins aussi important que le nombre de mots.

Concernant l'échantillonnage, nous avons déjà dit que ne garder que le contenu des exigences nous semblait être la meilleure option, et expliqué pourquoi les exigences avaient toutes les chances de constituer un genre. Si l'on admet alors que notre corpus est représentatif, il convient de se demander *de quoi* il est représentatif. Pour le savoir, il faut se rappeler la constitution des deux sous-corpus : comme nous l'avons dit, les exigences qui les composent appartiennent à deux projets différents, mais qui ont en commun d'être des projets spatiaux réalisés par le CNES. Il faut ajouter qu'au dire des experts du CNES, les profils des rédacteurs sont tout à fait semblables (formations similaires, etc.) et qu'un rédacteur du CNES peut très bien travailler ou avoir travaillé sur plusieurs projets différents. Il faut donc sans doute supposer l'existence d'une « culture d'entreprise », plutôt qu'une culture différente pour chaque projet. Nous avons répertorié tous les rédacteurs ayant travaillé sur au moins une des trente spécifications dont nous disposons : ils sont au nombre de trente-quatre (sans compter les génériques « équipe Microscope » et « équipe système »), ce qui nous paraît suffisamment varié. Enfin, tous les niveaux de l'« arbre produit » des projets sont représentés (à l'exception de la mission, le niveau le plus élevé) : système, segment, module, équipement ; auxquels s'ajoutent les spécifications d'interface.

Tout ceci nous amène à conclure que notre corpus peut raisonnablement être considéré comme représentatif *des exigences de projets spatiaux rédigées en français au CNES*. Il serait en revanche hasardeux, à ce stade, d'affirmer qu'il est représentatif des exigences en français ; d'une part parce qu'il ne contient pas d'exigences rédigées en dehors du CNES (ou de proches partenaires du CNES) et d'autre part parce qu'il ne contient pas d'exigences rédigées pour spécifier, par exemple, des logiciels informatiques ou des automobiles (et il se peut que des systèmes ou des domaines différents induisent des modes de formulation des exigences différents). Nous ne pouvons donc pas prétendre à un tel niveau de généralité, mais cela n'est pas strictement nécessaire pour notre projet, puisque notre ambition est d'améliorer la rédaction des exigences *au CNES*. Notre corpus est donc, à tout le moins, représentatif des usages que nous entendons décrire.

2.2.4 *Conclusion*

La linguistique de corpus nous offre la possibilité de confronter des hypothèses – et nous suggérons que les recommandations imposées par les langues contrôlées doivent être considérées comme telles tant qu'elles n'ont pas été vérifiées scientifiquement – à la réalité des usages, dont le corpus offre un échantillon représentatif. Les outils et méthodes de la linguistique de corpus et de la phraséologie permettent en effet des études systématiques et quantifiées et facilitent aussi l'analyse fine.

L'apport de ce travail est avant tout méthodologique : nous entendons combler une lacune dans la conception des langues contrôlées, dont l'utilité ne fait pour nous pas de doute, mais qui ne reposent pas toujours sur des fondations solides (ainsi que nous le montrerons dans le prochain chapitre). En combinant les approches dites *corpus-based* et *corpus-driven*, nous souhaitons exploiter au mieux le corpus pour mettre au jour des régularités langagières caractéristiques d'un genre textuel, qui nous permettent de mieux comprendre ce dernier. En outre, une étape de validation des propositions nous semble nécessaire pour en asseoir la légitimité. Cette méthodologie simple se veut facilement reproductible, y compris dans d'autres domaines ou pour d'autres types de textes.

Dans la seconde partie, nous mettrons en pratique nos propositions pour en montrer la faisabilité et l'intérêt, et nous commenterons les résultats obtenus sur les exigences du CNES.

Deuxième partie

ANALYSE ET ÉVALUATION DES PREMIERS RÉSULTATS

3

ANALYSE DU CORPUS : APPROCHE CORPUS-BASED

3.1 INTRODUCTION

Dans ce premier chapitre consacré à l'analyse du corpus d'exigences, nous nous concentrerons tout d'abord sur l'approche dite *corpus-based* (voir section 2.2.2). Autrement dit, dans l'optique de montrer les spécificités du genre des exigences, nous utiliserons le corpus pour vérifier des hypothèses – par définition, formulées *a priori* – portant sur les caractéristiques et la grammaire de ce genre. Plus concrètement, nous chercherons à mettre en évidence la présence (ou, au contraire, l'absence) de certaines formes langagières, mais aussi et surtout à montrer que la fréquence desdites formes dans le genre étudié varie significativement par rapport à la fréquence des mêmes formes dans la langue générale ou dans d'autres genres [1], qu'ils soient supposés proches ou non (CONDAMINES et WARNIER, 2014 ; 2017).

Ne souhaitant pas nous contenter d'une étude uniquement théorique et descriptive (qui nous paraît néanmoins être une première étape indispensable pour démontrer le bien-fondé de notre approche et des résultats qu'elle permet d'obtenir), nous essayerons ensuite de montrer comment ces spécificités liées au genre peuvent servir de bases pour améliorer les langues contrôlées destinées à la rédaction des exigences, voire pour proposer de nouvelles règles (ou, *a minima*, de confirmer l'intérêt des règles existantes). En effet, et comme déjà postulé, une norme langagière, pour être efficace, doit nécessairement tenir compte des usages qu'elle entend ensuite contraindre.

À ces deux objectifs – confirmer l'existence d'un genre représenté par notre corpus et étudier plus finement ses propriétés – correspondent schématiquement deux études, l'une essentiellement quantitative (qui considère le corpus dans sa globalité) et l'autre plus qualitative (qui nécessite de s'intéresser aux occurrences retrouvées dans le corpus).

Naturellement, une des difficultés majeures pour y parvenir (avec les limitations imposées par les moyens techniques dont nous disposons) est de définir, en amont, les hypothèses qui seront testées sur le corpus [2]. De fait, nous devons pour cela être en mesure d'avoir des intuitions sur les phénomènes linguistiques remar-

1. ↑Comme le signale BIBER (2010 : 162) : « One of the major general findings from corpus-based research is that *descriptions of grammatical variation and use are usually not valid for the language as a whole*. Rather, characteristics of the textual environment interact with register differences, so that *strong patterns in one register often represent weak patterns in other registers*. » (c'est nous qui soulignons).

2. ↑Problème qui ne se posera en revanche pas (ou en tout cas avec beaucoup moins d'acuité) avec l'approche *corpus-driven*.

quables qui sont susceptibles de se produire dans celui-ci et ce, sans avoir une connaissance fine et exhaustive de son contenu. Le serpent peut donc sembler se mordre la queue, mais nous disposons heureusement d'une solution toute trouvée pour nous éviter de chercher à l'aveugle : les règles de rédaction formulées dans les langues contrôlées existantes. En effet, nous avons toutes les raisons de penser que si ces règles ont été formulées, elles l'ont été pour remédier à certains problèmes spécifiques à la rédaction technique, ou du moins pour les prévenir ; rappelons que le principe de ces langues contrôlées est justement de se distinguer de la langue générale en excluant de celle-ci toutes les potentielles sources de confusion.

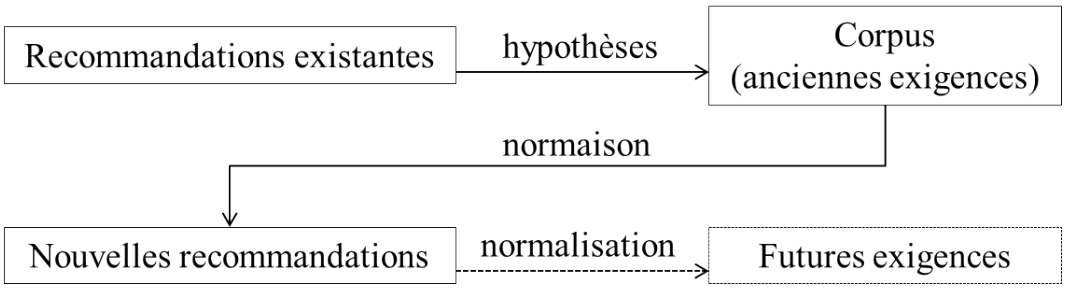
En d'autres termes, même si ces règles n'ont guère été appliquées par les rédacteurs (comme c'est très vraisemblablement le plus souvent le cas dans notre corpus), les phénomènes qu'elles entendent contraindre (parfois indirectement) n'en sont pas moins dignes d'intérêt pour notre étude : soit ils sont effectivement moins fréquents dans notre corpus qu'ailleurs, ce qui va dans le sens des règles édictées par les langues contrôlées (et pourrait indiquer qu'elles ont été dans une certaine mesure respectées), soit ils sont aussi fréquents (voire plus fréquents), ce qui signifie que les règles n'ont pas été suivies. En réalité, la situation est complexe et il faudrait probablement distinguer plusieurs cas de figure. Si les observations sur corpus vont dans le sens des recommandations, il se peut alors que certains rédacteurs aient une connaissance explicite de ces dernières, et qu'ils aient décidé de les appliquer, sans que cela ne leur ait été imposé (ni même demandé) ; mais il est également possible qu'ils en ignorent l'existence et rédigent malgré tout des exigences conformes à ces règles, parce qu'ils peuvent avoir des intuitions sur les formulations susceptibles de contenir de l'ambiguïté et sur les moyens de les éviter, et que ces intuitions concordent justement avec les règles (auquel cas on pourrait avancer que les règles en question sont finalement assez intuitives). Inversement, si les exigences trouvées dans notre corpus ne respectent pas les règles, cela peut simplement tenir à ce que les rédacteurs ne les connaissaient pas, ou bien à ce qu'ils les connaissaient, mais n'ont pas pu ou n'ont pas voulu les appliquer (ce qui pourrait impliquer qu'elles sont inadaptées ou trop contraignantes). Il ne nous sera pas toujours possible de trancher en faveur de l'une ou l'autre de ces hypothèses, mais là n'est de toute façon pas l'essentiel, puisque notre objectif n'est nullement d'évaluer le respect des règles existantes dans des spécifications rédigées par le passé, mais bien, comme nous l'avons dit, d'identifier des phénomènes langagiers sur lesquels nous pourrons effectuer nos analyses, en vue d'améliorer la rédaction de spécifications futures.

Nos hypothèses seront donc inspirées de règles trouvées dans la littérature ; certaines en sont assez directement adaptées, tandis que d'autres seront extrapolées grâce à notre connaissance du fonctionnement de la langue (notamment en contexte spécialisé). Nous suspectons en effet qu'interdire une structure en particulier peut amener les rédacteurs à en favoriser d'autres (stratégies de contournement), sans que le problème visé par l'interdiction ne soit forcément résolu. Par la

même occasion, nous serons amené à critiquer objectivement ces règles.

La FIGURE 3 reprend les principales étapes qui doivent permettre de passer des recommandations existantes à de nouvelles règles plus adaptées aux usages attestés par le corpus.

FIGURE 3 – Approche *corpus-based* pour la définition de nouvelles règles



3.2 CHOIX DES LANGUES CONTRÔLÉES

Afin de choisir les phénomènes langagiers à étudier dans notre corpus d'exigences, nous nous sommes intéressé de près à deux langues contrôlées distinctes, mais visant toutes deux à améliorer la rédaction de textes techniques destinés à être lus et interprétés par des êtres humains (« naturalistes », au sens de CLARK et al., 2009) : la première est le célèbre ASD *Simplified Technical English* (ASD-STE100, issue 4, 15/01/2007), parfois simplement nommé « anglais simplifié », et la seconde est le *Guide for Writing Requirements* (version 1, 30/01/2011)[3] proposé par l'International Council on Systems Engineering (INCOSE) (cf. section 1.5 pour une présentation plus générale des langues contrôlées).

Le *Simplified Technical English* (que nous abrégerons désormais ASD-STE), proposé et mis à jour par l'AeroSpace and Defence Industries Association of Europe[4], est en effet une référence incontournable de la rédaction technique. Elle est utilisée depuis de nombreuses années déjà dans des entreprises diverses et est fréquemment citée dans la littérature, ce qui justifie notre choix, quand bien même cette norme n'est pas originellement destinée directement à la rédaction des spécifications (mais de la documentation des procédures de maintenance[5], en particulier

3. ↑Pour ces deux langues contrôlées, les versions les plus récentes à l'heure où nous écrivons ces lignes datent de 2017 (issue 7 pour ASD-STE100 ; version 2.1 pour le *Guide for Writing Requirements*).

4. ↑Résultant de la fusion en 2004 de plusieurs entités, dont l'AECMA (Association Européenne des Constructeurs de Matériel Aérospatial – en anglais European Association of Aerospace Industries), d'où le nom d'AECMA SE (Simplified English) anciennement porté par ASD-STE (et finalement abandonné en 2005).

5. ↑« ASD Simplified Technical English is an international specification for the preparation of maintenance documentation in a controlled language. » (p. i) Leur site web (`http://www.asd-ste100.org/faq.html`) précise toutefois : « Can STE be applied to documents to all technical documents? [sic] Yes. Although STE was created to improve maintenance documentation, its principles can dramatically improve the reading quality of documentation in any industry. »

des aéronefs). « Simplified English » est d'ailleurs parfois utilisé comme terme générique pour « langue contrôlée » en anglais.

S'agissant du *Guide for Writing Requirements* (dorénavant GWR), outre le fait qu'il soit lui aussi promu par une association influente (l'INCOSE, fondé en 1990, se dit aujourd'hui fort de plus de 15 000 membres dans le monde), il nous a paru idéal, parce qu'il présente le double avantage d'être spécifiquement conçu pour améliorer la rédaction des exigences (ainsi que l'annonce son nom) – sans se limiter à un domaine particulier [6] – et de se présenter comme issu de l'état de l'art sur la question [7]. De ce fait, et étant par ailleurs assez récent, il doit être représentatif des recommandations généralement formulées dans les guides similaires et des tendances dans la rédaction technique en général ; à ce titre, il n'est d'ailleurs pas étonnant que certaines instructions se retrouvent à la fois dans le GWR et dans ASD-STE (bien que le premier ne fasse jamais référence au second). De surcroît, son existence et (au moins) certaines de ses règles sont connues de certains ingénieurs du CNES, qui ont pu y être sensibilisés durant leurs études ou à travers de brèves formations, bien qu'il ne leur soit en aucune manière imposé lors de la rédaction ou de la révision des spécifications.

Comme l'indiquent leurs noms, ces deux normes – tout comme de nombreuses autres – sont rédigées intégralement en anglais et proposent des langues contrôlées basées sur cette langue : leurs règles consistent à imposer ou au contraire à interdire certains éléments (lexicaux ou morphosyntaxiques) de la langue anglaise écrite, parce qu'il est supposé que certains d'entre eux sont moins adaptés (voire potentiellement préjudiciables) dans ce contexte. Bien entendu, cela signifie que ces règles ne peuvent pas être appliquées telles quelles à des textes rédigés en français (pour le lexique, cela n'aurait au demeurant aucun sens). Néanmoins, pour autant qu'elles ne soient pas trop dépendantes de la langue, un grand nombre de ces règles peuvent être adaptées d'une langue à l'autre avec plus ou moins de facilité (surtout lorsque ces deux langues sont typologiquement proches, comme c'est le cas pour l'anglais et le français [8]). Ce sont évidemment ces règles que nous avons retenues ; nous considérons donc que la différence de langue ne constitue pas un écueil majeur, même si nous avons pu être légèrement limité dans le choix des phénomènes à étudier. Une observation de Kittredge nous conforte dans l'idée qu'il serait possible de transposer un certain nombre de règles de l'anglais vers le français (et inversement), parce que des sous-langages parallèles restent proches en dépit de la différence de langue :

> Comparisons of linguistic features in several English and French sublanguages indicate a much stronger structural similarity between parallel (English

6. ↑« The guide is intended to cover the expression of requirements from across disciplines, such as verification, safety, security, reliability, etc., and across all stages of a product lifecycle, such as development, production, operation, maintenance, evolution and disposal. » (p. 12)

7. ↑« The aim is to draw together advice from a variety of existing standards into a single, comprehensive set of rules and objectives. » (p. 10)

8. ↑Pour ses recommandations, le GWR s'inspire d'ailleurs, outre de plusieurs sources anglophones, d'un guide en français rédigé par l'Association Française d'Ingénierie Système (AFIS).

> and French) sublanguages than one sees among disparate sublanguages in the same language. (KITTREDGE, 2005)

Notons qu'étonnamment, et contrairement à ASD-STE (lequel est défini comme un « controlled language »), le GWR n'est jamais présenté explicitement comme une langue contrôlée (mais simplement comme un « guide »). Il en a pourtant toutes les caractéristiques. Si l'on se réfère à la définition de KUHN (2014), on constate en effet (1) qu'il est basé sur une seule langue naturelle, l'anglais ; (2) qu'il est plus restrictif que la langue sur laquelle il est basé concernant le lexique, la syntaxe et la sémantique ; (3) qu'il reste compréhensible pour les locuteurs de celle-ci, dont il conserve les principales propriétés ; (4) qu'il a été consciemment défini. Malgré ce que peut laisser sous-entendre son nom, il ne s'agit pas d'un simple guide de style, parce que les recommandations qu'il contient sont de véritables règles, pas des conseils de bon sens (bien que les auteurs reconnaissent que les règles doivent être adaptées à l'usage [9]). C'est donc bien un langage différent (plus restreint) qui y est défini, car, toujours selon les auteurs :

> Natural language is an imperfect way of expressing requirements. It is hard work to be clear, precise and to avoid ambiguity. However, it remains the only universal means of expression that covers the huge variety of concepts needed. (p. 10)

Le GWR définit une série de quinze objectifs (complétude, compréhensibilité, nécessité, concision, précision, non ambiguïté, etc.) qui sont ensuite justifiés et auxquels sont rattachées des règles plus ou moins directes (par exemple, « Use definite articles » renvoie à l'objectif de précision) ; ces règles – au nombre de cinquante-neuf au total – sont ensuite généralement suivies d'explications complémentaires ou d'exemples. En vue de l'analyse semi-automatique du corpus, nous avons privilégié les règles qui étaient parmi les plus faciles à tester et avons nécessairement dû laisser de côté celles qui nous paraissaient trop complexes ou trop vagues (et ce parfois même pour un être humain), telles que « Express the level of detail appropriate to the layer in which the requirement lives » par exemple.

Il en va de même pour ASD-STE, où les règles sont regroupées par sections (intitulées « mots », « groupes nominaux », « verbes », « phrase », « ponctuation », etc.) et se composent d'une instruction (généralement courte), éventuellement suivie d'explications supplémentaires ou d'exemples ; certaines sont particulièrement précises (pour ne pas dire arbitraires [10]), là où d'autres restent assez difficiles à vérifier [11].

Suite à l'analyse détaillée de ces deux langues contrôlées, nous avons déterminé les phénomènes langagiers sur lesquels porteront nos hypothèses : les pronoms,

9. ↑« [...] the rules have to constantly be adapted to particular situations » (p. 11).
10. ↑Par exemple : « The maximum length of a paragraph is 6 sentences. Do not use one-sentence paragraphs more than once in every 10 paragraphs. »
11. ↑Par exemple : « Present new and complex information slowly » et « Make your instructions as specific as possible ». Ces règles étaient déjà citées à l'identique (pour des raisons similaires) par Huijsen en 1998.

les conjonctions, la longueur des phrases, les temps verbaux (et la modalisation) et la voix passive. Par extrapolation, nous y avons ajouté deux autres phénomènes : le pronom « on » et les nominalisations déverbales.

3.3 CHOIX DES PHÉNOMÈNES LANGAGIERS

3.3.1 *Pronoms*

Ce phénomène est directement inspiré d'une règle du GWR, désignée par le nom *Completeness/Pronouns* [12] :

> Repeat nouns in full instead of using pronouns to refer to nouns in other requirement statements [13].

Elle est expliquée ainsi :

> Pronouns are words such as 'it', 'this', 'that', 'he', 'she', 'they', 'them'. When writing stories, they (sic.) [14] are a useful device for avoiding the repetition of words; but when writing requirements, pronouns should be avoided, and the proper nouns repeated where necessary.

et un exemple est donné :

> *Unacceptable:* The controller shall send the driver his itinary [sic] for the day.
> *Acceptable:* The controller shall send the driver's itinary [sic] for the day to the driver.

La règle générale est donc plutôt contraignante, puisqu'elle indique qu'il n'est pas permis de remplacer un nom ou syntagme nominal par un pronom (mais qu'il faut systématiquement le répéter en entier à la place, ce qui contredit quelque peu l'objectif de concision), et ce, semble-t-il, quelles qu'en soient la longueur et la fréquence d'apparition. Il n'est par ailleurs rien dit des raisons motivant directement cette règle, mais on peut raisonnablement supposer qu'elle vise à éviter l'ambiguïté potentiellement créée par les pronoms, lorsque ceux-ci peuvent avoir plusieurs antécédents différents [15]. Cette règle vise plus particulièrement les références faites à des noms se trouvant dans d'autres exigences (qui violent clairement la règle de complétude), mais peut également se justifier au sein d'une même

12. ↑Il s'agit d'ailleurs de la seule règle reprise directement sous l'objectif *Completeness* (« A requirement statement should be complete in and of itself »).

13. ↑L'expression « requirement statement » est très fréquemment utilisée dans le document, sans jamais être précisément définie dans cette édition. On trouve toutefois, à son propos : « A statement is a single sentence, written in the active voice, and uncluttered by superfluous sub-clauses or auxiliary information, such as justifications, purposes and examples. »

14. ↑« (sic.) » présent dans le texte original.

15. ↑Exemple hors exigences : « Valentine, jeune mannequin, ramène chez *elle* une chienne égarée qu'elle vient de blesser avec sa voiture. » (synopsis sur la page Wikipédia du film *Trois couleurs : Rouge*). Par « elle », il faut comprendre non pas « Valentine », mais « une chienne » (qui est donc plus exactement un conséquent). Un utilisateur modifiera plus tard (29/03/2017) « elle » en « son propriétaire ».

exigence, voire d'une même phrase (comme c'est d'ailleurs le cas dans l'exemple donné, sur lequel nous reviendrons brièvement).

L'explication qui est donnée ensuite commence par citer (de façon non exhaustive, ce qui est regrettable pour une liste fermée) quelques exemples de pronoms, mais sans toutefois en donner de définition. Elle semble relativiser quelque peu la règle, de par l'emploi du modal « should » et de la précision « where necessary », mais, à nouveau, sans la moindre indication sur les cas où cette répétition serait « nécessaire ».

Enfin, il est flagrant que dans l'unique exemple donné pour illustrer la règle, l'élément problématique soit non pas un pronom, mais un déterminant possessif (« his », qui, dans la première version, peut effectivement référer tant au contrôleur qu'au conducteur). Cette observation entraîne deux réflexions. La première est que cette approximation nous conforte dans l'idée que les règles reprises dans le document n'ont de toute évidence pas été validées par des spécialistes de la langue, qui n'auraient pas laissé passer cette erreur (ni sans doute la forme erronée « itinary » pour « itinerary », à deux reprises). La seconde est que la règle vise en fait un phénomène plus général, celui de l'anaphore (et de la cataphore), qui ne se limite pas aux pronoms mais touche aussi, notamment, les adverbes et les déterminants [16], dont le GWR ne dit pourtant rien, ou presque [17]. Ces derniers peuvent pourtant se révéler tout aussi problématiques ; on pourrait donc par exemple vouloir affirmer qu'il ne semble pas possible qu'une exigence débute par un déterminant possessif ou démonstratif [18]. Une règle véritablement efficace devrait donc tenir compte de tous les cas possibles d'anaphore (ou au moins des plus courants) et pas seulement de ceux liés à une reprise pronominale.

ASD-STE semble beaucoup plus souple concernant l'emploi des pronoms (peut-être en raison de la nature différente des textes, les instructions devant nécessairement être lues séquentiellement, au contraire des exigences). Non seulement ils n'y sont pas interdits – « pronoun » étant d'ailleurs une catégorie du discours possible (et « carefully move the deflector unit down until *it* engages » un exemple autorisé) –, mais leur usage est même préconisé dans certains cas (liés à quelques-uns des autres phénomènes que nous présenterons par la suite). C'est notamment permis, sous conditions, avec les pronoms personnels sujets « we » et « you », lors-

16. ↑Exemple hors exigences : « Les adultes voyageant avec des enfants sont priés de placer *leur*(s ?) masque(s ?) en premier » (consignes orales enregistrées d'une compagnie aérienne) : les masques des adultes ou ceux de leurs enfants ? L'ambiguïté ne se pose pas dans la version anglaise de cette même compagnie : « their *own* masks first ».

17. ↑Une règle préconise bien : « Use definite articles », mais pas pour les mêmes raisons et par opposition aux articles indéfinis, jugés moins « précis ». Du reste, on le sait, l'anaphore demeure possible avec un article défini. (Notons au passage que la règle concernant les pronoms cite « this » et « that », qui peuvent également être des déterminants démonstratifs, mais qu'ils ne sont pas envisagés comme tels.)

18. ↑Ce qui est pourtant loin d'être rare dans notre corpus, par exemple : « *Ce passage en survie de la PF* entraîne le passage des CU en survie [...] », renvoyant selon toute vraisemblance par anaphore conceptuelle à l'exigence précédente, « Toute panne PF entraîne le passage en repli du satellite sur des unités B. ».

qu'il s'agit de passer de la voix passive à la voix active et que l'agent n'était pas précisé :

> In descriptive text, when the agent (the person or thing that does the action) is not identified in the sentence, you can use the pronoun subjects "you" or "we" in the active sentence if the agent is the reader ("you") or the manufacturer ("we").
>
> *Example:*
> *Non-STE: On the ground, the valve can be opened with the override handle.* (Passive)
> *STE: On the ground,* you *can open the valve with the override handle.* (Active)
>
> *Example:*
> *Non-STE: The bypass filter is not used in the system because...* (Passive)
> *STE:* We *do not use the bypass filter now because...* (Active)

Et lorsqu'il s'agit de garder une phrase suffisamment courte en la scindant, ce qui peut impliquer une reprise anaphorique dans la deuxième phrase ainsi créée :

> If you have to break up a long sentence, look for connecting words such as "and", "but", "if", or "also". You can usually break a sentence at these words, although you must add a pronoun ("it", "they", "this", "these"), or a demonstrative adjective ("this", "these") and a noun.
>
> *Example:*
> *Non-STE: The side stay assembly has two main members and two folding toggles hinged together and attached with hinges between the main gear strut and the side stay bracket.*
> *STE: The side stay assembly has two main members and two folding toggles.* These *are attached with hinges between the main gear strut and the side stay bracket.*

Les règles d'ASD-STE semblent donc, sur ce point, plus développées et mieux étayées que celles du GWR.

3.3.2 *Conjonctions*

Ce phénomène est, lui aussi, tiré d'une règle lapidaire du GWR (bien que les conjonctions n'y soient pas désignées par ce terme), nommée *Singularity/Propositional* :

> Avoid combinators.

Cette règle est expliquée ainsi :

> Combinators are words that join clauses together, such as 'and', 'or', 'then', 'unless'. Their presence in a requirement usually indicates that multiple requirements should be written.

Et illustrée comme suit :

> *Such as:* 'and', 'or', 'but', '/', 'as well as', 'but also', 'however', 'whether', 'meanwhile', 'whereas', 'on the other hand', 'otherwise'.
>
> *Unacceptable:* The user shall either be trusted or not trusted.
> *Acceptable:* The security officer shall be able to categorize users as (*either* Trusted *or* Not Trusted).
> *Acceptable:* The security officer shall be able to categorize users as exactly one of the following: Trusted, Not Trusted.
>
> *(At the sub-system specification level)*
> *Unacceptable:* "The function 'command the sidelights' knows the demands of the lights, front and back fog lights, to insure the regulatory consistency of lights ignition: the ignition of sidelights is maintained during the ignition of the side lamps, or the lights, or the fog lights or the back fog lights or a combination of these lights."
> *Acceptable:* "The 'control side lamps' function shall illuminate the side lamps while any combination of the following lights is illuminated: side lamps, head lamps, front fog lamps, rear fog lamps.

Cette règle semble aussi contraignante qu'elle est courte, et les deux mots qui la composent peuvent être discutés. « Avoid », au sens strict [19], revient à interdire complètement l'usage de ce qui suit, en l'occurrence les « combinators ». Le GWR donne cependant deux exceptions dans la (sous-)règle qui suit, nommée pour sa part *Singularity/Propositionals/InConditions* :

> Use an agreed typographical device to indicate the use of propositional combinators for expressing a logical condition within a requirement.
>
> Sometimes the use of 'and' and 'or' is unavoidable. This is because there are times when several things need to be present to achieve a correct or desired result. This is often the case when compound conditions have to be defined. [...]
>
> While the rule */Singularity/Propositions* states that 'and' and 'or' are to be avoided, such words may be used to express a logical condition within a single requirement, as described in */Singularity/Propositionals/InConditions*.

Autrement dit, seuls « et » et « ou » peuvent être utilisés, et seulement pour exprimer des conditions complexes dans lesquelles ils seraient indispensables. On pourrait pourtant trouver d'autres cas où l'on aurait peine à s'en passer ; par exemple, lorsque deux couleurs différentes sont attribuées indissociablement à une même entité (« un drapeau blanc et rouge » ; « une photographie en noir et blanc »).

19. ↑ « to prevent something from happening or to not allow yourself to do something » (*Cambridge Dictionary*, en ligne).

Le terme « combinator » pose également question. Aucun des dictionnaires de l'anglais que nous avons consultés n'enregistre l'acception qui lui est donnée ici, à savoir « word that join[s] clauses together ». En tenant compte à la fois de cette définition et des exemples qui sont donnés juste après, on peut logiquement supposer qu'il s'agit en fait de ce que l'on nomme traditionnellement en grammaire française « conjonctions » [20, 21] et « locutions conjonctives » [22] (de coordination : « et », « ou », etc., et de subordination : « à moins que », etc.), et c'est donc à celles-ci que nous nous intéresserons lors de l'analyse du corpus. Il n'en reste pas moins que le choix d'un terme nouveau a de quoi surprendre ; peut-être ce dernier est-il jugé plus simple pour des non-spécialistes ?

Les autres critiques que l'on peut formuler à l'encontre de cette règle sont sensiblement les mêmes que pour celle qui concerne les pronoms (bien que cette fois, on l'a vu, une définition soit proposée) : la liste (non exhaustive) des conjonctions est donnée en vrac, et l'élaboration de la règle semble une nouvelle fois moins directive et plus évasive (« usually indicates », « should ») que la règle elle-même. Cette élaboration permet cependant de justifier l'intérêt de la règle, laquelle vise directement à préserver le principe d'atomicité. Notons enfin que certains exemples de reformulation ne correspondent pas à la définition de « combinator », puisque les éléments qui sont liés ne sont pas des propositions (« clauses ») : « the side lamps, *or* the lights, *or* the fog lights [...] », et qu'un exemple jugé acceptable contient pourtant une conjonction (« while »). Plus généralement, il paraît impensable de se passer complètement des conjonctions : sans « if » (qui répond bien à la définition « word that joins clauses together »), par exemple, comment exprimer une condition de façon naturelle ? [23]

De nouveau, ASD-STE paraît beaucoup plus flexible que le GWR. Aucune règle n'interdit l'usage des conjonctions [24], qui sont l'une des huit catégories du discours autorisées, et l'exemple « Make sure that the flaps are retracted *and* the aileron trim handwheel is in the zero position » (où la conjonction « and » coordonne les deux

20. ↑La définition de *conjunction* est d'ailleurs étonnamment proche de celle donnée par le GWR : « word such as 'and', 'but', 'while', or 'although' that connects words, phrases, and clauses in a sentence » (*Cambridge Dictionary*, en ligne), « A word used to connect clauses or sentences or to coordinate words in the same clause (e.g. and, but, if). » (*Oxford Dictionaries*, en ligne).

21. ↑Notons par ailleurs que le GWR possède une règle nommée *NonAmbiguity/Conjunction*, mais celle-ci est assez éloignée de la thématique abordée ici, puisqu'elle précise seulement : « Use the construct 'only X and Y' instead of 'both X and Y'. » Le terme semble emprunté à la logique plus qu'à la grammaire.

22. ↑Ou plus largement mots conjonctifs, mots de liaison ou encore connecteurs (l'adverbe « alors », par exemple). Ces derniers sont sans doute, dans l'idée (et de par leur nom), les plus proches des « combinators » du GWR et des « connecting words » d'ASD-STE (cf. *infra*), mais ne se limitent pas à une seule catégorie du discours et ont donc pour inconvénient majeur qu'il est plus difficile d'en dresser une liste exhaustive.

23. ↑Une autre règle précise par ailleurs « Use the construct 'if not B then A' instead of 'A unless B' ».

24. ↑Une « recommandation générale » précise même que la conjonction de subordination « that » après un verbe (qui peut être omise en anglais) devrait être systématiquement insérée.

subordonnées complétives [25]) est considéré parfaitement licite.

Seule la règle « Keep to one topic per sentence », que nous avions déjà partiellement citée en évoquant les pronoms, signale qu'il est souvent possible (et préférable, dans la mesure où cela rend les phrases plus courtes et moins denses) de couper une phrase en deux lorsque certains mots sont rencontrés :

> If you have to break up a long sentence, look for connecting words such as "and", "but", "if", or "also" [26]. You can usually break a sentence at these words, although you must add a pronoun ("it", "they", "this", "these"), or a demonstrative adjective ("this", "these") and a noun.
>
> *Example:*
> *Non-STE: The side stay assembly has two main members and two folding toggles hinged together* and *attached with hinges between the main gear strut and the side stay bracket.*
> (The underlined connecting word links two sentence parts that can stand on their own.)
>
> *STE: The side stay assembly has two main members and two folding toggles. These are attached with hinges between the main gear strut and the side stay bracket.*

Il ne s'agit cependant pas d'une injonction stricte comme celle du GWR, mais plutôt d'un conseil au rédacteur, qui reste libre dans le choix de sa formulation (pour autant que la phrase ne dépasse pas une certaine longueur, comme nous le verrons).

Plus encore, quelques règles d'ASD-STE encouragent explicitement l'usage de ce qu'elles nomment « connecting words » [27] (dont la définition est plutôt large), lesquels sont censés rendre le texte à la fois plus clair et plus agréable à lire. C'est en particulier le cas de la règle suivante :

> Use connecting words to join consecutive sentences that contain related thoughts.
>
> Connecting words are words or phrases that connect one thought with another one that follows. To build information slowly and logically, use connecting words to show the relationship between ideas. Such connecting words are: "and", "thus", "also", "but", and "then".

25. ↑Remarquons que « that » n'est pas répété pour la seconde, ce qui semble aller à contresens de la recommandation que nous venons d'évoquer.

26. ↑Bien qu'il n'y soit pas fait référence, ce conseil rejoint clairement une recommandation similaire de FLESCH (1949 : 129) : « Look for the joints where the conjunctions are—*if*, *because*, *as*, and so on—and split your sentences up ».

27. ↑On trouve aussi parfois, dans la terminologie anglaise, les expressions « linking » et « transition(al) words (and phrases) ».

> The old idea that a sentence must not start with "but" or "and" is dead. Just make sure that you do not start too many sentences with "and". If you use "and" too often to start sentences, you will weaken your text.
>
> *Example:*
> *These safety precautions are the minimum necessary for work in a fuel tank.* But *the local regulations can make other safety precautions necessary.*

Si l'on peut, là aussi, regretter un certain manque de systématicité dans la règle (qui ressemble davantage à un conseil de style), elle a au moins le mérite de mettre en évidence le fait que ces mots (parmi lesquels on retrouve des conjonctions) permettent d'expliciter les liens logiques qui unissent les propositions, ce qui s'avère souvent très utile pour rendre le texte plus compréhensible.

En conclusion, le GWR et ASD-STE s'opposent partiellement sur ce que l'on peut assimiler aux conjonctions. Le premier, qui défend à tout prix le principe d'atomicité des exigences [28], semble en interdire presque totalement l'utilisation (quoique les exemples contredisent déjà la règle), tandis que le second en reconnaît (et en souligne) l'utilité pour rendre le texte plus compréhensible – tout en imposant une certaine concision dans les phrases (pour ne pas surcharger le lecteur en informations différentes), ce qui peut impliquer de les scinder là où se trouvent des conjonctions. Il apparaît ici que les approches sont assez différentes : le GWR impose des règles qui, en théorie, peuvent être appliquées mécaniquement, mais qui, en contrepartie, paraissent trop contraignantes, là où ASD-STE propose des règles beaucoup plus flexibles, laissant davantage de latitude au rédacteur, au risque, sans doute, de n'être pas suffisamment suivies.

3.3.3 *Longueur des phrases*

La longueur des phrases – en nombre de syllabes ou, plus fréquemment, de mots – est une variable souvent prise en compte par les formules de lisibilité, y compris les plus célèbres : la formule de Flesch et celle de Dalle-Chall (qui s'en inspire), par exemple, sont en partie basées sur le nombre moyen de mots par phrase dans le texte, un nombre plus élevé caractérisant un texte moins facile. Probablement partiellement sous l'influence de celles-ci, les guides de style préconisent habituellement des phrases courtes (et, idéalement, dont la longueur alterne dans le texte, afin de le rendre moins monotone), donnant pour certains d'entre eux une longueur à ne pas dépasser (généralement de l'ordre de vingt à vingt-cinq mots pour l'anglais [29]).

28. ↑Tel qu'énoncé notamment dans la norme ISO : « The requirement statement includes only one requirement *with no use of conjunctions* » (ISO/IEC/IEEE29148, 2011 : 11 ; c'est nous qui soulignons).

29. ↑C'est le cas, par exemple, du guide de style du site officiel du gouvernement du Royaume-Uni (`https://www.gov.uk/guidance/style-guide/a-to-z-of-gov-uk-style`) : « Do not use long sentences. Check sentences with more than 25 words to see if you can split them to make them clearer. »

Effectivement, il parait intuitif qu'une phrase longue impose au lecteur (en particulier s'il n'est pas natif) une charge cognitive plus grande (puisqu'il lui est nécessaire de retenir plus de mots pour la saisir dans son entièreté), et ce d'autant plus qu'elle est susceptible de contenir davantage d'informations nouvelles qu'une phrase brève. Cette intuition semble confirmée par certaines études, qui ont montré que des phrases (en particulier les phrases les plus longues) obtenaient un meilleur score de lisibilité une fois séparées en phrases plus courtes (COLEMAN, 1962). On ne s'étonne donc pas vraiment de la récurrence de ce type de recommandation, en apparence simple d'application (surtout si la vérification est automatisée), quand bien même il ne s'agit pas là d'une corrélation parfaite et que le nombre maximal de mots autorisé est défini au mieux empiriquement, au pire arbitrairement. À titre d'exemple, dans l'étude menée par O'BRIEN (2006) sur huit langues contrôlées, cette règle était la seule à se retrouver (sous une forme ou sous une autre) dans chacune d'entre elles.

Cela n'a bien sûr pas empêché certaines remises en question (BOTEL et GRANOWSKY, 1972). RICHAUDEAU (1969), notamment, fait tout d'abord remarquer que même des lecteurs lents sont tout à fait capables de lire, comprendre et retenir des phrases contenant plus de mots qu'ils ne sont capables d'en placer en mémoire immédiate [30] (nombre correspondant à leur *empan* de lecture). Il avance que c'est la structure des phrases (plus que leur seule longueur) qui détermine leur lisibilité – concédant que « plus la phrase est courte, plus sa structure a des chances d'être élémentaire et correcte ; mais ce n'est là qu'une caractéristique secondaire » (1969 : 24). Ajoutons à cela qu'insérer des mots supplémentaires dans une phrase peut permettre d'en faciliter la compréhension ; que l'on pense, pour s'en convaincre, aux connecteurs logiques, déjà évoqués. HARBER (1979), s'appuyant sur des études antérieures, faisait ainsi remarquer que la phrase « I went to the store and I bought bread for lunch », malgré un nombre de mots plus élevé, n'est pas plus complexe que son équivalent en deux phrases, « I went to the store. I bought bread for lunch ». Inversement, retirer des mots d'un énoncé pour le raccourcir tend dans bien des cas à l'obscurcir ; un exemple extrême est fourni par le style dit « télégraphique » (que l'on retrouve aussi dans les petites annonces et le langage SMS), mais c'est plus largement le cas de toutes les structures elliptiques, forme de « déviance » fréquente dans les corpus spécialisés (associés à une connaissance particulière) (CONDAMINES, 2003).

On sait les formules de lisibilité, guides de style et langues contrôlées (qui partagent un objectif commun de compréhensibilité) liés de près ou de loin par leurs objectifs et il n'est pas surprenant de retrouver pareille recommandation dans les langues contrôlées. Sur ce point, c'est ASD-STE qui se révèle de loin le plus contraignant.

En effet, le GWR n'impose aucune limite formelle de longueur aux exigences. Il définit bien un objectif de *Concision* (« A requirement statement should be concise »),

30. [↑]Ces mots placés en mémoire immédiate correspondent à un sous-ensemble de la phrase qu'il nomme dès lors simplement « sous-phrase ».

mais celui-ci ne passe par aucune mesure (du nombre de mots ou autre) ; tout au plus le rédacteur est-il enjoint à éviter tout élément non indispensable [31] : « A statement is a single sentence, written in the active voice, and *uncluttered by superfluous sub-clauses or auxiliary information, such as justifications, purposes and examples*. »

ASD-STE, au contraire, recommande explicitement des phrases courtes et limite catégoriquement le nombre de mots par phrase : vingt au maximum dans les « procedures », vingt-cinq pour le « descriptive writing » :

> Keep procedural sentences as short as possible (20 words maximum).
>
> The maximum length of a procedural sentence is 20 words.
>
> Keep sentences in descriptive writing as short as possible (25 words maximum).
>
> In descriptive writing, a sentence length of 25 words is permitted. This is because descriptive text is more complex than procedural text.

Conscients que compter les mots est en réalité loin d'être trivial et que le résultat peut sensiblement varier en fonction des critères retenus pour décider du début et de la fin d'un mot ou d'une phrase, les auteurs d'ASD-STE se sont vus obliger de dédier une section entière (recensant pas moins de onze règles) à la ponctuation et au comptage des mots (par exemple, « un nombre compte pour un mot »). Dans la pratique, on imagine que ces règles ne peuvent être respectées systématiquement que si un logiciel (configuré selon les principes d'ASD-STE [32]) se charge de compter les mots pour le rédacteur, sans quoi la tâche se révélerait des plus ingrates. Elle peut même occasionnellement nécessiter une réflexion linguistique, puisqu'une règle précise :

> When you count words for sentence length, each word in a hyphenated group counts as a separate word unless it is a prefix.
>
> Full words connected with hyphens count as the same number of words (as if there were no hyphens). A prefix is not a word by itself. Therefore, a word with a prefix counts as only one word.
>
> *Example:*
> The air-conditioning system inspection is not part of the pre-flight check.

31. ↑La seule règle visant directement cet objectif concerne les infinitifs superflus ; par exemple, la phrase « The system shall allow the user to *be able to* print invoices » devrait, selon le GWR, être reformulée en « The system shall allow the user to print invoices ». (On peut cependant se demander si, en contexte, la phrase « The user shall be able/allowed to print invoices », encore plus concise, ne lui serait pas équivalente et préférable.)

32. ↑Comme c'est déjà le cas pour plusieurs d'entre eux, même si aucun n'est directement géré par l'ASD (voir section 2.1.3).

> "Air-conditioning" counts as two words, but "pre-flight" counts as only one word.

Pour appliquer la règle correctement, il est donc nécessaire de procéder à une rapide analyse morphologique pour déterminer le statut de l'élément précédant le trait d'union (préfixe ou non).

Cette règle n'est donc pas sans conséquences pratiques et il est de toute façon évident que cette borne supérieure (dont l'origine n'est jamais précisée) ne peut être qu'arbitraire : difficile en effet de justifier pourquoi une phrase de vingt-six mots devrait absolument être réduite, tandis qu'une phrase de vingt-cinq mots serait, elle, parfaitement acceptable dans du texte descriptif. Il est aussi à craindre que cette règle ait un effet pervers, celui de vouloir à tout prix réduire la longueur d'une phrase, en y supprimant des mots dont la présence serait pourtant bénéfique. Une autre règle, donnée dans une section distincte, vise manifestement à dissuader le rédacteur de céder à cette tentation : « Do not omit words to make your sentences shorter ». Pour ce dernier, tout l'art est donc de rester sous la barre fatidique des vingt ou vingt-cinq mots par phrase, tout en respectant toutes les autres recommandations d'ASD-STE et sans tronquer son propos.

Lors de l'analyse de notre corpus d'exigences, nous chercherons à savoir si la proportion de phrases « longues » y est moindre que dans les autres types de textes. Pour cela, il nous faut donc nous aussi fixer une limite au-delà de laquelle une phrase est jugée (trop) longue. Afin d'adapter les recommandations d'ASD-STE, deux éléments seraient à considérer : d'une part, les exigences ne correspondent strictement ni à des instructions (vingt mots maximum), ni à du texte descriptif (vingt-cinq mots maximum) et, d'autre part, il n'est pas certain que la longueur moyenne des phrases en français soit la même qu'en anglais. Kandel et Moles (1958, cités par François, 2011), dans leur tentative d'adapter la loi de Flesch au français, observaient que les mots étaient en moyenne 15 % plus longs dans les textes francophones que dans les textes en anglais, mais ils ajoutaient :

> Quant à la longueur moyenne des phrases, les différences paraissent beaucoup moins systématiques et certainement beaucoup moins reliées à des variables proprement linguistiques qu'au style propre de tel ou tel traducteur (Kandel et Moles, 1958 : 257, cités dans François, 2011 : 141).

Une étude de Gylling (2013) montre des différences considérables dans la longueur moyenne des phrases en danois (21,67 mots), anglais (24,12) et italien (33,82) au sein du corpus parallèle Europarl (transcriptions de discours tenus au Parlement européen). En effectuant la même démarche sur les textes de la paire anglais-français (Koehn, 2005), nous obtenons une moyenne de 23,84 mots par phrase en anglais et 25,51 mots par phrase en français ; autrement dit, dans ce corpus, les phrases en français sont en moyenne 7 % plus longues que les phrases en anglais. Pour cette raison, nous choisirons – là aussi arbitrairement, faute de mieux – la limite « haute » d'ASD-STE (et des langues contrôlées qui s'en inspirent), soit vingt-cinq mots par phrase au maximum.

3.3.4 *Temps verbaux et modalisation*

Nous nous intéresserons ensuite à la façon dont s'exprime le caractère « obligatoire » des exigences. Rappelons en effet que ces exigences ont préalablement fait l'objet d'un accord entre les parties prenantes et qu'à ce titre, leur respect constitue une obligation contractuelle pour le maître d'œuvre externe (et une condition du succès du projet pour la maîtrise d'œuvre interne). Il est donc crucial que le système réalisé soit conforme en tous points à la « description » qui en est donnée à travers les spécifications. Cette description a toutefois ceci de particulier qu'elle énonce – de façon extrêmement contraignante – comment devra se comporter quelque chose (le système ou l'un de ses composants) qui n'existe pas encore au moment de la rédaction (sinon virtuellement, grâce à des simulations, par exemple), mais devra exister plus tard, dans un futur plus ou moins déterminé par les échéances fixées. En cela, elle s'apparente donc surtout à une *injonction* pour le maître d'œuvre, et donc, *in fine*, pour les ingénieurs chargés de lire et de respecter les spécifications.

Pour savoir comment se traduit cette injonction, nous envisagerons deux phénomènes linguistiques (traités séparément par commodité, mais néanmoins partiellement liés, puisqu'ils relèvent tous deux de la modalité au sens large) : le choix des tiroirs verbaux, d'une part, et l'utilisation du verbe semi-auxiliaire modal « devoir » (pour sa valeur déontique), d'autre part.

Avant de nous concentrer sur les recommandations des langues contrôlées, il nous paraît important d'insister sur la nature des exigences, qui présentent des traits assez spécifiques[33]. Comme déjà dit, les exigences (de même d'ailleurs que la plupart des écrits) fonctionnent en différé : le moment de leur rédaction (de leur énonciation, si l'on préfère) est ainsi à distinguer de celui de leur réception et de leur mise en application – mise en application nécessairement postérieure à la rédaction, dont elle dépend fortement. Ce qui est particulier, c'est que la rédaction consiste à décrire en détail un référent qui n'est alors qu'une vue de l'esprit, mais en imposant au (futur) récepteur de mettre en œuvre la réalisation de ce référent, sans qu'il s'agisse pour autant d'une instruction : l'injonction (dont on cherchera à savoir si elle est plus ou moins explicitée) est en quelque sorte indirecte, puisqu'elle porte sur les conditions et les performances de fonctionnement de l'objet décrit – ce à quoi il doit être conforme –, pas sur la manière d'y parvenir (au contraire d'un mode d'emploi ou d'un manuel, par exemple). Les exigences ne sont donc jamais formulées à la première ou à la deuxième personne, seulement à la troisième[34] (la « non-personne » de Benveniste) ; en réalité, s'il y a bien sûr une énonciation, les traces en sont autant que possible effacées, conférant

33. ↑Mais pas totalement uniques, puisqu'on les retrouve en partie dans les contrats et les textes législatifs, qui, sur ces points, en sont assez proches, comme nous le verrons.

34. ↑« [...] dans la communication « en différé », où les acteurs sont le plus souvent absents (les lois, règlements et autres textes prescriptifs), on emploie la troisième personne. Dans le dernier cas, le futur peut – toutes choses égales par ailleurs – être remplacé par le verbe devoir. » (STAGE, 2002 : 53)

aux exigences une apparente objectivité et, surtout, les rendant indépendantes de l'émetteur et du récepteur particuliers (lesquels ne sont du reste pas toujours nommément désignés). Si injonction il y a, c'est donc bien en raison de la nature des exigences et de l'obligation légale qu'elles représentent (une exigence étant l'expression d'un besoin que le maître d'œuvre s'est engagé à respecter) – et non pas de leur formulation. En d'autres termes, une exigence [35] est une obligation *par défaut* et n'a, de fait, pas nécessairement à rappeler à son destinataire qu'elle en est une [36]. Cette injonction indirecte, du maître d'ouvrage au maître d'œuvre, pourrait être paraphrasée ainsi : « *Réalisez un système tel que…* [contenu de l'exigence] ». En raison du contexte et du cadre légal, le maître d'œuvre, une fois l'accord pris, n'est plus en mesure de contester l'obligation créée par l'exigence (il reste toutefois libre du choix de la solution à mettre en place pour la respecter).

Ces considérations prises en compte, deux temps (ou plus précisément tiroirs) verbaux du français semblent s'imposer comme « temps de base » pour la rédaction des exigences [37] : le présent et le futur simple de l'indicatif. Du second, on peut intuitivement penser qu'il paraîtra plus naturel au rédacteur, puisque, par définition, il marque bien la distance temporelle entre le moment de la rédaction et celui de la réalisation de l'objet, l'ultériorité par rapport au moment de l'énonciation (Bres, 2009) étant généralement marquée par un temps du futur. Mais on peut à l'inverse faire remarquer que le présent a pour lui l'avantage de sembler moins marqué (puisque, justement, en tant que temps de base du discours, il efface encore plus les traces de l'énonciation), plus « neutre » et donc plus « générique » : des exigences rédigées au présent devraient être valables à tout moment, là où des exigences rédigées au futur ne devront être valables que « plus tard ». Rappelons en outre qu'après avoir été mises en œuvre, les exigences devront aussi être vérifiées et qu'à ce moment-là, le présent paraîtra sans doute plus naturel que le futur.

On peut donc trouver des arguments en faveur de l'un ou de l'autre de ces temps. Ce qui importe surtout, finalement, c'est que tous les deux permettent bien d'exprimer (voire contiennent en eux) la modalité déontique, inhérente aux exigences. C'est le cas – en contexte – du présent, bien que les nombreuses valeurs qu'il peut recevoir font que la modalité est souvent renforcée par plusieurs marqueurs, qu'ils soient morphologiques, lexicaux ou syntaxiques (Le Querler, 2004). C'est aussi le cas du futur, qui combine valeur temporelle et valeur modale [38] :

35. ↑Dont le nom dans la langue générale est d'ailleurs assez évocateur : « Ce que l'on exige ; ce que l'on veut impérativement. » (*TLFi*).

36. ↑Contrairement à des énoncés hors contexte, qui pourraient être interprétés différemment. Nef (1986 : 127), cité par Stage (2002 : 61), fait ainsi remarquer, à propos de l'énoncé « Vous quitterez votre femme », que « [l']interprétation impérative n'est […] pas indépendante du contexte », car on peut l'annuler, par exemple en replaçant la phrase dans le cadre d'une prédiction astrologique. On pourrait bien sûr donner de nombreux autres exemples : « Tu as fini ? », simple question ou (sur un ton exaspéré) rappel à l'ordre ?

37. ↑Évidemment, nous laissons totalement de côté la concordance des temps.

38. ↑Le lien entre futur et obligation s'observe aussi en diachronie, le futur français (et dans d'autres langues romanes) s'étant formé sur la base d'une expression périphrastique latine (composée de l'infinitif et d'une forme du verbe auxiliaire *habere* (avoir)) qui exprimait à l'origine l'obliga-

> [...] le futur simple dans *Vous préparerez ce dossier avec soin* fait bien sûr référence à une époque d'un futur proche où l'action se réalisera, mais établit surtout un rapport intersubjectif marquant la modalité de l'ordre. Le futur ici marque une relation d'autorité entre le locuteur et l'interlocuteur de façon analogue à un impératif, par exemple (« Préparez ce dossier avec soin »). Temporalité et modalité dans cet énoncé sont très liées et le tiroir verbal du futur exprime donc bien autre chose que la temporalité, ici la modalité intersubjective. (Le Querler, 2004 : 651)

Dans le cas des exigences (comme dans d'autres textes prescriptifs par nature), en raison de l'effacement délibéré de l'énonciateur et du co-énonciateur, nous nous garderons de parler de modalité intersubjective (modalité de type IV dans la typologie de Culioli (Douay, 2003)) et lui préférerons simplement « modalité déontique » (obligation, interdiction, permission) – mais le principe reste le même : il s'agit bien, indirectement, d'un acte de langage exercitif (Austin, 1991) ou directif (Searle, 1972), dont la force dépend avant tout du rapport entre le locuteur et l'interlocuteur ; en l'occurrence, c'est par obligation professionnelle que l'interlocuteur est tenu de s'y soumettre. Les valeurs du futur simple sont donc finalement assez proches de celles de l'impératif :

> Comme une injonction porte sur l'avenir, le futur peut en exprimer, avec différentes forces, les diverses nuances [...]. Le futur simple permet d'expliciter l'époque où doit se réaliser l'ordre, qui est généralement moins strict qu'à l'impératif, à cause de la part d'incertitude inhérente au futur [...]. (Riegel et al., 1994 : 313, cité par Stage, 2002 : 53)

Ce qui distingue impératif et temps de l'indicatif (futur simple ou présent qui, d'après Cornu (1990) tend désormais à le supplanter dans les textes législatifs), c'est la possibilité, avec les seconds, de spécifier ou au contraire d'omettre l'agent – qui est forcément la deuxième personne (le lecteur) dans un texte rédigé à l'impératif :

> [...] dans la communication écrite, le donneur d'ordre est invisible bien qu'omniprésent et les autres « acteurs » sont seulement virtuellement présents. Souvent l'agent n'est pas nommé explicitement. Les textes prescriptifs portent tous la marque de la généralité. [...] Dans le langage juridique, les énoncés au futur déontique sont caractérisés par l'emploi de la troisième personne, par des syntagmes nominaux génériques, par l'emploi du passif et par l'omission de l'agent. (Stage, 2002 : 56)

Une autre différence importante est que l'impératif exprime toujours une injonction ou une interdiction (au sens large : conseil, ordre, menace, etc.), tandis que le présent et le futur simple permettent chacun d'exprimer tant une assertion qu'une injonction. Pour cette raison, il n'est donc pas étonnant que, dans nombre de phrases dans lesquelles le verbe est conjugué à l'un de ces deux temps,

tion : « j'ai à [faire] ». Plus généralement, il faut noter qu'un ordre (ou une interdiction), même à l'impératif, porte toujours sur un procès postérieur au moment de l'énonciation (et donc virtuel).

l'injonction soit renforcée par des modalisateurs (verbes, adverbes, tournures syntaxiques, etc.). Nous nous focaliserons sur le plus évident d'entre eux, à savoir le semi-auxiliaire « devoir ». On peut suspecter sa présence dans les exigences, car il présente le double avantage d'expliciter l'injonction – on peut sans doute y voir la marque de l'obligation par excellence en français, plus encore qu'avec l'impératif – et d'être d'un emploi extrêmement simple et léger [39] (c'est-à-dire forme conjuguée + infinitif). Son inconvénient est une relative polysémie, puisqu'il peut aussi (bien que plus rarement) indiquer une valeur épistémique (plausibilité) ou même alléthique (certitude) (DENDALE, 1994 ; VETTERS, 2012). Remarquons surtout qu'il est ambigu dans les phrases négatives, dans lesquelles il peut signifier tantôt l'absence d'obligation (« il ne doit pas se lever tôt aujourd'hui »), tantôt l'interdiction (« il ne doit en aucun cas s'absenter »). Nous vérifierons donc si les rédacteurs l'emploient dans les exigences ou si, le jugeant superflu, ils préfèrent s'en passer.

Sur ces points, le GWR reste très évasif ; la question des temps et des modaux n'y est jamais clairement évoquée. Un objectif (*Uniformity of Language*) précise bien : « A set of requirement statements should use consistent grammar and vocabulary », mais seule une règle (nommée *UniformityLanguage/Convention*) recommande d'être cohérent dans la façon d'indiquer les exigences obligatoires [40] :

> Use an agreed convention for distinguishing mandatory requirements from non-mandatory.

Elle n'impose aucune convention en particulier, mais fait référence à MoSCoW, un acronyme et moyen mnémotechnique bien connu basé sur certains modaux anglais :

> There are several conventions used for classifying requirements according to their indispensability. One example is the "MoSCoW" convention, where each requirement is tagged with one of:
>
> M = MUST have this.
> S = SHOULD have this if at all possible.
> C = COULD have this if it does not affect anything else.
> W = Won't have this time but WOULD like in the future.

ASD-STE, en revanche, se révèle bien plus précis et contraignant concernant l'usage des temps. D'abord en établissant une liste fermée des temps autorisés dans la documentation, à savoir :

— The infinitive
— The imperative

39. ↑D'autres structures seraient envisageables, par exemple les tournures impersonnelles « Il faut que… » ou « Il est nécessaire que… » + subjonctif. En raison de leur lourdeur, il paraît cependant très peu probable qu'elles soient répétées au début de chaque exigence.

40. ↑Notons que cette indication ne passe pas forcément par des formes langagières ; il peut tout aussi bien s'agir d'une métadonnée.

— The simple present tense
— The simple past tense
— The past participle as an adjective[41]
— The future tense[42].

Mais aussi en imposant que les instructions (qui font partie des procédures) soient rédigées à l'impératif :

> In an instruction, write the verb in the imperative ("commanding") form.

Cette règle tient à ce que, selon eux, l'impératif enjoint plus directement l'opérateur à effectuer la tâche demandée :

> It is better to write the verb in the imperative ("commanding") form because this emphasizes that the operator must do a certain action. Other, less direct forms of instructions leave confusion as to whether something:
>
> — Must be done
> — Is already done
> — Must be done in the future by someone else.

Il est vrai que l'impératif impose la première interprétation (« doit être fait (maintenant, par vous) »), alors que, dans un contexte d'obligation, le présent de l'indicatif permet également les deux autres interprétations (« est déjà fait » ou « devra être fait ») et que le futur peut orienter vers la dernière (« devra être fait (éventuellement par quelqu'un d'autre) ») ; cette règle semble donc avoir tout son sens pour des procédures, de maintenance ou autre. Nous avons cependant expliqué plus haut pourquoi les exigences n'étaient pas des instructions au sens classique du terme et donc pourquoi elles n'étaient pas rédigées à l'impératif. Cette dernière règle ne peut dès lors pas leur être appliquée directement.

Il nous reste à ajouter que dans la tradition anglophone de rédaction des exigences, il est très souvent d'usage d'inclure le modal « shall » (probablement hérité des textes législatifs, où il abonde), comme c'était déjà le cas dans certains exemples tirés du GWR que nous avons cités précédemment, mais surtout comme suggéré par exemple dans la norme internationale consacrée à l'ingénierie des exigences (ISO/IEC/IEEE29148, 2011 : 9) :

> It is important to agree in advance on the specific keywords and terms that signal the presence of a requirement. A common approach is to stipulate the following:
> — Requirements are mandatory binding provisions and use 'shall'. [. . .]

41. ↑En théorie, le participe passé ne peut être utilisé que comme adjectif ; d'autres règles interdisent spécifiquement de l'employer pour former d'autres temps, et les formes passives, comme nous le verrons, sont fortement découragées. Certaines formes irrégulières (« been », « gotten ») sont par ailleurs interdites.
42. ↑Avec l'auxiliaire « will ».

Pour avoir discuté avec plusieurs ingénieurs spécialistes de la rédaction et de la lecture des exigences (au CNES et ailleurs), nous savons que certains d'entre eux vont jusqu'à considérer qu'en l'absence du mot-clé « shall », le texte (en anglais) n'est tout simplement pas une exigence correctement formulée. Ces échanges nous ont aussi appris que ce même « shall » (qui ne possède pas d'équivalent exact en français) est généralement ressenti comme une marque de futur (en même temps que d'obligation). De ce fait, il n'est pas exclu qu'il puisse influencer les rédacteurs francophones et qu'il leur fasse privilégier le futur au présent. Pourtant, même en anglais, le présent simple peut assez efficacement se substituer à « shall » :

> [...] il [= le présent simple] peut exprimer une grande partie des emplois de *shall* déontique car il permet de présenter un fait comme étant avéré au moment de l'énonciation. Or, dans un contexte normatif, cela équivaut à le poser d'une manière impérative, puisque l'autorité de l'énonciateur est absolue. Le critère d'impérativité de la norme est donc rempli.
>
> [...] Le présent indique que c'est ainsi que les choses sont – donc doivent être dans tous les contextes – et seront. [...]
>
> Il serait toutefois possible d'utiliser V[erbe] au présent simple pour formuler des obligations, comme c'est le cas en français avec le présent de l'indicatif, qui est tacitement reconnu comme ayant valeur impérative en droit français. (RICHARD, 2006 : 6-8)

Si « shall » s'est longtemps imposé dans les textes juridiques en anglais, c'est finalement plus par habitude que pour ses valeurs réelles, et nombreux sont aujourd'hui ses détracteurs[43], qui le voient comme inhabituel et ambigu :

> Several of those who call for the elimination of *shall* suggest its replacement in cases of obligation by *must* (e.g. the Australian Office of Parliamentary Counsel or the Plain Language Association) on the grounds that *must* is "more precise and unambiguous" (Cheek 2003); "To state legal obligation, use 'must'. There is no ambiguity in this word" (Lauchman 2002: 47). [...] others encourage a wider use of the present simple [...]. (WILLIAMS, 2005 : 202)

Cette remise en question ne semble toutefois pas avoir eu lieu chez les rédacteurs d'exigences, où sa double valeur (de futur et d'obligation) en fait peut-être le modal le plus « naturel ». Quoi qu'il en soit, le problème reste entier dans d'autres langues, dont le français. Quant à l'AFIS, à travers son guide *Bonnes Pratiques en Ingénierie des Exigences* (AFIS, 2012 : 43), elle indique qu'« une exigence est une phrase simple, écrite au présent » ; mais, en l'absence de justification, cette recommandation ressemble tout autant à un parti pris.

43. †Le *Plain Language Movement*, en particulier, est opposé à « shall » et lui préfère « must » pour marquer l'obligation (https://plainlanguage.gov/guidelines/conversational/shall-and-must/).

3.3.5 *Voix passive*

À l'instar des règles limitant la longueur des phrases, celles qui interdisent la voix passive (ou qui imposent la voix active, ce qui revient au même) sont extrêmement fréquentes : une seule des huit langues contrôlées étudiées par O'Brien (2006) ne formulait pas de recommandation à ce sujet, et leur succès ne s'est pas démenti depuis. Ce n'est sans doute pas non plus un hasard si la voix passive aussi est un phénomène testé de façon récurrente par les études sur la lisibilité (en tout cas par celles qui ne se limitent pas au niveau lexical), et ce sur des textes de natures très diverses (Bostian, 1983 ; Ownby, 2005 ; Heilman et al., 2007 ; Randall et Graf, 2014).

En réalité, il y a principalement deux raisons distinctes qui font de la voix passive la mal aimée des langues contrôlées. La première est justement due aux études psycholinguistiques : les phrases passives sont vues par beaucoup d'auteurs comme marquées[44] et plus complexes à appréhender que leurs contreparties actives, dont elles seraient issues par transformations grammaticales et avec lesquelles elles partagent (selon la grammaire transformationnelle) la même structure profonde. Les phrases actives, « canoniques » (non marquées) et plus faciles à traiter, seraient donc à privilégier.

Lorsque l'on soumet cette hypothèse à l'épreuve des tests, les résultats ne sont pas si tranchés : bien que le temps de lecture soit effectivement généralement plus court (Gough, 1965) et la compréhension meilleure avec une phrase active[45] (Coleman, 1964), il arrive, dans certains cas, que les phrases passives soient plus simples à traiter, parce qu'elles placent le focus sur un actant différent ou qu'elles permettent de respecter un ordre particulier. Hupet et Costermans (1976 : 13-14) citent notamment l'exemple de la phrase passive « Ma sœur est effrayée par le noir », préférée par une large majorité des sujets interrogés à son équivalent à la voix active (« Le noir effraie ma sœur »), probablement car le sujet animé y est placé avant l'objet inanimé. Cette préférence pour le passif est encore plus forte lorsque le syntagme nominal sujet est « nul », comme dans la phrase « On annonça l'heure du départ », à laquelle les sujets interrogés préfèrent largement « L'heure du départ fut annoncée », sans complément d'agent.

Cet exemple permet d'illustrer la seconde raison – sans doute bien plus importante pour la rédaction des exigences –, à savoir que, contrairement à la phrase active (qui impose la présence de l'agent comme sujet du verbe pour être grammaticale : « L'agent doit faire l'action »), la voix passive permet l'omission de l'agent[46] : la phrase « L'action doit être faite » est en effet parfaitement correcte d'un point

44. ↑Les arguments en ce sens sont multiples ; l'un d'entre eux est que les phrases à la voix active (de loin les plus fréquentes) sont toujours possibles, tandis que les phrases à la voix passive (plus longues en français et en anglais si l'agent est précisé) ne le sont qu'avec les verbes d'action.

45. ↑D'après les expérimentations de Charrow et Charrow (1979), fréquemment citées, c'est essentiellement vrai dans les propositions subordonnées.

46. ↑Pour quelques verbes faisant exception, avec lesquels l'agent est obligatoire, voir (Mathieu, 1993).

de vue grammatical, mais paraît incomplète ou inachevée car il n'est pas précisé par qui ou par quoi l'action mentionnée doit être faite. Soit cette information peut être déduite grâce au co(n)texte[47] (GARCÍA-FLORES et al., 2006), soit elle est tout simplement manquante. Le second cas n'est évidemment pas acceptable dans une exigence, où la complétude est un principe majeur. Le premier cas peut éventuellement être toléré, mais on ne s'étonnera pas qu'il soit indésirable, dans la mesure où il n'y a pas de certitude que le lecteur dispose bien de la connaissance nécessaire pour restituer (correctement) l'agent implicite.

En tant que menace à la complétude, on comprend mieux que l'usage de la voix passive soit souvent découragé dans le cadre de la rédaction technique. Cela dit, il est important de préciser que ce problème ne trouve pas son origine dans le passif lui-même, mais uniquement dans le passif non agentif. La phrase « L'action doit être faite par l'agent » est en effet tout aussi complète que son équivalent actif. Si les rédacteurs utilisent des tournures passives, il convient donc premièrement de vérifier si l'agent y est précisé. Le cas échéant, il faut alors se demander pourquoi le passif a été préféré, au détriment de l'actif, pourtant jugé plus « simple » par la plupart des chercheurs. Dans le cas contraire (réduction de l'agentivité), on peut s'interroger sur les raisons qui ont motivé cette omission. Simple oubli ? Agent considéré comme suffisamment évident, de par le contexte ? Ou bien choix plus ou moins délibéré, visant à laisser en suspens la question de la responsabilité (l'acteur exact n'étant peut-être pas encore clairement identifié) ? Remarquons enfin que ne pas préciser l'agent peut aussi signifier que, par défaut, la responsabilité incombe (potentiellement) à tous : « Toute modification des données de vol doit être signalée au poste de commande [par celui, quel qu'il soit, qui est responsable de cette modification] ».

Une fois encore, on peut trouver plusieurs raisons poussant les rédacteurs à opter (ponctuellement ou de façon plus systématique) pour le passif. Les plus connues d'entre elles sont la thématisation (ou topicalisation) (KAIL, 1979) du second argument (qui se voit ainsi antéposé au verbe), lorsque celui-ci a déjà été évoqué (« J'ai croisé un vieil ami. *Le pauvre* a été attaqué par un chien. ») et la focalisation (GUȚĂ, 2006), quand l'agent, déplacé en fin d'énoncé, est par là même mis en relief en raison de l'importance relative que l'on veut lui accorder au sein de celui-ci (« La sécurité du président est assurée *par une équipe de professionnels*. »).

Mais les effets du passif ne s'y limitent pas. Il nous semble à nouveau opportun de rapprocher les exigences (vues notamment comme obligations contractuelles) des textes juridiques, et en particulier des contrats, où les passifs sont apparemment bien plus fréquents que dans la langue générale et d'autres genres de textes (RASMUSSEN, 1999). ROUSKI (2015 : 2) émet l'hypothèse que « les formes passivantes sont exploitées en vue d'imposer de façon à la fois incontournable, incontestable et indirecte, voire anonyme, la volonté de l'autorité ». Elle ajoute que :

47. [↑]De la même façon que certains verbes transitifs peuvent s'utiliser sans complément, en emploi absolu (« Il aime *lire* [des romans, de la poésie, etc.]. ») (NOAILLY, 1996).

> Le désir d'autorité et de pouvoir absolus, de même que l'expression brute et directe de la volonté d'imposer sa puissance, sont inadmissibles dans les sociétés démocratiques contemporaines [...]. Aussi convient-il de faire parler la loi plutôt que le législateur, de substituer la troisième personne à la première, de taire souvent le nom de celui qui parle [...]. Cet état des choses facilite l'occultation très fréquente de l'agent dans les tours passifs. (Rouski, 2015 : 4)

En contexte déontique, les formes passives (d'autant plus si l'agent n'est pas spécifié) auraient donc la propriété de paraître moins directes, moins autoritaires que les formes actives – sans pour autant que l'ordre ne soit en rien affaibli. Cette idée, appliquée ici à la loi, rejoint en grande partie ce que nous avons déjà écrit à propos de l'expression de la modalité dans les exigences et autres textes injonctifs « qui ne se disent pas ».

> Si les modalités déontiques qui sont des modalités orientés [sic] vers le futur apportent une appréciation quant à la réalisation d'une action par rapport aux exigences de la norme, les tours passifs, eux, désignent les résultats obtenus à la suite de la réalisation de la norme. (*ibid.*)

Enfin, le passif, quand on le compare aux formes attributives[48], contribuerait à mettre en avant la résultativité. Le résultat ne peut être obtenu que par l'action de l'agent, qui se révèle indispensable :

> Le passif canonique[49] assure une mise en valeur du résultat. Celui-ci est le terme final du processus exprimé par le verbe. La différence entre les formes attributives et les tours passifs est facilement sentie : « le règlement est nul » *vs* « le règlement est annulé », « le délai est de trois mois » *vs* « le délai est fixé à trois mois ». Avec le passif c'est une nouvelle situation qui est instaurée, un nouvel état qui est obtenu à la suite de décisions prises, de démarches effectuées à différentes étapes. C'est l'effet de l'action (ou des actions) d'un certain agent (ou de plusieurs agents) qui est mis en valeur. (Rouski, 2015 : 5)

En conclusion, toujours selon Rouski (2015 : 8), « Le passif pose le résultat, l'état des choses envisagé, voulu, attendu. Dans cette action l'autorité veut se faire discrète, mais sa volonté est entendue comme inconditionnelle. » Le passif – tant décrié par les chercheurs en lisibilité et les auteurs de langues contrôlées – semble avoir des effets multiples (Gledhill, 2000 : 66-67), que les rédacteurs d'exigences, consciemment ou non, pourraient vouloir mettre à profit.

Le GWR et ASD-STE ne font bien sûr pas exception et incluent tous les deux une ou plusieurs règles concernant la diathèse. Le GWR rappelle tout d'abord

48. ↑Par rapport aux formes actives, c'est la valeur aspectuelle d'accompli qui semble être renforcée (Carlier, 2002).

49. ↑Nommé ainsi par opposition au passif impersonnel (« il a été rappelé que... ») et au passif pronominal (« ... se lit comme suit... »), moins fréquents.

dans l'objectif *Precision* : « A statement is a single sentence, *written in the active voice* [...] ». Il contient aussi une règle laconique (*Precision/ActiveVoice*) qui impose explicitement l'emploi systématique de la voix active :

> Use the active voice with the actor clearly identified.

Aucune autre information (ni même exemple) n'étant donnés, on ne peut savoir avec certitude pourquoi la voix active est imposée de façon aussi catégorique. Au mieux, grâce au complément « with the actor clearly identified », peut-on supposer que c'est avant tout pour empêcher l'omission de l'agent. Il est probable également qu'il soit attendu que l'agent en question (donc le sujet du verbe) soit suffisamment spécifique (tel composant particulier, plutôt que « le système » ou « l'environnement », trop génériques et trop flous). Outre l'absence de justification quelconque, on peut regretter que pas la moindre explication ne soit donnée sur ce qu'est la voix active, à quoi elle s'oppose et sur comment l'identifier ou l'utiliser. Il faut donc espérer que les rédacteurs maîtrisent suffisamment cette notion[50].

Il en va tout autrement dans ASD-STE, qui élabore beaucoup plus longuement sur les voix active et passive. Rappelons en premier lieu que le participe passé est autorisé, mais uniquement utilisé comme adjectif (et non comme forme verbale). Cette règle est valable même après le verbe « être » :

> Use the past participle only as an adjective, either with a noun or after the verbs TO BE, TO BECOME.
>
> [...] When used as an adjective, the past participle shows the state, or condition, that something is in. (Do not confuse this with the passive voice, where it shows an action.)

Plusieurs exemples d'utilisations correctes sont ensuite donnés, dont « The wires are disconnected », où « disconnected » est donc ressenti comme statif.

Ensuite, tout comme le GWR, ASD-STE encourage la voix active. Le début de la règle laisse supposer qu'elle est la seule voix autorisée, mais la suite laisse entrevoir la possibilité d'utiliser occasionnellement la voix passive :

> Use the active voice. Use only the active voice in procedural writing, and as much as possible in descriptive writing.

Dans les cas où la voix active n'est pas utilisée, il ne semble logiquement pas possible d'éviter l'emploi d'un participe passé (non adjectival) pour former le verbe passif. Plus que cette contradiction, c'est surtout le flou entourant l'utilisation de

50. ↑Assez simple à appréhender sur le principe, il est vrai, l'opposition actif *vs* passif peut cependant parfois se révéler très délicate, par exemple quand il s'agit de distinguer « être + participe passé [par...] » (passif) de « être + participe passé employé comme adjectif » (actif). En français, il faut y rajouter une autre possibilité de confusion, « être (auxiliaire d'un temps composé) + participe passé » (actif) (Creissels, 2000). Par ailleurs, l'anglais permet des structures impossibles en français, par exemple « Paul was given a book ».

la voix passive qui pose question. Celle-ci semble autorisée (avec parcimonie) dans les textes descriptifs, mais jamais dans les textes procéduraux (peut-être parce que la bonne identification de l'agent est cruciale dans ceux-ci, mais légèrement moins importante dans ceux-là). Dans la section « Descriptive Writing », on trouve la remarque suivante, qui ne nous éclaire pas vraiment davantage :

> In descriptive writing, you can use the active voice in most cases. Use the passive voice only when absolutely necessary.

Rien n'est dit sur les cas où la voix active ne serait pas possible (parce que l'agent est inconnu ?), ni sur ceux où la voix passive est « absolument nécessaire ». De même dans le dictionnaire, à l'entrée correspondant à la préposition « by », où l'on peut lire, en note :

> Use as a function word to show agency when a passive construction is necessary in descriptive writing.

Il semblerait que les auteurs d'ASD-STE aient voulu laisser ouverte la possibilité d'utiliser la voix passive, mais sans préciser dans quels contextes celle-ci se justifierait ou s'imposerait. En revanche, ils fournissent une longue explication et plusieurs exemples[51] permettant de différencier la voix active de la voix passive. Le premier test qu'ils proposent pour identifier une voix passive est de se demander si la réponse à la question « par qui ou quoi » est donnée dans le texte (ce qui est un indice sûr). Mais ils ne manquent pas d'ajouter qu'une construction passive est possible même en l'absence de groupe commençant par « by » ; dans ce cas, « the second and confirming test for the passive is the way the text is written. [...] we write the passive voice by using some form of the verb *TO BE plus the past participle* of the main verb ». L'inconvénient de ce second test est que des phrases telles que « The wires are disconnected » correspondent aussi à cette structure (il s'agit d'une condition nécessaire, mais non suffisante). Ce constat montre à quel point il peut être difficile d'identifier avec certitude toutes les voix passives – et on peut dès lors se demander s'il est réellement productif de les traquer systématiquement.

Finalement, ils suggèrent plusieurs méthodes pour réécrire une phrase passive à la voix active. La première est évidemment de réorganiser les actants (le complément d'agent devenant sujet) ; la deuxième est plus spécifique : « Change an infinitive verb to an active verb »[52] ; la troisième concerne les procédures, dans lesquelles, comme déjà dit, l'impératif est obligatoire ; la dernière, déjà citée, est la possibilité de remplacer un agent manquant par les pronoms « we » (pour « le fabricant ») ou « you » (pour « le lecteur »).

51. ↑Certains critiquables : « The manufacturer made the procedures the same [...] » devenant « The procedures have been made the same by the manufacturer [...] », passant ainsi du *past simple* actif dans la première phrase au *present perfect* passif dans la seconde.

52. ↑L'exemple donné est le suivant : « These values are used by the computer *to calculate* the altitude » devient « The computer *calculates* the altitude from these values. » On ne peut s'empêcher d'observer qu'il s'agit pourtant d'une transformation plus complexe ; la véritable forme active de la première phrase serait « The computer uses these values to calculate the altitude ».

Ceci nous amène à nous demander si les rédacteurs, sachant que la voix passive est déconsidérée, ne sont pas tentés d'utiliser d'autres stratégies pour formuler leurs phrases à la voix active sans pour autant devoir préciser l'agent. Nous pensons en particulier au pronom « on » et aux nominalisations déverbales.

3.3.6 *Pronom « on » et nominalisations déverbales*

Les deux derniers phénomènes langagiers que nous avons choisi d'envisager ne trouvent pas directement leur origine dans les deux langues contrôlées que nous avons étudiées, mais plutôt dans nos intuitions de linguiste. Plus précisément, nos hypothèses sont extrapolées des règles interdisant l'utilisation de la voix passive, très probablement dans le but d'éviter le passif non agentif. Nous voudrions savoir si cette interdiction n'a pas pour conséquence d'inciter les rédacteurs à trouver d'autres moyens de formuler leurs exigences, formellement à la voix active, mais en restant malgré tout vagues quant à l'agent chargé de la réalisation de l'action.

Ces phénomènes sont le pronom[53] « on » (que nous traiterons donc à part des autres pronoms dans l'analyse qualitative) et les nominalisations déverbales. Le premier étant propre au français – bien que l'on puisse lui trouver des équivalents en anglais, en particulier certains usages du pronom « one » (KITAGAWA et LEHRER, 1990), qu'il a pu influencer –, il est tout à fait logique qu'il ne soit pas spécifiquement pris en compte par les langues contrôlées anglophones (même si l'on pourra sans doute objecter qu'il constitue une infraction à l'objectif de précision du GWR, et en particulier à la règle qui stipule que l'acteur de la voix passive doit être clairement identifié). Ses propriétés sont assez bien connues : il ne peut être utilisé qu'en position de sujet et son référent est normalement toujours humain, mais, bien que correspondant formellement à la troisième personne du singulier, il peut en fait se substituer à tous les autres pronoms personnels sujets (première, deuxième et troisième personnes du singulier et du pluriel), ce qui en fait, de loin, le plus ambigu de tous (BOUQUET, 2007 ; MALRIEU, 2007). RABATEL (2001 : 32) conclut à son propos :

> Sa valeur de base, indéfinie, n'est jamais totalement supprimée : soit que le locuteur veuille faire entendre que l'identification ne peut être plus précise ; soit, plus sûrement, qu'il veuille nous faire entendre qu'il ne souhaite pas l'être.

L'ambiguïté qui peut découler de cette plasticité, quand elle est recherchée par le rédacteur, en fait finalement le candidat idéal pour occuper la fonction obligatoire de sujet de la phrase active lorsque l'on ne veut, ou l'on ne peut l'identifier clairement : « *On* vérifiera que les tables sont reçues en intégralité », où « on », dont la valeur sémantique est nulle, peut être n'importe quel agent humain : « tout le monde » ou juste « quelqu'un », avec l'avantage d'être bien plus discret que ce der-

53. À l'instar de BOUQUET (2007 : 146), qui définit simplement le morphème « on » comme « un clitique verbal apparaissant en position argumentale sujet », nous évacuons complètement de nos analyses les interrogations sur la nature de ce pronom, personnel ou indéfini.

nier – car il est évident qu'une exigence débutant par « *Quelqu'un* vérifiera que… » apparaitrait immédiatement aux yeux du lecteur comme imprécise et donc inacceptable.

Contrairement à ce qui est parfois affirmé, même les genres écrits peuvent présenter une forte proportion de « on » (Poudat, 2006 ; Malrieu, 2007 ; Gjesdal, 2008) et nous pouvons donc nous attendre à en retrouver dans des spécifications. Du reste, nous savons grâce à nos entretiens avec eux qu'il a mauvaise presse parmi les ingénieurs du CNES, ce qui indique que le risque qu'il représente leur est bien connu, mais sans doute aussi que son emploi tend(ait) à être fréquent. Il nous faudra déterminer si c'est toujours le cas (notamment comme option concurrente de la voix passive) ou si, au contraire, il a fini par disparaître des spécifications.

Pour finir, nous évoquerons brièvement les nominalisations déverbales, dont il a été montré qu'elles sont « anormalement » fréquentes dans les corpus techniques, au détriment des formes verbales (Condamines et Picton, 2014). Elles non plus ne sont pas interdites ou déconseillées par le GWR et ASD-STE, si ce n'est de façon indirecte : ASD-STE préconisant d'utiliser une liste fermée (définie à l'avance par chaque compagnie, en plus du dictionnaire commun) de noms et de verbes techniques autorisés, il en résulte que si un verbe y figure mais pas la nominalisation correspondante, cette dernière est de fait écartée (mais l'inverse est également possible !). De plus, une autre règle d'ASD-STE précise :

> If there is an approved verb to describe an action, use the verb (not a noun or other part of speech).
>
> If there are different ways to write something, use the construction in which an approved verb shows the action. This is because the primary function of a verb is to express action.

Cette règle consistant à privilégier les verbes pour exprimer une action lorsque cela est possible nous paraît sensée, mais les exemples donnés[54] montrent que ce ne sont pas les nominalisations déverbales qui sont visées ici. Plus encore, en conséquence des règles mentionnées juste avant, il peut arriver que l'emploi d'un verbe soit écarté au profit du nom correspondant, introduit par un verbe support (ce qui rappelle clairement la philosophie du Basic English) :

> For example, in STE you must write "Do the leak test" instead of "Test for leaks". This is because "test" is only approved as a noun in the Dictionary.

Ce qui nous intéresse principalement ici concernant les nominalisations issues d'un verbe transitif, c'est que lorsqu'elles dénotent un procès, il est possible – tout comme avec le passif – de préciser ou non par quel agent il est réalisé. Les

54. ↑Préférer « show… » à « give an indication », « you hear… » à « … is audible », « push… » à « … is depressed ».

nominalisations déverbales étant significativement moins connues des rédacteurs techniques que la voix passive, il est raisonnable de penser que l'omission de celui-ci ait plus de chance de passer inaperçue, même quand il s'agit d'une information importante[55].

Quelques études psycholinguistiques se sont intéressées aux nominalisations déverbales qui, comme la voix passive, sont vues par les auteurs comme « secondes » par rapport aux formulations avec un verbe à la voix active (VINET, 1978) ; là aussi, les résultats semblent indiquer un temps de traitement accru ou une moins bonne compréhension (COLEMAN, 1964 ; BOSTIAN, 1983). C'est pourquoi on peut parfois retrouver des règles qui en découragent explicitement l'utilisation, comme c'est le cas par exemple dans les recommandations du *Plain Language Movement* (qui les appelle « hidden verbs »[56]) ou dans les *Guidelines for Document Designers* (FELKER et al., 1981), qui en proposent (à la section « Avoid nouns created from verbs ; use action verbs ») une longue et intéressante présentation, puis indiquent comment les repérer et pourquoi les éviter.

Ajoutons, pour en finir avec les inconvénients potentiels des nominalisations déverbales, qu'un certain nombre d'entre elles sont polysémiques (et donc ambiguës) parce qu'elles peuvent désigner tantôt le processus lui-même (« la *construction* [= action de construire] a commencé il y a longtemps »), tantôt son résultat (« la *construction* [= ce qui a été construit] a été inaugurée il y a peu »). Cela se vérifie même en contexte spécialisé, où ces nominalisations peuvent être des termes à part entière (CONDAMINES et PICTON, 2014).

Une des raisons qui pourraient néanmoins inciter les rédacteurs d'exigences à utiliser des nominalisations déverbales, c'est que certaines tournures semblent renforcer la responsabilité de l'agent, par exemple dans une phrase attributive comprenant le groupe adjectival « responsable de… » : « L'opérateur est responsable de la gestion des autorisations » (*vs* « L'opérateur gère les autorisations »).

Pour conclure sur la problématique de l'agentivité, le TABLEAU 2 représente de façon très simplifiée la structure argumentale et la possibilité de faire figurer ou non l'agent avec la voix active, le passif, le pronom « on » (sujet d'une phrase active) et une nominalisation déverbale. [X] indique que l'agent ne peut pas être déterminé sans l'aide du contexte. Comme déjà dit, ce cas de figure peut se présenter avec « on » et certains usages du passif et des nominalisations. Il sera intéressant de voir si la fréquence relative d'utilisation de la voix passive est (inversement) proportionnelle à celle des nominalisations et du pronom « on ».

55. ↑Comparer : « Il s'est assuré de la bonne réception des données. » *vs* « Il s'est assuré que les données ont bien été reçues. ».

56. ↑https://www.plainlanguage.gov/guidelines/words/avoid-hidden-verbs/

TABLEAU 2 – Positions respectives de l'agent et du patient par rapport au procès

Actif	L'opérateur	[Agent]	vérifie	[Procès]	le niveau	[Patient]
Passif	Le niveau	[Patient]	est vérifié	[Procès]	par l'opérateur	[Agent]
	Le niveau	[Patient]	est vérifié	[Procès]		[X]
« On »	On	[X]	vérifie	[Procès]	le niveau	[Patient]
Nomina-	La vérification	[Procès]	du niveau	[Patient]	par l'opérateur	[Agent]
lisation	La vérification	[Procès]	du niveau	[Patient]		[X]

3.3.7 *Conclusion*

En guise de synthèse de l'analyse des langues contrôlées que nous avons étudiées, nous voudrions rappeler les principaux défauts que nous avons pu identifier de façon récurrente au sein de celles-ci. Il ne s'agit plus ici de rentrer dans le détail des règles (étant entendu que ces dernières varient de toute façon d'une langue contrôlée à l'autre, et, pour une même langue contrôlée, d'une version à une autre), mais simplement de mettre en évidence les principaux points (de fond ou de forme) sur lesquels nous pensons qu'elles pourraient être sensiblement améliorées[57] :

— Certaines règles paraissent trop contraignantes, et donc difficilement applicables à la lettre. C'est le cas par exemple des règles qui interdisent de façon stricte l'utilisation d'une catégorie de mots fréquents (pronoms, conjonctions, etc.) ;
— D'autres, au contraire, notamment parce qu'elles sont relativement vagues (« rendez vos instructions aussi spécifiques que possible »), peuvent être facilement contournées et se révéler peu efficaces. Il faut probablement y voir des conseils, plus que des règles à strictement parler ;
— En particulier, certaines règles ne traitent que partiellement le problème qu'elles entendent solutionner, en contraignant l'emploi de certaines formes, mais pas d'autres qui présentent pourtant le même risque (les pronoms pour le phénomène plus général de l'anaphore) ;
— En lien avec le point précédent, il est regrettable qu'il ne soit pas toujours clairement expliqué pourquoi une règle doit être suivie et quel(s) risque(s) elle entend prévenir. Il est fort probable que les rédacteurs préféreraient comprendre l'« esprit » de la règle, plutôt que de l'appliquer mécaniquement[58],

57. ↑Nous ne sommes du reste pas les premiers à souligner quelques-unes des insuffisances des langues contrôlées ; LUX (1998) a déjà proposé nombre de critiques de l'anglais simplifié et de son pendant francophone, le français rationnalisé (ce dernier étant aujourd'hui d'un usage nettement plus confidentiel), dont la plupart restent d'actualité vingt ans plus tard : absence de définitions des objets à contrôler, possible inadéquation de la liste des termes autorisés avec la réalité des textes, manque de complétude des règles, etc.

58. ↑Comprendre le sens de la règle nous paraît en effet essentiel pour bien l'utiliser. Il en va de même pour les lois, où « [l]orsque le texte n'est pas clair, voire absurde, il appartient au juge d'en retrouver le sens exact et de faire prévaloir l'esprit sur la lettre. [...] [On a] condamné un voyageur qui était descendu du train en marche alors que le texte d'incrimination interdisait aux voyageurs de descendre "ailleurs que dans les gares et lorsque le train est complètement arrêté..." (article 78

et qu'ils pourraient plus facilement s'adapter à des situations inédites si c'était le cas ;
— Certaines règles peuvent peut-être être difficiles à appréhender par des non-spécialistes de la langue, soit parce qu'elles font appel à des notions plus complexes (distinguer un participe passé utilisé comme adjectif après « être » d'une voix passive ; identifier un préfixe), soit parce qu'elles sont insuffisamment expliquées (surtout en l'absence de définition claire et/ou d'une liste exhaustive des éléments visés) ;
— D'autres, sans forcément être complexes individuellement, peuvent être indéniablement rébarbatives (multiples règles spécifiant comment compter les mots dans une phrase) ou sembler arbitraires (nombre maximum de phrases par paragraphe) ;
— Quelques règles peuvent paraître superflues ou redondantes [59] ;
— Plus généralement, les incohérences ou erreurs ne sont pas absentes, tant s'en faut, qu'il s'agisse d'un exemple contradictoire (ASD-STE : « You can do this by using connecting words such as [...] "so" », qui n'est pourtant pas un mot approuvé dans le dictionnaire), d'un renvoi à une règle inexistante (GWR : « see rule Ambiguity/Conjunction/Scope ») ou encore d'une terminologie inadaptée.

Ces différentes observations (et particulièrement la première) nous confortent dans l'idée que ces règles ne sont, pour certaines, pas totalement adaptées aux usages réels et que même si l'intention est bonne, elles courent le risque d'être au mieux inutiles, au pire contre-productives. On imagine facilement qu'elles ont pu être définies empiriquement, par expérience, pour tenter de mettre fin à un problème rencontré régulièrement lors de la révision des exigences (ou d'autres textes techniques), qu'elles ont éventuellement pu être remaniées de version en version ou d'une langue contrôlée à une autre, mais qu'il a manqué une étape pourtant nécessaire de confrontation à la réalité de la pratique – ce que nous nous proposons de faire au travers de l'étude sur corpus. Les imprécisions et incohérences que l'on retrouve dans ces règles laissent aussi à penser qu'elles ont été proposées par des spécialistes du domaine (ingénieurs ou autres), dont l'expérience est à ce titre extrêmement précieuse, mais qu'il a manqué l'expertise et les conseils de spécialistes de la langue (linguistes). Il ne fait pourtant aucun doute que les connaissances des

d'un décret du 11 novembre 1917). » (SALVAGE, 2016 : §49). C'est moins vrai dans la *common law*, où la « plain meaning rule » prévaut (« the meaning is determined **only** from the text » (GIBBONS, 2003 : 17)) : « La fonction principale du juge est de décider de la bonne interprétation des textes. Pourtant, dans la 'common law' du moins, afin de ne pas juger au cas par cas, les juges évitent l'interprétation pragmatique. Selon la "règle littérale", l'intention doit toujours s'appréhender à travers les mots utilisés. Cette règle traditionnelle pose naturellement des problèmes. Dans la pratique, en l'absence de critères contextuels, il s'avère souvent difficile de déterminer le sens des expressions référentielles. En effet, si tout énoncé est associé quasiment par définition à un contexte, ce contexte doit toujours jouer un rôle dans la compréhension. » (CHARNOCK, 2008 : 109)

59. ↑Dans ASD-STE, la règle 1.8 « Do not use different Technical Names for the same thing » semble n'être qu'un cas particulier de la règle 1.12 « Once you choose the words to describe something, continue to use these same words (particularly Technical Names) » et on pourrait donc probablement en faire l'économie.

uns et des autres gagneraient à être partagées.

Il faut au demeurant admettre que l'objectif des langues contrôlées dites « naturalistes » – contraindre une langue naturelle afin que les locuteurs, natifs ou non, puissent produire exclusivement des énoncés exempts d'ambiguïté ou de flou – est pour le moins ambitieux, tant la langue est riche et diverse et tant l'ambiguïté semble omniprésente en discours, y compris en contexte spécialisé. Il est impossible que les restrictions soient exhaustives et les concepteurs de langues contrôlées se retrouvent face à un dilemme : celles-ci doivent offrir suffisamment de souplesse et d'expressivité pour pouvoir exprimer des besoins potentiellement infinis, avec des nuances de sens parfois fines, mais elles prennent par la même occasion le risque d'être contournées. Il s'agit donc pour eux de trouver un compromis efficace, c'est-à-dire un ensemble de règles dont le bénéfice à long terme est plus grand que le coût (représenté par le temps nécessaire pour apprendre, appliquer et vérifier ces règles). En pratique, les recommandations de formes langagières qui constituent les langues contrôlées sont bien souvent présentées dans des guides de rédaction qui font abondamment recours aux exemples. Ces exemples prototypiques suppléent en partie les règles, lorsque celles-ci apparaissent incomplètes, insuffisamment claires ou tout simplement absentes ; ils représentent des formes supposées correctes dont les rédacteurs peuvent s'inspirer (pour s'en rapprocher autant que possible) et ont, de fait, un statut assez important, au-delà de leur intérêt strictement didactique.

Les quelques phénomènes que nous venons de présenter et à partir desquels nous avons formulé nos hypothèses de travail ne représentent bien sûr qu'une petite partie (que nous espérons néanmoins variée et représentative) de ceux qui sont abordés dans la vaste littérature que composent les langues contrôlées (aux règles généralement nombreuses, souvent récurrentes sur le fond, mais dont l'expression peut sensiblement varier) et les études de lisibilité. Nous espérons aussi que la présentation que nous avons faite de chacun d'eux a permis de montrer comment il est possible de réinterpréter certaines des préconisations qu'elles contiennent pour proposer ensuite une analyse plus proprement « linguistique » (scientifique et descriptive) du corpus, ou même de s'en inspirer pour étudier des phénomènes qu'elles n'abordent pas directement, mais qui semblent relever d'une même problématique (dans notre cas, l'omission de l'agent pour la voix passive).

Nous avions déjà rapidement évoqué, en introduction, certaines des raisons qui ont orienté notre choix. Nous avons principalement écarté trois types de règles pour notre sélection. Tout d'abord, celles qui portent sur des propriétés non linguistiques des documents ; on en trouve un nombre non négligeable dans le GWR (par exemple, rassembler les exigences mutuellement dépendantes) ainsi que dans les guides qui incluent des recommandations de mise en forme, lesquelles relèvent davantage de la « legibility » en anglais (par exemple, utiliser des graphiques ou des tableaux en complément du texte). Nous n'en nions bien entendu pas le bienfondé, mais elles se trouvent clairement en dehors de notre champ d'analyse.

Ensuite, certaines recommandations, même si elles peuvent être considérées comme relevant de phénomènes langagiers, nous ont paru présenter assez peu d'intérêt pour une analyse approfondie ; se limitant souvent au niveau lexical, c'est le cas avant tout de celles qui imposent d'utiliser des mots ou expressions présents dans une liste déterminée ou, au contraire, en interdisent d'autres (souvent appelés génériquement *weak words* en anglais). Plus que la règle elle-même, c'est sans doute l'utilisation effective de ces mots en contexte (ou les difficultés engendrées par l'absence d'un mot) qu'il faudrait étudier finement, ce que font par exemple Krisch et al. (2016) en proposant une méthode automatisée basée sur des règles pour distinguer les occurrences réellement ambiguës de celles qui, en contexte, ne le sont pas. Tout aussi dignes d'intérêt selon nous sont les méthodes utilisées pour identifier semi-automatiquement les termes qui devraient intégrer les listes de mots déconseillés : Merdy, Kang et Tanguy (2016) recourent pour cela à l'analyse distributionnelle sur un corpus d'exigences, un corpus construit automatiquement grâce au Web et un grand corpus générique.

On peut y ajouter les règles qui, souvent vagues, sont difficiles à vérifier (par exemple, se limiter à une seule idée par phrase ou par paragraphe) ou encore celles qui portent sur des phénomènes par trop spécifiques (utiliser telle structure particulière plutôt que telle autre, dans tel cas précis). Enfin, le dernier critère limitant est que certains phénomènes sont tout simplement hors de portée d'une analyse (semi-)automatique, ce qui est un problème à la fois pour notre étude sur corpus (quantitative et qualitative) et dans l'optique d'une implémentation future des règles dans un correcteur automatique. Nous avons dès lors dû en tenir compte.

Il est par ailleurs intéressant de remarquer que parmi la très grande diversité de phénomènes qui ont pu être envisagés dans les études et recommandations de toutes sortes, nombreux sont ceux qui se rejoignent plus ou moins fortement. S'intéressant à la notion de complexité de la phrase en français (en particulier dans une perspective d'application à la recherche d'information), Tanguy et Tulechki (2009) ont identifié une série de cinquante-deux traits (dont quelques-uns tirés directement d'ASD-STE) pour lesquels ils ont calculé automatiquement la valeur correspondante en corpus. Il en ressort une assez grande redondance des traits, qui sont pour certains fortement corrélés ; c'est notablement le cas de la plupart des formules de lisibilité et de la longueur des phrases (quelle que soit l'unité retenue). Cette première étape leur a permis de réduire considérablement le nombre de traits à envisager, et une analyse en composantes principales sur ceux-ci d'identifier cinq facteurs principaux (c'est-à-dire comptant pour plus de 5 % de la variance totale) – dont le premier, sans réelle surprise, est la longueur des phrases (en nombre de mots).

3.4 CHOIX DES TEXTES ET OUTILS

3.4.1 *Textes de comparaison*

La première phase de notre analyse sera quantitative, et plus exactement comparative : notre premier objectif étant, rappelons-le, de montrer la spécificité du genre des « exigences » (voir section 1.4), nous poserons, sur la base des phénomènes que nous venons de décrire, une série d'hypothèses simples qui reposent toutes sur l'idée que la fréquence relative de tel ou tel trait dans notre corpus est différente (supérieure ou inférieure) à la fréquence observée dans d'autres genres de textes. Nous avons donc tout d'abord dû rassembler des collections de textes à la fois suffisamment homogènes et distincts les uns des autres pour pouvoir effectuer des comparaisons pertinentes.

Nous avons déjà décrit notre corpus d'exigences (cf. section 2.2.3), qui sera au centre des analyses, et les deux sous-corpus qui le composent, nommés Pléiades et Microscope. Nous le comparerons à d'autres exigences, également fournies par le CNES, mais relatives à un autre domaine, celui du logiciel informatique. Bien que nous aurions initialement souhaité intégrer ces dernières directement au corpus (et ainsi le rendre représentatif de plusieurs domaines développés au CNES), nous avons préféré y renoncer pour plusieurs raisons. La principale d'entre elles est qu'il nous a été rapporté qu'à la différence de domaine s'ajoute une différence de « culture » et de pratiques dans la rédaction des exigences : quoique tous soient employés par le CNES, il ne s'agit pas des mêmes rédacteurs et il n'y aurait guère d'interactions entre ces différentes équipes (projets spatiaux et logiciels informatiques), ce qui laisse penser qu'il peut s'agir de (sous-)communautés professionnelles distinctes. En particulier, les exigences informatiques seraient rédigées dans un style plus « littéraire » et en apparence moins « technique », les éloignant des (supposées) exigences « prototypiques » qui nous intéressent au premier chef, ce qui risque de biaiser les résultats. Elles sont en outre nettement moins nombreuses ; celles que nous avons obtenues sont issues de deux spécifications de besoins différentes, contenant 338 exigences pour l'une et 509 pour l'autre (contre plus de 1 000 pour Microscope et plus de 2 500 pour Pléiades), pour un total de 55 670 mots[60]. Leur identification se fait ici par des marques de style (et non par des balises comme dans Pléiades et Microscope), ce qui complique quelque peu leur extraction automatique. Par facilité, nous les traiterons à part dans cette première étude, mais il sera intéressant de voir si des rapprochements restent possibles ; l'idée étant, à terme, d'intégrer des projets plus divers au corpus pour gagner en généralité.

Le deuxième texte de comparaison présente également des similitudes avec le corpus d'exigences, puisqu'il concerne le même domaine (spatial) et a lui aussi été rédigé au CNES. Il s'agit de la troisième et plus récente (2002) édition en date

60. [†]Un premier constat est que ces exigences sont en moyenne beaucoup plus longues : 65,73 mots par exigence, contre 46,69 pour Pléiades et 40,64 pour Microscope. Cela va dans le sens de ce qui nous a été dit à leur propos.

d'un cours de technologie spatiale en trois volumes nommé *Techniques et Technologies des Véhicules Spatiaux* (TTVS)[61]. Rédigé par plus de quatre-vingts experts du CNES, il est destiné à des semi-experts et contient quatorze chapitres, correspondant chacun à un domaine de compétence du CNES (PICTON, 2009). Les extraits que nous avons récupérés sont issus de l'introduction et d'un chapitre consacré à l'optique et l'optoélectronique, dont nous avons préalablement retiré plusieurs éléments (titres et sous-titres, notes, figures, références bibliographiques) pour ne conserver que le corps du texte. Le résultat, composé de 120 086 mots au total (8 229 pour l'introduction et 111 857 pour le chapitre), doit illustrer un genre différent des exigences – technique lui aussi, mais plus didactique et avec d'autres destinataires et une autre finalité :

> Another type of communication takes place between experts and semi-experts, such as students or experts from related fields. In such cases, the experts will probably use the same highly specialized terms that they would use when communicating with another expert from their own field, but they will accompany these terms with explanations where necessary (e.g. in text books). In this way, they help the semi-expert to learn the terms and concepts of the subject field. (BOWKER et PEARSON, 2002 : 28)

Enfin, le troisième et dernier texte de comparaison est censément le plus éloigné des exigences, avec lesquelles il ne présente, *a priori*, pas de similarités ; il s'agit d'articles du quotidien français *Le Monde* datés de décembre 2000. Là aussi, nous avons tout d'abord retiré certains éléments (titres, résultats de tirages de loterie, légendes d'images, etc.). Nous avons de plus réduit sa taille (par simple troncation), de plus de deux millions à 161 403 mots, soit très exactement la taille de notre corpus d'exigences. L'objectif étant simplement de comparer les exigences à des écrits de nature très différente, nous nous garderons de qualifier ce texte de « corpus de référence » du français car, tout comme d'autres auteurs travaillant sur des corpus spécialisés et ayant procédé de la même manière avant nous (KÜBLER et FRÉROT, 2003 ; DROUIN, 2007 ; TODIRAŞCU et al., 2008), nous sommes bien conscient des limites de cette approche. Il ne faut donc pas le voir comme représentatif de la langue générale, mais plutôt comme une illustration de la langue journalistique[62], que l'on suppose beaucoup moins spécialisée que celle utilisée dans notre corpus d'étude.

61. ↑Disponible à la vente, ce cours sert aussi de support lors des sessions de formation organisées annuellement sur dix jours par le CNES : « Le cours TTVS présente les caractéristiques et les contraintes de conception spécifiques à la technologie spatiale et décrit les techniques et les technologies des sous-systèmes à équipements embarqués. Son objectif pédagogique est de vous permettre de visualiser l'ensemble des techniques et des technologies qui concourent à la constitution et au fonctionnement des véhicules spatiaux. Public concerné : Toute personne qui souhaite élargir l'éventail de ses connaissances pour acquérir une vision globale des métiers intrinsèques à la conception des véhicules spatiaux. » (https://www.ttvs.fr).

62. ↑Ce qui n'empêche pas les citations (que nous avons conservées) d'apporter un peu de diversité, qu'elles transcrivent (plus ou moins fidèlement) de l'oral ou qu'elles reprennent d'autres textes.

Le Tableau 3 reprend, pour chaque sous-corpus ou texte de comparaison, le nombre de mots (occurrences) et le nombre de types qu'il contient. Les textes n'étant pas tous de longueur égale, nous serons contraint, dans les analyses quantitatives, de fournir les fréquences relatives en plus des fréquences absolues.

Tableau 3 – Composition du corpus et des textes de comparaison

Corpus	**Nombre d'exigences**	**Nombre d'occurrences**	**Nombre de types**
Pléiades (EX_PLEIADES)	2 528	118 036	5 944
Microscope (EX_MICROSCOPE)	1 067	43 367	3 020
Total (EX_SPATIAL)	3 595	161 403	7 211
Textes de comparaison			
Exigences informatique (EX_INFO)	847	55 670	2 857
Cours TTVS (COURS)	-	120 086	6 323
Le Monde (PRESSE)	-	161 403	14 388

Sans vraiment de surprise, on constate (à taille égale) une bien plus grande diversité lexicale dans les textes journalistiques que dans les exigences de projets spatiaux, avec, pour ceux-là, deux fois plus de types différents utilisés que dans celles-ci (lesquelles se caractérisent donc, de ce point de vue tout du moins, par une certaine redondance).

3.4.2 *Outils*

S'agissant des outils à utiliser pour analyser le corpus (et secondairement les autres textes), la contrainte forte qui a limité notre choix est la clause de confidentialité que nous nous sommes engagé auprès du CNES à respecter. Celle-ci prévoit entre autres que les documents qui nous ont été fournis (dont font partie les spécifications), en raison de leur importance stratégique évidente, ne peuvent être distribués à une tierce partie sans accord préalable. Il en découle que nous nous sommes strictement limité à des applications hors ligne et avons évité les interfaces web, pour lesquelles le risque que les données envoyées soient conservées à notre insu est malheureusement non négligeable. Cette restriction a, de fait, éliminé d'emblée un certain nombre de possibilités (entre autres, le Lexicoscope (Kraif et Diwersy, 2012)).

De plus, nous avons privilégié celles qui étaient suffisamment connues, totalement gratuites et idéalement libres. Nous avons logiquement opté pour celles qui répondaient le plus directement à nos besoins pour cette étude, ce qui explique pourquoi certaines n'ont pas été retenues (mais nous n'excluons évidemment pas d'en faire usage à l'avenir). Parmi celles-ci, on peut citer Le Trameur (Fleury et Zimina, 2014), Analec (Landragin, Poibeau et Victorri, 2012), AntConc (Anthony, 2018), NoSketch Engine (Rychlý, 2007) – la version libre mais limitée du Sketch

Engine (KILGARRIFF et al., 2014) – pour l'exploration et la caractérisation des textes, ou encore YaTeA (AUBIN et HAMON, 2006) ou Acabit (DAILLE, 2003) pour l'extraction de termes dans l'approche *corpus-driven*. Nous regrettons de n'avoir pas pu essayer SubCAT (TEMNIKOVA et al., 2014) ou son successeur annoncé SuperCAT (COHEN, BAUMGARTNER et TEMNIKOVA, 2016) : ces outils d'analyse de corpus particulièrement adaptés aux sous-langages, qui doivent selon leurs auteurs permettre d'évaluer la représentativité d'un corpus, auraient peut-être pu nous être utiles, mais le premier n'est hélas plus disponible depuis des années[63] et le second ne l'a, à notre connaissance, jamais été.

La plupart de nos résultats chiffrés ont été obtenus grâce à l'analyseur syntaxique probabiliste Talismane, développé au sein du laboratoire CLLE-ERSS par URIELI (2013) et diffusé sous licence libre. Il comprend quatre modules linguistiques appelés en cascade (mais qui peuvent au besoin être utilisés indépendamment) : le premier pour la segmentation en phrases, le deuxième pour la segmentation en mots, le troisième pour l'étiquetage morphosyntaxique et le dernier pour l'analyse syntaxique. Bien qu'étant avant tout un *parser* statistique (paramétrable finement), Talismane peut aussi être enrichi de règles linguistiques, améliorant ainsi sa précision sur des phénomènes difficiles (URIELI, 2015) ou spécifiques. Les ressources nécessaires pour le français (et l'anglais) sont fournies par l'auteur, ce qui dispense de l'entraîner pour ces langues.

Nous nous sommes donc basé sur les résultats fournis par Talismane pour détecter les pronoms[64], les conjonctions, le verbe modal « devoir », les voix passives et le pronom « on ». Pour les nominalisations déverbales, nous avons, grâce à un script *ad hoc*, combiné l'étiquetage et la lemmatisation de Talismane (quand elle est disponible) avec la ressource Verbaction (HATHOUT, NAMER et DAL, 2002), mise à disposition sur le site de CLLE-ERSS, qui recense près de 10 000 paires verbe / nom apparentés[65]. Pour déterminer la longueur des phrases, nous les avons tout d'abord segmentées grâce à la suite libre Unitex/GramLab[66] et à ses graphes, puis créé un autre script pour les découper en mots[67]. Nous avons également fait usage

63. ↑En dépit de plusieurs demandes en ce sens auprès de leurs développeurs.

64. ↑Étiquettes CLO (clitique objet), CLS (clitique sujet – dont fait partie « on ») et PRO (pronoms) uniquement ; nous avons donc exclu certaines étiquettes correspondant à des pronoms qui nous semblent généralement non ambigus, tels que les pronoms relatifs (PROREL), interrogatifs (PROWH) ou réfléchis (CLR).

65. ↑Nous avons également essayé d'identifier des nominalisations supplémentaires grâce à une recherche par suffixes (*-age*, *-erie*, *-tion*, etc.) (TANGUY et HATHOUT, 2001), mais la précision était trop faible et l'apport insuffisant, malgré quelques « vrais positifs » intéressants (« calibration » et « paramétrage », par exemple). Nous n'en avons donc finalement pas tenu compte.

66. ↑http://unitexgramlab.org

67. ↑Nous avons considéré qu'un retour à la ligne forcé correspondait à une nouvelle phrase ; ainsi :
« Le CCC doit mettre en place un mécanisme permettant de :
- traiter la TM après chaque passage,
- détecter les éventuels trous de mesures, [...] »
sont traitées comme trois phrases différentes (ce qui respecte les recommandations d'ASD-STE). C'est entre autres pour cette raison que nous avons préféré Unitex à Talismane (dont la sortie ne conserve pas cette information) pour cette analyse.

de la plateforme libre TXM (Heiden, Magué et Pincemin, 2010) pour parcourir et comparer le corpus et les autres textes. Enfin, nous avons eu recours au célèbre étiqueteur TreeTagger (Schmid, 1995), qui peut être intégré directement dans TXM, pour déterminer la fréquence des tiroirs verbaux. Le *tagset* du français par défaut de Talismane, emprunté au *French Treebank* (Crabbé et Candito, 2008) n'inclut en effet pas cette information sur les formes verbales (exception faite des participes passés et présents), mais seulement leur mode.

Il va de soi que, comme pour toute étude de traitement automatique du langage, l'exactitude de l'intégralité des résultats obtenus ne peut être garantie (et ce d'autant plus que le corpus contient des mots techniques ou mal orthographiés, qui peuvent poser problème) ; ces derniers peuvent d'ailleurs varier selon l'outil utilisé. Cela importe peu, l'objectif étant ici de détecter des tendances. La condition pour y parvenir est de rester cohérent dans les choix méthodologiques qui sont faits.

3.5 ANALYSES

Pour chaque phénomène envisagé, l'analyse quantitative sera suivie de la présentation de quelques exemples tirés du corpus d'exigences, puis d'une réflexion sur la façon d'améliorer les règles.

3.5.1 *Pronoms*

3.5.1.1 *Analyse quantitative*

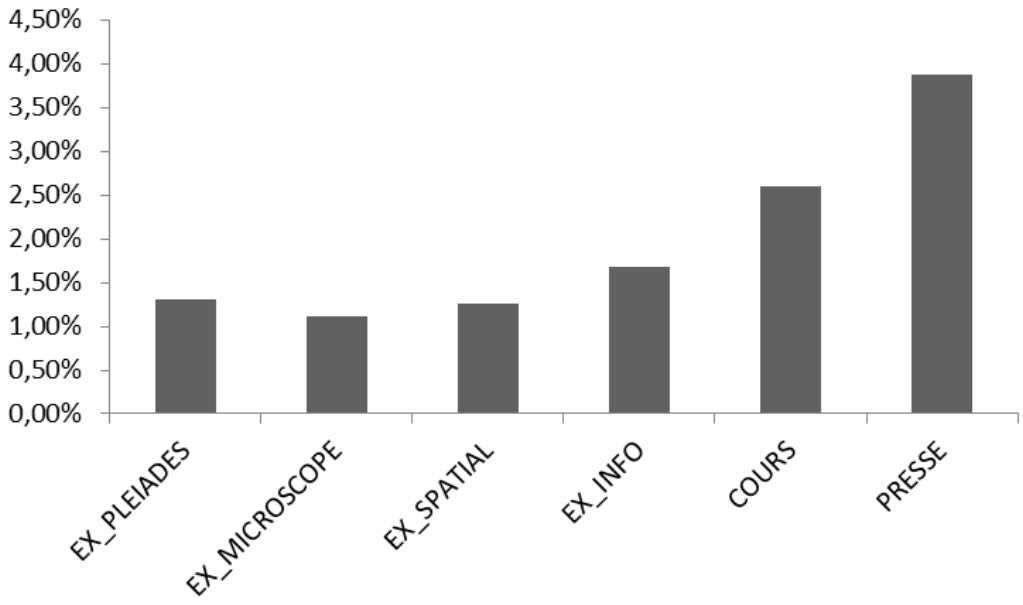
Hypothèse : la fréquence relative des pronoms est plus faible dans le corpus d'exigences que dans les autres textes.

TABLEAU 4 – Fréquence des pronoms (par rapport au nombre total de mots)

	Pronoms
EX_PLEIADES	1543 (1,31 %)
EX_MICROSCOPE	485 (1,12 %)
EX_SPATIAL	2028 (1,26 %)
EX_INFO	936 (1,68 %)
COURS	3114 (2,59 %)
PRESSE	6264 (3,88 %)

Ces premiers résultats correspondent bien à notre hypothèse : ce sont les exigences (Pléiades, Microscope, informatique) qui contiennent le moins de pronoms, suivies du cours et enfin de la presse, où ils sont clairement plus nombreux. Cela n'est pas une surprise, puisqu'ils sont une source possible d'ambiguïté et doivent donc être employés avec précaution dans les textes techniques. Dans les genres où le style se doit d'être « agréable » pour le lecteur (presse, littérature, etc.), en revanche, ils sont un outil indispensable pour éviter les répétitions, très mal perçues dans la plupart des cas (surtout en français). Il faut ajouter que les exigences sont des textes indépendants et assez courts ; en conséquence, le besoin de faire référence plusieurs fois à un même élément au sein du texte pour en assurer la cohésion (au sens de HALLIDAY et HASAN (1976)) se fait sans doute moins souvent sentir – même si, d'un autre côté, on peut aussi supposer que, l'univers du discours étant plus restreint, ce sont souvent les mêmes objets auxquels il est fait référence. Les pronoms ne sont cependant pas totalement absents des exigences, loin de là.

FIGURE 4 – Fréquence des pronoms (par rapport au nombre total de mots)



Tous les chiffres (fréquences absolues et relatives) sont donnés dans le TABLEAU 4 et représentés visuellement sur la FIGURE 4. Le test du χ^2 de Pearson montre une différence fortement significative dans la fréquence d'emploi des pronoms selon les textes [68] (autrement dit, celle-ci n'est très probablement pas due au hasard). Toutefois, bien que ce test soit, encore aujourd'hui, très souvent utilisé dans les études linguistiques sur corpus, plusieurs chercheurs ont montré qu'il aboutissait presque toujours à rejeter l'hypothèse nulle (erreur de Type I), dès lors que les effectifs sont suffisamment nombreux (KILGARRIFF, 2005 ; BESTGEN, 2013). Par conséquent, nous jugeons que cette validation ne présente que très peu d'intérêt (puisqu'elle ira toujours dans notre sens) et faisons le choix de nous en passer pour les analyses suivantes, considérant que la lecture des chiffres est plus éclairante.

La sous-utilisation des pronoms dans les exigences (par rapport aux autres textes) apparait également lorsque le test du χ^2 est effectué sur le tableau de contingence [69] (proportion de pronoms *vs* autres mots selon le type de texte ; valeur du $\chi^2 = 2471{,}1$; $p < 0{,}001$). Les attendus sont donnés par le TABLEAU 5 et les résidus par le TABLEAU 6.

68. ↑Valeurs pour le test du χ^2 d'indépendance par paires (degré de liberté = 1) : EX_PLEIADES-EX_MICROSCOPE : 9,12 ; EX_PLEIADES-EX_INFO : 37,64 ; EX_PLEIADES-COURS : 513,34 ; EX_PLEIADES-PRESSE : 1662,94 ; EX_MICROSCOPE-EX_INFO : 54,63 ; EX_MICROSCOPE-COURS : 321,81 ; EX_MICROSCOPE-PRESSE : 818,51 ; EX_INFO-TTVS : 140,47 ; EX_INFO-PRESSE : 624,53 ; COURS-PRESSE : 354,59. La *p-value* (arrondie à 3 décimales) est systématiquement de 0,000, sauf pour la paire EX_PLEIADES-EX_MICROSCOPE (0,003).

69. ↑Nous remercions Basilio Calderone (Ingénieur de Recherche CNRS) pour ses conseils et son aide concernant ce test.

TABLEAU 5 – Effectifs théoriques (attendus) pour les pronoms

	Pronoms	Non pronoms
EX_PLEIADES	2922,004	115114,00
EX_MICROSCOPE	1073,559	42293,44
EX_INFO	1378,122	54291,88
COURS	2972,752	117113,25
PRESSE	3995,563	157407,44

TABLEAU 6 – Résidus pour les pronoms

	Pronoms	Non pronoms
EX_PLEIADES	-25,510866	4,0644479
EX_MICROSCOPE	-17,962915	2,8618916
EX_INFO	-11,909625	1,8974680
COURS	2,590607	-0,4127413
PRESSE	35,887050	-5,7176046

Pour les exigences, le signe des résidus pour la colonne « pronom » est négatif (répulsion, la valeur des effectifs observés étant inférieure à la valeur des effectifs théoriques), tandis que pour les deux autres textes, il est positif (attraction).

3.5.1.2 *Analyse qualitative*

Il est évident que certains pronoms ne peuvent pas être évités simplement en les remplaçant par un syntagme nominal. C'est le cas, entre autres, des pronoms relatifs (proches dans leur fonctionnement des conjonctions de subordination) :

> Sur réception de cette TC, le LVC met à jour le paramètre *qui*[a] donne la taille maximum d'un paquet TM de type dump [...]
>
> *a.* ↑C'est nous qui soulignons à chaque fois.

Ici, « qui » est indispensable pour introduire la relative et ainsi préciser quel paramètre est mis à jour. C'est aussi le cas des « il » impersonnels, assez nombreux dans le corpus :

> [...] A noter que certaines tables se terminent par un paquet vide (N=0). *Il* ne sera pas utile de vérifier ce paquet " vide " pour contrôler la fin de réception d'une table [...]

Ces pronoms impersonnels ne remplacent pas un nom, ne prêtent pas à équivoque et ne sont donc pas problématiques de ce point de vue[70]. Nous ne nous

70. ↑Pour chaque analyse, nous nous focalisons sur un seul phénomène à la fois. D'aucuns pourraient arguer, par exemple, que cette phrase relève davantage du commentaire que de l'exigence.

attarderons pas sur cette catégorie de pronoms, qui ne relèvent de toute évidence pas de la règle du GWR.

Parmi les pronoms qui remplacent effectivement un syntagme nominal que nous avons relevés dans le corpus, la plupart ne posent pas non plus de problème d'interprétation ; soit parce qu'ils n'ont qu'un seul antécédent possible, soit parce que l'antécédent est rendu évident (même pour des non-spécialistes du domaine) par le contexte ou le cotexte. Ainsi, dans :

> [...] Le paquet ne sera généré que s'*il* est activé par le LVC.

Non seulement « Le paquet » est le seul antécédent possible pour « il », mais le remplacement de ce dernier (« Le paquet ne sera généré que si le paquet... ») produirait probablement une incongruité, voire une possible confusion (s'agit-il bien du même paquet ?).

> Cette TC est limitée à la programmation des paramètres réservés, i.e. *ceux* ayant un champ ALIAS vide en BDS [...]

Même si le remplacement de « ceux » par « les paramètres réservés » serait possible, il ne ferait qu'alourdir une phrase déjà parfaitement claire. Nous ne voyons pas non plus d'inconvénient à ce que l'antécédent se trouve dans une phrase précédente (tant qu'il n'est pas trop distant) :

> Cette TC permet au système de démarrer automatiquement la séquence de déploiement des panneaux solaires. *Elle* permet aussi au sol de redémarrer la séquence [...]

« Cette TC » s'impose logiquement comme l'antécédent d'« elle » (au détriment de « la séquence... », également féminin singulier) en raison de la reprise du même verbe (« permettre »), suivi de « aussi », parce que tous deux ont la même fonction syntaxique (sujet), et parce que « la séquence » est justement l'objet du verbe « redémarrer » dans la deuxième phrase. Il n'y a donc ici qu'une seule interprétation raisonnable. Ce cas est fréquent et nous paraît justifié :

> Le générateur de TCH vérifiera que le nombre d'octets à écrire en mémoire est pair. *Il* vérifiera aussi les bornes de ce paramètre. [...]

« Il » remplaçant sans le moindre doute « Le générateur de TCH », et non « le nombre d'octets... ». Parfois, l'utilisation du pronom démonstratif « celui-ci » (ou « ce dernier ») permet de lever une possible ambiguïté :

> [...] Le générateur de TC ne rejettera pas la création du PARAM_ID diagnostic si *celui-ci* est déjà défini à bord.

Cette question est ici laissée de côté ; rappelons simplement que tous les extraits sont repris à l'identique (exception faite de la mise en forme) du corpus et ont donc été écrits et validés comme exigences par les experts du CNES.

« celui-ci » se réfère nécessairement à l'élément (masculin singulier) le plus proche, donc « PARAM_ID diagnostic » (et non « Le générateur de TC »). Le rédacteur a donc préféré, ici, éviter l'usage du pronom « il » (lequel aurait pu référer au sujet de la phrase précédente). Enfin, c'est souvent le sens de la phrase qui impose une interprétation :

> Les télémesures asynchrones dont la liste est donnée ci dessous doivent respecter les contraintes suivantes :
> - *elles* sont vidées à la demande du sol en HKTM-P et/ou en HKTM-R,
> - il est demandé de *les* décommuter en Temps Réel (cas HKTM-P), [...]

« elles » et « les » réfèrent en toute logique à « les télémesures... », pas « les contraintes... ».

L'antécédent des pronoms est parfois explicité :

> Le DUPC devra attendre au minimum DUREE_MIN_TRANS_ECR_LEC cycles 8Hz après la fin d'écriture d'un fichier avant de pouvoir *en* commander la lecture (du même fichier). [...]

On constate ici que le rédacteur, craignant sans doute que le pronom « en », référant à « un fichier », ne soit pas suffisamment clair, a préféré ajouter une précision : « (du même fichier) ». Celle-ci ne nous semblait pourtant pas indispensable (et un équivalent sans pronom reste possible : « de pouvoir commander la lecture de ce même fichier »).

Il semblerait par ailleurs que certains pronoms soient ajoutés justement afin de rendre la phrase plus claire :

> [...] Dans ce cas, la configuration de repli devient tout B sauf l'équipement en panne qui, *lui*, reste sur A.

Le pronom tonique « lui » (que l'on retrouve parfois dans la structure « quant à lui ») permet d'insister sur l'opposition entre « l'équipement en panne » (qu'il reprend) et le reste de la configuration de repli.

> [...] Le demi-angle de poussée dépend de la durée de la poussée qui *elle-même* dépend du vecteur incrément de vitesse commandé. [...]

Dans cette structure récursive du type « N1 dépend de N2, qui dépend de N3 », « elle-même » après le relatif permet de clarifier la situation de N2 (à la fois objet indirect de N1 et sujet de N3).

Certains pronoms reprennent tout un segment de texte :

> Les valeurs des paramètres MIN et MAX seront saisis en valeurs physiques. Pour *cela* on utilisera le type [...] de la variable [...]

L'expression « Pour cela » est fréquente dans le corpus. Elle reprend ce qui précède en ajoutant des informations plus détaillées.

> [...] Elle permet aussi au sol de redémarrer la séquence au cas où la séquence automatique ne démarrerait pas - *ce qui* ne doit pas arriver nominalement – [...]

« ce qui » reprend avantageusement la proposition précédente (« la séquence automatique ne démarrerait pas ») et commente la situation qu'elle évoque.

Certaines confusions peuvent toutefois naître suite à l'utilisation de pronoms. Dans quelques cas, c'est simplement parce que ceux-ci sont mal choisis et qu'il s'agit vraisemblablement d'une erreur :

> Les format des commandes 1553 à destination du chiffreur/déchiffreur bord (DCU) ne seront pas *toutes* détaillées en BDS ([DR3]). [...]

Ici, l'accord voudrait que « toutes » soit en réalité « tous », « les format [sic] des commandes » étant masculin. (L'erreur est probablement due à « commandes ».)

> Le CECT génère les produits Noa, dès que la télémétrie est décommutée et *le* met immédiatement à disposition du CMSM sur le SEF. [...]

« le » ne pouvant référer qu'à « les produits Noa », il devrait être accordé au pluriel.

> [...] Les noms et versions des sessions citées dans le scénario correspondent à *celles* qui sont présentes dans la bibliothèque des fiches de session opérationnelles (sous le TAG SVN) et qu'elles sont bien qualifiées. [...]

Ici, la grammaire traditionnelle du français contemporain privilégierait un accord au masculin pluriel (« noms » étant masculin), et donc l'emploi du pronom « ceux ». (On peut cependant y voir un accord de proximité avec « versions », effectivement féminin.)

Nous avons relevé quelques cas où un pronom pouvait éventuellement être ambigu, en tout cas en l'absence de connaissances suffisantes sur le domaine. Il appartient aux experts de déterminer si ces pronoms représentent ou non un risque sérieux pour l'interprétation :

> Le PHR Access Unit permet à l'opérateur de dépôt d'extraire les ICRs qu'*il* n'a pas encore vérifiées, avec ses éléments ICR liés.

« il » réfère probablement à « l'opérateur », mais peut-être aussi à « Le PHR Access Unit ».

> La GSD est produite par un outil qui prend en compte la base de données satellite fournie par le constructeur satellite et qui permet de *l'*enrichir par des paramètres et des règles associées définis par le CNES.

« l' » réfère probablement à « la base de données… », mais peut-être aussi à « La GSD ».

> Les mots "envoyer", "recevoir", "extraire" n'impliquent pas une copie des fichiers à l'intérieur d'un Unit : d'autres mécanismes peuvent être utilisés, par exemple notification et disques partagés ; *ils* sont définis lors de la conception de la chaîne de programmation.

« ils » peut référer à « Les mots… » ou à « d'autres mécanismes ».

> […] En routine, le COO utilise ces données pour désigner les stations. En cas de non mise à disposition de *celles-ci* par le CCC, le COO est capable de désigner les stations avec ses moyens propres […]

« celles-ci » réfère probablement à « ces données », mais la proximité de « les stations » sème le doute (*celles-ci* référant normalement à l'élément le plus proche). Nous pensons que « celui-ci », plus difficile à interpréter, devrait être délaissé au profit du couple « il »/« ce dernier » : « Si elles [= ces données] ne sont pas mises à disposition par le CCC… » *vs* « En cas de non mise à disposition de ces dernières [= les stations] par le CCC… ».

> Le PHR Access Unit avertit l'utilisateur s'*il* soumet une demande de production liée à une ICR dont la production n'est pas autorisée (état de traitement différent de " actif-prod " et différent de " actif-tout ") : l'utilisateur peut passer outre, et soumettre la demande au PHR Programming Unit qui effectuera l'envoi, conformément à la requête.

De nouveau, on peut se demander si « il » remplace « le PHR Access Unit » (sujet dans la proposition principale) ou « l'utilisateur » (objet) ; la confusion est d'autant plus grande que « l'utilisateur » est répété en entier par la suite.

L'utilisation d'un pronom (même non ambigu) peut aussi parfois suggérer qu'il serait préférable d'écrire plusieurs exigences distinctes :

> Le suivi des demandes de modifications est effectué conformément à DA7, *celui* des anomalies conformément à DA8.

Cette exigence pourrait être scindée en deux, une pour « le suivi des demandes » et une autre pour « le suivi des anomalies ».

Pour finir, les cas les plus problématiques sont en fait représentés par les exigences contenant un pronom dont l'antécédent est absent (parce qu'il apparaît avant dans la spécification) :

> *Il* a le format suivant : [...]

où « Il » renvoie au sujet de l'exigence précédente, « L'ordre délai court permet de gérer des délais courts [...] ». Au total, nous avons relevé sept exigences présentant ce défaut, dont deux qui se réfèrent à la même exigence :

> Le simulateur sera interfacé au centre de contrôle du satellite suivant un protocole TCP/IP. Il reçoit du centre de contrôle des paquets de télécommande et émet des paquets de télémesure.

suivie de

> *Il* peut être également utilisé en autonome, les TC étant émises en local.

et

> *Il* devra supporter au moins la cadence du temps réel.

Le problème, cette fois, n'est donc pas que l'exigence est ambiguë (parce que le pronom pourrait avoir plusieurs antécédents), mais qu'elle est incomplète (parce qu'il n'en a aucun). De fait, envisagée isolément, l'exigence n'est plus interprétable, ce qui est un problème parce que la phase de vérification et de validation doit se faire exigence par exigence et surtout parce que le texte qui précède pourrait être déplacé, modifié ou supprimé par la suite[71]. L'indépendance des exigences est un principe de base en ingénierie des exigences, qui se voit ici violé. Un cas similaire est illustré par l'extrait de spécification suivant :

> Une surveillance de fonction doit posséder les caractéristiques suivantes :
>
> N£ PHR-ST-0-12-CNES_1400 £N
> T£ - active ou inactive : *elle* ne s'exécute que si elle est active. L'état actif/inactif n'est positionné que par le logiciel en fonction des modes auxquels la surveillance appartient, £T
>
> N£ PHR-ST-0-12-CNES_1500 £N
> T£ - autorisée ou inhibée : *elle* ne s'exécute que si elle est active et autorisée. Cet état est commandable par le sol. £T

où les deux exigences sont totalement ininterprétables en l'absence du segment (hors exigence) qui les précède. Il est probable que le rédacteur ait voulu respecter à la lettre le principe d'atomicité des exigences (une fonctionnalité = une exigence), mais il l'a fait au détriment d'autres principes tout aussi importants.

71. ↑Problème qui ne tient d'ailleurs pas qu'aux pronoms. Dans une des exigences du corpus, on trouve ainsi une référence à une « règle énoncée précédemment », sans plus de précision.

3.5.1.3 *Synthèse*

Il y a une vraie pertinence à proposer une recommandation à propos de l'emploi des pronoms dans les exigences, car ceux-ci sont une réelle source d'ambiguïté. Lux (1998) proposait d'ailleurs déjà trois consignes concernant l'anaphore dans le cadre de son *Français Rationalisé Étendu Modulaire*. Dans la pratique, on constate cependant que le nombre de pronoms pouvant poser des problèmes d'interprétation est assez faible (ce qui indique sans doute que le risque est perçu intuitivement par les rédacteurs ou en tout cas qu'en contexte technique, il est moins fréquent qu'on le pense) et il nous semblerait donc tout à fait contreproductif de les remplacer systématiquement par un syntagme nominal, ainsi que le recommande sans nuance le GWR.

Remarquons par ailleurs que, d'une part, ce problème ne touche pas tous les types de pronoms et que, d'autre part, il touche aussi d'autres parties du discours (déterminants, par exemple). Par conséquent, il serait peut-être plus simple d'utiliser une expression plus générique que « pronom », par exemple « mot qui remplace ou qui renvoie à d'autres » (en partant du principe que le concept d'anaphore a peu de chance d'être connu des rédacteurs d'exigences). Il conviendrait également d'établir une liste de ces mots, dans laquelle on retrouverait en particulier « il(s) », « elle(s) », « le », « la », « l' », « les », « lui », « leur », « y », « en » pour les pronoms et « son », « sa », « ses », « leur(s) » pour les déterminants. La plupart des problèmes rencontrés dans le corpus concernent les pronoms personnels de troisième personne. Du reste, comme on l'a constaté, les cas les plus problématiques sont ceux où l'antécédent ne se trouve pas dans l'exigence elle-même, mais ailleurs.

Si l'on se limite aux pronoms, la règle générale pourrait être formulée ainsi : « N'utilisez un pronom que s'il a un et un seul antécédent possible au sein de l'exigence ». En termes plus simples, elle pourrait devenir : « Si vous utilisez un mot qui remplace un autre (groupe de) mot(s), assurez-vous (1) que ce (groupe de) mot(s) soit bien présent dans la même exigence et soit facilement identifiable et (2) qu'il ne puisse pas être confondu avec un ou plusieurs autres (groupes de) mot(s). Si l'une de ces conditions n'est pas remplie, ou si vous n'êtes pas sûr, réécrivez le (groupe de) mot(s) visé en entier. » Nous pensons que cette règle est suffisamment souple pour éviter les cas problématiques sans être trop contraignante ou amener à des répétitions incongrues.

Si l'on souhaite que la recommandation soit plus générale et ne se limite pas aux pronoms, on pourrait économiquement rajouter : « un mot qui remplace *ou renvoie à* un autre (groupe de) mot(s) ». Si l'on souhaite éviter la cataphore, plus difficile à interpréter, on pourrait remplacer : « soit bien présent dans la même exigence » par « *apparaisse plus tôt* dans la même exigence ».

Pour la plupart des pronoms (« il(s) », « elle(s) », « le », « la », etc.), la liste de leurs antécédents possibles est grossièrement constituée des syntagmes *de même nombre*

et de même genre présents dans le cotexte. C'est une propriété à exploiter par un outil de vérification semi-automatique et cela peut également être un conseil à formuler à destination des rédacteurs pour les détecter. Néanmoins, selon nous, le rédacteur (humain) reste seul capable de décider si un syntagme est – ou non – un antécédent possible : comme nous l'avons vu, beaucoup sont trop invraisemblables pour représenter un réel risque d'ambiguïté. Il faudra toutefois le mettre en garde de ne pas trop se reposer sur ses connaissances du contexte, qui ne seront peut-être pas toutes partagées par le destinataire.

Quelques exemples :

— « Si le paquet était déjà défini, *il* est remplacé. » : acceptable, car « il » ne peut remplacer que « le paquet ».
— « La donnée est archivée dans la base. *Elle* est conservée pendant une semaine. » : acceptable, car le contexte permet de savoir que « elle » remplace « la donnée » et non « la base ». (Si le doute est permis, remplacer par « Cette donnée » ou « La base ».)
— « Le logiciel de bord génère un rapport sur la base des informations contenues dans le fichier envoyé par l'opérateur. *Il* respecte le format décrit dans le document de référence. » : non acceptable, car « il » peut remplacer soit « un rapport », soit « le fichier envoyé par l'opérateur ». Reprendre le groupe de mots visé : « Ce rapport » ou « Ce fichier ».
— (Éventuellement :) « Lorsqu'*il* est complet, le fichier est renommé. » : non acceptable, car « il » remplace un groupe de mots situé après (et non avant). Reformuler en « Lorsque le fichier est complet, il est renommé. ».

Nous recommandons en outre de mettre explicitement en garde contre « celui-ci » (parfois mal utilisé par les rédacteurs du CNES). Si le pronom doit renvoyer au dernier élément mentionné, on préférera « ce dernier ».

L'implémentation d'une règle de ce type (qui se limiterait à l'anaphore pronominale – l'anaphore nominale étant nettement plus complexe à détecter – et éventuellement aux déterminants) dans un outil de vérification est possible. Il suffit, sur le principe, de vérifier, lorsqu'un pronom est utilisé, qu'il y a un (et un seul) syntagme nominal du même genre et du même nombre (quand c'est pertinent) apparaissant avant lui dans l'exigence[72]. S'il n'y en a aucun, un avertissement est émis. S'il y en a plusieurs, le logiciel le signale pour que le rédacteur vérifie rapidement qu'il n'y a pas d'ambiguïté. Le principal risque consiste en un nombre élevé de faux positifs (par exemple, signaler « L'opérateur » comme antécédent possible de « *Il* est allumé »). Les avertissements doivent donc être aussi peu invasifs que possible.

72. ↑S'agissant des pronoms pluriels, il faut tenir compte des cas où ils remplacent plusieurs noms en même temps. « Ils », par exemple, peut remplacer plusieurs éléments à la fois (qui peuvent être au singulier), dont au moins un est masculin. Notons le choix fait par les concepteurs d'ACE : « [...] the ACE system replaces any anaphoric reference by *the most recent accessible noun phrase that agrees in gender and number* » (manuel d'ACE ; c'est nous qui soulignons).

3.5.2 *Conjonctions*

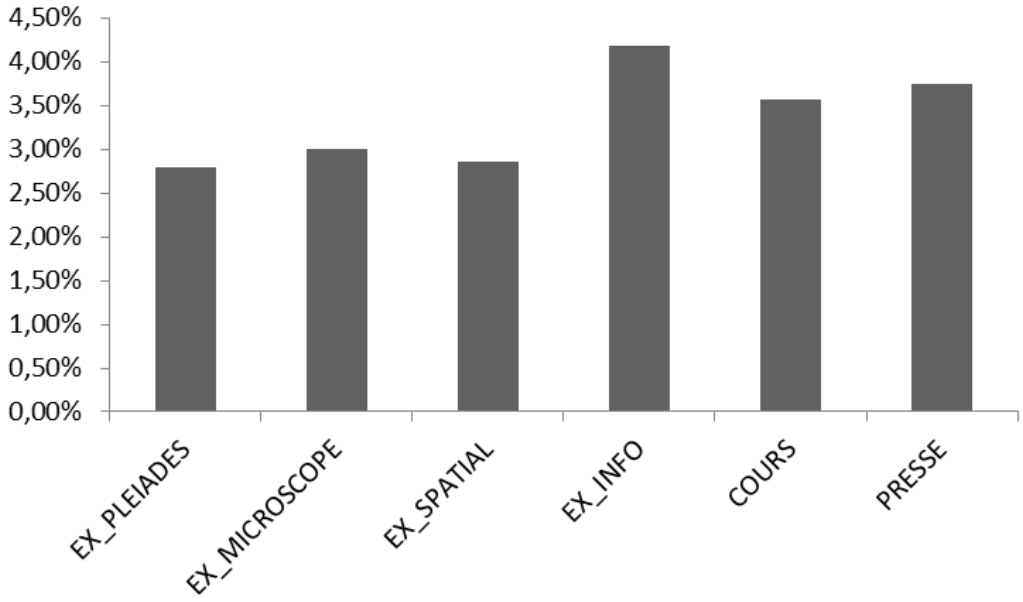
3.5.2.1 *Analyse quantitative*

Hypothèse : la fréquence relative des conjonctions est plus faible dans le corpus d'exigences que dans les autres textes.

TABLEAU 7 – Fréquence des conjonctions (par rapport au nombre total de mots)

	Conjonctions coordination	Conjonctions subordination	Total conjonctions
EX_PLEIADES	2482 (2,10 %)	816 (0,69 %)	3298 (2,79 %)
EX_MICROSCOPE	1092 (2,52 %)	211 (0,49 %)	1303 (3,00 %)
EX_SPATIAL	3574 (2,21 %)	1027 (0,64 %)	4601 (2,85 %)
EX_INFO	1781 (3,20 %)	548 (0,98 %)	2329 (4,18 %)
COURS	3228 (2,69 %)	1047 (0,87 %)	4275 (3,56 %)
PRESSE	3944 (2,44 %)	2106 (1,30 %)	6050 (3,75 %)

FIGURE 5 – Fréquence des conjonctions (par rapport au nombre total de mots)



À nouveau, Pléiades et Microscope confirment notre hypothèse puisque le corpus contient en moyenne moins de conjonctions (de coordination et de subordi-

nation) que les autres textes, comme le montrent le TABLEAU 7[73] et la FIGURE 5. Les exigences informatiques, par contre, constituent une exception sur ce point, puisque ce sont elles qui contiennent le plus de conjonctions (en particulier de coordination). On constate en effet que des éléments (verbes, adjectifs, etc.) y sont souvent coordonnés lorsqu'ils sont sémantiquement proches : « comme un outil convivial, intuitif *et* facile d'utilisation » (« intuitif » et « facile d'utilisation » étant selon nous des quasi-synonymes dans ce contexte), « les messages d'erreur émis par l'outil sont contextualisés *et* explicités ». Il s'agit donc d'une différence majeure entre les deux types d'exigences. Enfin, il n'est pas surprenant que les extraits du *Monde*, au style plus élaboré, soient ceux qui contiennent le plus de conjonctions de subordination et de phrases complexes.

La plus faible proportion de conjonctions dans les exigences spatiales peut s'expliquer par la volonté de respecter le principe d'atomicité (une exigence ne doit exprimer qu'un seul besoin), ce qui n'empêche pas leur présence. On peut s'attendre à ce qu'elle aille de pair avec des phrases plus courtes.

3.5.2.2 *Analyse qualitative*

Nous constatons d'emblée que de nombreuses conjonctions (qu'elles lient ou non des propositions entre elles) ne peuvent tout simplement pas être évitées. C'est le cas des conjonctions de subordination :

> [...] *Si* plusieurs paquets TM sont nécessaires, le LVC génèrera un paquet par cycle 8Hz.

> [...] Le générateur de TC ne rejettera pas la création du PARAM_ID diagnostic *si* celui-ci est déjà défini à bord.

Dans ces deux exigences, la conjonction « si » est indispensable pour pouvoir préciser à quelle(s) condition(s) l'action doit être effectuée. Elle est parfois concurrencée dans ce rôle (improprement, selon nous) par la conjonction « lorsque » :

> [...] Cette TC est refusée *lorsque* :
> * Le mode télécommandé est déjà actif,
> * Le logiciel SCAO n'est pas en mode Normal.

où « lorsque » pourrait tout aussi bien être remplacé par « si ». Nous pensons qu'il serait préférable de s'en tenir uniquement à « si » pour les conditions, et de réserver « lorsque » pour la survenue d'un événement (équivalent de « quand ») :

73. ↑À titre de comparaison, les fréquences absolues obtenues avec l'étiqueteur TreeTagger (dont le *tagset* français par défaut ne distingue pas entre conjonctions de coordination et conjonctions de subordination) sont les suivantes : EX_PLEIADES : 3304 (+ 6 par rapport à Talismane) ; EX_MICROSCOPE : 1326 (+ 23) ; EX_SPATIAL : 4630 (+ 29) ; EX_INFO : 2376 (+ 47) ; COURS : 4379 (+ 104) ; PRESSE : 6157 (+ 107).

> [...] Ces paramètres seront transmis dans les DAS au COME sélectionné (par l'ordre MPO WRITE_DAS ou par défaut à la dernière tranche de COME mise ON) *lorsque* le COME sera mis ON.

Mais la distinction n'est pas toujours des plus aisées. Les conditions, lorsqu'elles sont nombreuses, sont parfois exprimées sous la forme d'une liste à puces ou tirets. Cette solution est généralement plus lisible, mais le risque (qui se vérifie régulièrement dans le corpus) est d'oublier de préciser si toutes les conditions doivent être remplies (« et »), seulement une (« soit ») ou au moins une (« ou ») :

> [...] Cet ordre est rejeté *si* :
> - le mode NORM automatique est actif
> - le satellite est en mode MAN
> - le satellite n'est pas en mode convergé (GAP ou SUP)
> - un ordre MAN/CAP est déjà en attente d'exécution

Il n'est évidemment pas non plus possible (sauf éventuellement changement de la structure profonde de la phrase) de se passer de la conjonction « que » introduisant une complétive, que l'on retrouve très souvent, dans le corpus, après les verbes « vérifier » et « contrôler » notamment :

> Le générateur de TCH vérifiera *que* l'APID du paquet TM correspond à un paquet TM de type asynchrone (type REPORT).

> [...] Le générateur de TCH contrôlera *que* le type de TM sera diagnostic (DIAGNOS) ou cyclique (HK).

C'est aussi le cas de la plupart des conjonctions de coordination. La conjonction « ou », par exemple, est souvent indispensable :

> Le générateur de TCH vérifiera que la surveillance (définie par PROCESS_ID et SM_ID) correspond à une surveillance de type analogique (ANA_RAW *ou* ANA_PHYS) *ou* flottant (FLOAT). [...]

Ici, la surveillance peut être de type « analogique » *ou* « flottant » ; les deux types sont acceptés. De même, au type analogique peuvent correspondre « ANA_RAW » et « ANA_PHYS », qui doivent donc être mentionnés tous les deux ensemble et être coordonnés ; il ne serait pas possible de faire deux exigences distinctes, puisqu'elles seraient alors toutes les deux incomplètes.

> Sur réception de cette TC, le LVC recherche puis supprime dans le buffer MDP correspondant au MDP_ID la *ou* les séquences ayant une heure bord supérieure *ou* égale à HB.

Les deux « ou » présents dans cette exigence sont eux aussi inévitables : le premier précise qu'il peut y avoir une ou plusieurs séquences, le second que l'heure

bord peut être soit supérieure, soit égale à HB (« supérieure à » aurait laissé penser qu'elle devait être strictement supérieure). S'en passer entraînerait une perte d'information.

Il existe toutefois deux risques associés à l'utilisation de « ou ». Le premier tient à son ambiguïté : indique-t-il une disjonction inclusive (« et/ou ») ou une disjonction exclusive (« soit… soit… » ; « ou bien ») ? Cette distinction est parfois explicitée :

> Les fenêtres d'indisponibilité de la charge utile définissent les créneaux indisponibles pour la prise de vue *et/ou* le vidage des demandes d'acquisition des clients : […]

> […] Les objets obtenus à partir d'une demande PHR particulière peuvent être *soit*[a] insérés dans l'espace de travail courant (i.e. ouvert), *soit* dans un nouvel espace de travail.
>
> *a.* [↑]Il aurait été préférable de placer le premier « soit » après « insérés ».

Mais cela n'est généralement pas le cas. Il faut alors supposer que cela peut être déduit du contexte, ou bien que cela est sans incidence. Le second risque est qu'une alternative est parfois proposée sans qu'il soit précisé comment ou pourquoi opter pour l'une ou l'autre des options possibles :

> L'altitude doit être exprimée en mètre *ou* en pied.

> […] Pour cela, on utilisera les données BDS (LENGTH et LOCATION_UNIT) de la table des OBCD (globaux) *ou* la description (LONGUEUR) des paramètres diagnostic déjà crées.

Rien ne dit dans quel cas choisir une mesure ou l'autre, ou pourquoi il est possible d'utiliser des données différentes pour arriver au même résultat. À nouveau, il faut espérer que si cette information n'est pas précisée, c'est parce qu'elle n'est pas vraiment nécessaire ou peut être facilement déduite.

En fin de compte, il semblerait que « et » et « mais » soient les seules conjonctions qui puissent être réellement évitées. « Mais » (assez rare) a cependant l'intérêt d'apporter une information logique (opposition, restriction, exception, etc.) et peut permettre de préciser quelque chose qui n'irait pas de soi :

> […] Ces contrôles seront implémentés dans GOTLIB *mais* devront être explicitement appelés par le DUPC.

En tant que connecteur logique, il participe de la cohérence de l'exigence en tant que (court) texte ; ces phrases coordonnées sont probablement plus claires pour le lecteur qu'une simple juxtaposition.

Quant à « et », il y a également de nombreux cas où il est inévitable :

> [...] Remarque : la surveillance ne sera exécutée que si la surveillance fonctionnelle est autorisée *et* activée par le LVC.

Il y a deux conditions nécessaires pour que la surveillance soit exécutée, celles-ci doivent donc toutes les deux être exprimées[74].

> [...] Avant calibration, le taux d'off-modulation moyen est entre 0,67 *et* 1 (DR1), ce qui génère un étalement de poussée entre 0 *et* 20%. [...]

Les deux « et » utilisés dans ce fragment d'exigence sont indispensables pour indiquer les bornes inférieure et supérieure d'un intervalle. Plus généralement, « et » est inévitable après la préposition « entre » (ici pour une « interface ») :

> Cette TC permet de sélectionner le MSI utilisé comme interface TM/TC entre le LV *et* l'instrument. [...]

Il peut aussi servir à exprimer une relation partitive (méronymie) :

> [...] Ce vecteur d'état est composé d'une position, d'une vitesse *et* du temps associé. [...]

Ici, les trois co-méronymes (dont l'holonyme est « ce vecteur d'état ») doivent être énumérés ensemble, sans quoi, une fois de plus, l'exigence serait incomplète.

> [...] En conséquence, les durées de PDV XS *et* PAN ne seront pas identiques.

Pour pouvoir indiquer que des éléments (ne) sont (pas) identiques, il est commode de pouvoir les lier. On recense encore d'autres exemples, qu'il serait inutile de lister de façon exhaustive.

On trouve par ailleurs de nombreux cas où l'emploi de « et », bien qu'il pourrait être évité, peut sembler justifié :

> Les valeurs des paramètres MIN *et* MAX seront saisis en valeurs physiques. [...]

> [...] Les contrôles " premier paquet " *et* " dernier paquet " consistent à vérifier les valeurs du champ CCSDS " Group flag ". [...]

Dans ces deux extraits, coordonner deux éléments avec « et » permet d'éviter de répéter des phrases (parfois longues) qui ne différeraient que très peu (ici par un seul mot). Les regrouper au sein d'une même exigence nous paraît d'autant plus justifié que ces éléments (les paramètres « MIN » et « MAX », les contrôles « premier paquet » et « dernier paquet ») semblent très clairement liés. Dans l'exemple

74. ↑En logique des propositions, cette phrase pourrait être formalisée ainsi : $(p \wedge q) \iff r$.

qui suit, en particulier, si l'exigence était scindée en deux pour éviter « et », les deux phrases qui en résulteraient ne se distingueraient que par un seul caractère – et il n'est dès lors pas improbable qu'elles puissent être considérées à tort comme des doublons par un œil insuffisamment attentif :

> Les champs SM_ID *et* FM_ID seront extraits à partir de la BDS, aucun contrôle n'est donc nécessaire (on considère la BDS validée).

La conjonction « et » peut occasionnellement lier plus que deux éléments (ici, trois adjectifs épithètes) sans que cela n'obscurcisse l'exigence :

> [...] Les paramètres sont des nombres entiers, non signés, *et* codés sur 16 bits. [...]

Il semblerait indéniablement artificiel de rédiger trois exigences distinctes pour exprimer ces trois propriétés des paramètres.

Nous pensons même qu'il peut être justifié de coordonner deux actions différentes, à condition que celles-ci soient clairement liées et qu'elles aient leur sujet en commun (ici, « le LVC ») :

> Sur réception de cette TC, le LVC met à jour la table des surveillances standards de l'application destinataire *et* ré-initialise le compteur d'erreur (remise à 0) associé à cette surveillance. [...]

> Sur réception de cette TC, le LVC supprime de la table des historiques les N premiers évènements *et* met à jour le paramètre " nombre d'historiques ".

Dans les deux exemples ci-dessus, il ne fait pas de doute que les deux actions sont liées ; en fait, la seconde n'aurait vraisemblablement pas lieu d'être sans la première. On peut discuter de la pertinence de rédiger deux phrases au lieu d'une, mais nous pensons en tout cas qu'il est approprié de les regrouper au sein de la même exigence (mais ce choix reste au final celui des rédacteurs). Dans tous les cas, il serait toutefois préférable de préciser si ces actions doivent être réalisées simultanément ou (comme c'est sans doute le cas ici) séquentiellement. Dans ce dernier cas, « et » pourrait judicieusement être remplacé par « puis » – comme c'est d'ailleurs le cas dans certaines exigences :

> Sur réception de cette TC, le LVC envoie les commandes de sélection de l'horloge demandée à l'instrument pour l'ensemble des 4 voies *puis* met à jour la donnée contenant l'identificateur de l'horloge sélectionnée.

Il arrive que les séquences de ce genre comportent plus de deux actions :

> Sur réception de cette TC, le LVC entre dans l'état " GIMB_MAN ", calcule le profil de manœuvre *et* exécute la manœuvre cardan pour atteindre la position canonique. [...]

Dans ce cas, il peut être opportun d'opter pour une liste à puces, solution parfois choisie par les rédacteurs :

> Sur réception de cette TC, la fonction de gestion du COME est mise ON et le LVC :
> - Exécute la procédure de mise ON de la fonction contrôle [...]
> - Exécute la procédure de mise ON des convertisseurs sélectionnés [...]
> - [...]

Les listes permettraient également d'éviter les phrases longues et moins lisibles :

> Cette TC permet de passer contrôle thermique plate-forme en mode REDUCED, c'est-à-dire de sélectionner des seuils de régulation " larges " pour le contrôle thermique grossier (pour limiter la puissance consommée), *et* de modifier la valeur d'écrêtage de la puissance injectée pour le contrôle thermique fin. [...]

En revanche, on a plus de mal à justifier la coordination de propositions qui ne partagent aucun élément commun :

> Sur réception de cette TC, le LVC exécute la procédure de mise ON de la fonction IOT. L'automate de la fonction IOT passe à ON, *et* l'ensemble des traitements de niveau fonction est séquencé.

Dans l'hypothèse où les deux événements décrits dans la seconde phrase correspondent à la « procédure » évoquée dans la première, il conviendrait de le signaler plus clairement. Si, au contraire, il s'agit d'actions supplémentaires, il semblerait préférable de séparer clairement les deux phrases, qui n'ont en apparence rien en commun.

Même lorsque le sujet est commun, il est préférable de scinder en deux phrases lorsque les circonstances (exprimées par des subordonnées) sont différentes :

> Un ensemble fonctionnel passe en alarme dès qu'un paramètre qui le compose répercute une alarme sur lui (alarme en cascade), *et* redevient normal lorsque les alarmes de tous les paramètres qu'il recouvre ont disparu.

Ici, la première action (« passe en alarme ») correspond à un événement introduit par « dès que », qui est clairement distinct de celui (introduit par « lorsque ») qui provoque la seconde action (« redevient normal »). Il serait sans doute plus clair de rédiger une phrase précisant quand passer en alarme et une autre quand redevenir normal, pour qu'à chacune corresponde une seule condition ou série de conditions.

Enfin, nous avons relevé quelques cas où l'utilisation de « et » semble plutôt injustifiée, tant les éléments qu'il lie semblent exprimer des besoins ou des contraintes différents :

> Les demandes sont saisies sur le FOS *et* le logiciel ARPE gère les conflits entre les demandes Spot, Hélios et Pléïades. [...]

En dépit du lien entre les deux propositions (représenté par « les demandes »), il s'agit clairement de besoins différents pour lesquels tous les actants sont différents.

> [...] Le format des données de mesure angulaire et Doppler est conforme au standard CCSDS décrit dans le document DA9 *et* le schéma XML respecte le standard décrit dans DA11.

Même si dans les deux cas, il s'agit de respecter une norme, la norme en question autant que ce qui doit s'y conformer sont différents d'une proposition à l'autre.

3.5.2.3 *Synthèse*

Il est cette fois plus compliqué d'établir une règle opérationnelle. Il appert que la recommandation du GWR est irréaliste : il n'est pas concevable de se passer par exemple des conjonctions « si » et « ou ». En réalité, seule la conjonction de coordination « et » pourrait être évitée – et encore, pas toujours. Du reste, même quand elle peut l'être, il n'est pas certain que cela soit bénéfique, car elle permet très souvent d'éviter des répétitions qui alourdiraient excessivement les spécifications au point sans doute, si l'on poussait la logique à l'extrême, de les rendre indigestes.

Il est vrai cependant que dans certains cas, son utilisation peut sembler assez inopportune et qu'il conviendrait sans doute de s'en passer pour privilégier la rédaction de plusieurs phrases, voire de plusieurs exigences distinctes. Mais lorsqu'il s'agit d'identifier ces cas, la difficulté est double : d'une part, parce qu'à peu près tous les éléments peuvent potentiellement être coordonnés, quels que soient leur nature (adjectif, adverbe, nom, etc.), leur fonction dans la phrase (sujet, complément d'objet, circonstancielle, etc.) ou leur niveau syntaxique (mot, syntagme, proposition, phrase) et, d'autre part, parce que décider s'il est justifié de coordonner plusieurs éléments ne peut être le fait que d'un être humain disposant d'une bonne connaissance du domaine. Seul le rédacteur peut en effet réellement savoir s'il est acceptable de regrouper deux phrases dans une même exigence, voire deux éléments dans une même phrase[75]. Le cas qui semble être le plus souvent problématique est logiquement celui où « et » lie deux propositions qui n'ont aucun élément en commun[76].

75. ↑Pour faire une analogie : « Je dois faire des courses et le ménage » paraît être une phrase tout à fait sensée, alors que la phrase « Je dois faire des courses et le bien », qui lui est pourtant semblable d'un point de vue syntaxique, semble dénuée de logique apparente et la coordination injustifiée (sauf figure de style).

76. ↑Mais on peut arriver au même résultat sans utiliser de conjonction, par juxtaposition (« Les valeurs des arguments saisis par le CMSM sont les " libellés pour fiches session ", les valeurs brutes correspondantes sont déduites par le CECT au moment de la création (encodage) de la TC. »).

Pour simplifier, on pourrait se contenter de distinguer, au sein de la phrase, le sujet et le prédicat, et décider qu'il est permis de coordonner plusieurs sujets différents si le prédicat est le même[77], ou, inversement, d'attribuer plusieurs prédicats différents à un même sujet[78] – mais dans ce second cas, il n'est pas certain que ce regroupement soit toujours justifié d'un point de vue pragmatique, les prédicats pouvant être complètement différents. Cette distinction binaire sujet/prédicat pourrait alors être élargie à une relation ternaire sujet/verbe/objet (ou attribut) selon la même logique, en exigeant que deux des trois constituants soient communs pour que le seul constituant variable puisse faire l'objet d'une coordination ; qu'il s'agisse du sujet[79], du verbe[80] ou bien de l'objet[81]. Il est possible que la présence de deux constituants en commun (sur trois)[82] garantisse une certaine similarité, qui elle-même justifierait le regroupement au sein d'une même phrase.

Néanmoins, à supposer que cette règle ait un intérêt, il serait illusoire de vouloir imposer aux rédacteurs une telle réflexion et il n'est pas certain non plus qu'une détection automatique par un analyseur syntaxique soit vraiment efficace, en particulier face aux phrases – évidemment nombreuses – qui s'éloignent de la structure canonique sujet/verbe/objet. Pour cette raison, nous préférons la laisser de côté pour le moment et privilégier une recommandation beaucoup plus générale et moins stricte pour les rédacteurs, qui leur conseillerait, lorsqu'ils souhaitent utiliser « et », de considérer la possibilité d'utiliser plutôt une liste à puces (surtout s'il y a plus de deux éléments ou que la phrase est longue, ce qui peut leur être signalé par un outil) si cela est possible ou de faire deux phrases différentes, voire deux exigences différentes si cela se justifie. En partant du principe qu'ils doivent avoir été sensibilisés au principe d'atomicité des exigences et savoir comment ces exigences seront mises en œuvre, tracées et vérifiées, il faut admettre qu'ils sont les mieux placés pour savoir s'il convient ou non d'y coordonner des éléments. Pour les aider concrètement dans leur activité, nous pensons que la meilleure solution serait de leur signaler automatiquement tous les « et » qui coordonnent deux propositions indépendantes, puisque ceux-ci sont souvent les moins justifiés. Si l'on part du principe qu'un verbe correspond à une action et que chaque action doit être traitée séparément, on peut également signaler les « et » qui coordonnent des verbes.

On peut également proposer quelques règles plus spécifiques, liées à certaines conjonctions en particulier. Toujours à propos de « et », lorsqu'il coordonne plu-

77. ↑« Ntuy *et* Fscao sont des paramètres de la méthode de calcul de la manœuvre ».
78. ↑« Ces messages sont répartis de manière équidistante (tous les deux slots) *et* numérotés de 1 à 16. ».
79. ↑« Le masque (MASQUE) *et* la valeur attendue (COMP_VAL) devront être saisi au format hexadécimal. ».
80. ↑« Sur réception de cette TC, le LVC extrait *et* vérifie l'ensemble des séquences du buffer plan mission préalablement chargé. ».
81. ↑« Le LVC met à jour les éléments de la matrice d'inertie satellite *et* les gains des différents types de contrôleur en normal mode ».
82. ↑Il y a d'autres cas, comme cette structure elliptique où seul le verbe (d'état) est le même, ce qui entraîne son élision dans la deuxième partie : « La dynamique est donc de [0-63] et la précision de 1/1024 (LSB). ».

sieurs actions et si ces actions forment une séquence, il pourrait être remplacé par « puis » pour signaler que l'ordre de leur réalisation n'est pas indifférent. Concernant « ou », il peut être utile de préciser s'il s'agit d'une disjonction exclusive, par exemple en le remplaçant par « soit ». Nous recommandons que les conditions soient introduites uniquement par « si » ; « lorsque » devrait être utilisé uniquement pour introduire les événements qui déclenchent une action, et « tant que » pour les états qui peuvent durer plus longtemps. Enfin, il serait très important de préciser, pour chaque liste de conditions, si toutes, seulement une ou au moins une doivent être remplies.

3.5.3 *Longueur des phrases*

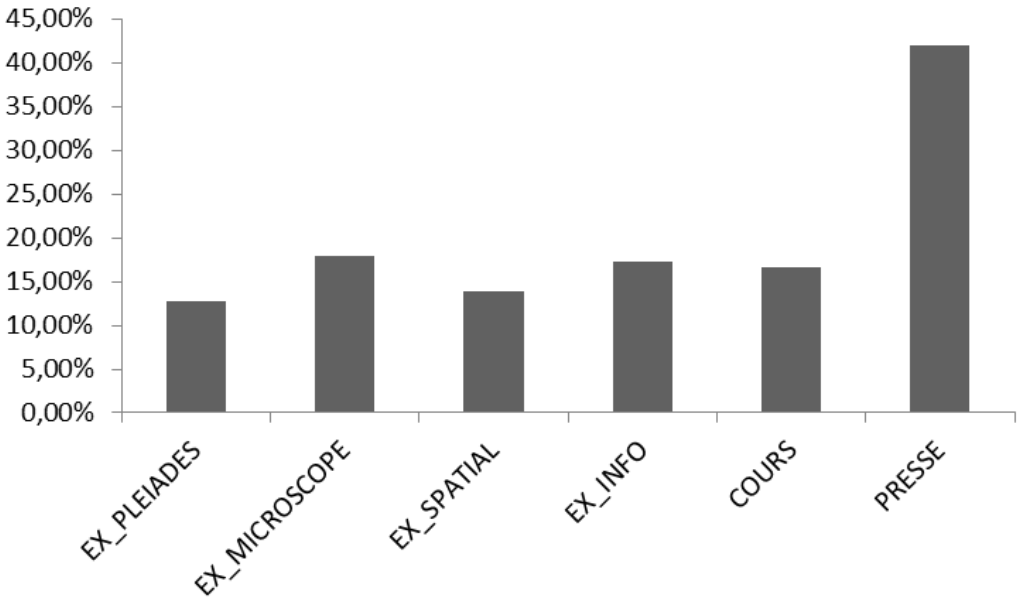
3.5.3.1 *Analyse quantitative*

Hypothèses : la fréquence relative des phrases longues (plus de vingt-cinq mots) et la longueur moyenne des phrases sont plus faibles dans le corpus d'exigences que dans les autres textes.

TABLEAU 8 – Longueur des phrases et fréquence des phrases longues (par rapport au nombre total de phrases)

	Nombre de phrases	Nombre de phrases > 25 mots	Longueur moyenne	Longueur maximale
EX_PLEIADES	8779 (3,47/exigence)	1120 (12,76 %)	13,44	92
EX_MICROSCOPE	2572 (2,41/exigence)	460 (17,88 %)	16,86	56
EX_SPATIAL	11351 (3,16/exigence)	1580 (13,92 %)	15,15	92
EX_INFO	3494 (4,13/exigence)	604 (17,29 %)	16,52	81
COURS	7645	1271 (16,63 %)	15,53	85
PRESSE	6198	2603 (42,00 %)	25,4	131

FIGURE 6 – Fréquence des phrases longues (par rapport au nombre total de phrases)



Nos deux hypothèses (liées) se confirment grâce au TABLEAU 8 et à la FIGURE 6 : les exigences spatiales contiennent les phrases les plus courtes en moyenne (15,15 mots) et la plus faible proportion de phrases « longues ». C'est néanmoins essentiellement dû à Pléiades, Microscope se situant au même niveau (et même légèrement au-dessus) que les exigences informatiques et le cours. On constate donc ici une différence entre les deux sous-corpus, mais elle est sans commune mesure avec la différence entre les exigences et la presse, où les phrases longues sont très fréquentes (près d'une sur deux), la moyenne de mots par phrase étant même supérieure à vingt-cinq (ce qui indique la présence de phrases très longues). Quoi qu'il en soit, la proportion de phrases « longues » est loin d'être négligeable dans chacun des textes, y compris dans les exigences spatiales (près d'une sur sept).

On ne peut savoir avec certitude si les phrases en moyenne plus courtes dans les exigences résultent d'un effort conscient de concision de la part des rédacteurs ou s'il s'agit d'une tendance naturelle, mais on constate que cela n'empêche pas la présence de phrases exceptionnellement longues, telles que celle-ci :

> Pour chaque maille de chaque demande d'acquisition qui intersecte au moins une "zone d'accès restreint" définie avec des restrictions d'accès, lorsque les deux champs "ORG" et "INTERL" sont vides ou lorsque l'organisme n'est pas autorisé ou lorsqu'au moins une de ses stations d'inventaire n'est pas autorisée, le logiciel doit invalider toutes les mailles de cette demande d'acquisition (même celles qui ne sont pas en intersection) c'est-à-dire indiquer qu'aucun de ces essais ne doit être programmé ainsi que la raison de leur inactivation. [...]

qui compte pas loin de cent mots (et seulement deux virgules !). Cette exigence comporte même une seconde phrase presque aussi longue que la première.

3.5.3.2 *Analyse qualitative*

Les phrases considérées comme « longues » peuvent se répartir dans deux grandes catégories : celles qui peuvent facilement être raccourcies et celles qui ne le peuvent pas vraiment. Dans la première, on retrouve principalement celles qui contiennent une conjonction de coordination additive, ce qui nous renvoie directement au phénomène précédent :

> [...] (l'étape B2) introduit ensuite les demandes d'acquisition civiles, en fonction de leur priorité, *et* partage le reste des ressources disponibles, s'il y en a, entre tous les acteurs, tant civils que défense, suivant les règles de partage des ressources. [...]

> Dans le Segment sol Utilisateur civil, le recensement routine et le recensement civil crise (sous-système " PHR Programming Request Management ") calculent la note des essais conformément à la méthode décrite dans [DOC205], *et* génèrent les demandes d'acquisition en initialisant leurs données de priorité en fonction de celles des essais liés.

Dans ces deux exemples, on pourrait facilement diminuer la longueur de la phrase en la scindant en deux phrases plus courtes. (Remarquons que dans ce type de cas, le « et » est fréquemment précédé d'une virgule.) Comme déjà suggéré, il est parfois possible de remplacer la conjonction par une liste à puces :

> [...] Le contrôle consiste à vérifier que l'on n'a pas perdu de paquet TM pendant le vidage et que l'on a bien reçu le premier et le dernier paquet. [...]

Ici, les deux actions correspondant à la vérification, toutes deux introduites par « que », pourraient être remplacées par autant de tirets et de retours à la ligne. On peut néanmoins s'interroger sur l'absolue nécessité de réduire la longueur de cette phrase qui, quand bien même elle compte plus de vingt-cinq mots, ne semble pas vraiment poser de difficultés.

On pourrait ajouter les phrases qui contiennent des mots superflus :

> Le contenu des TCH étant défini complètement en BDS, le générateur de TC s'attachera à utiliser autant que possible la BDS de manière à générer les contenus binaires des TCH.

Dans cette exigence, « s'attachera à utiliser autant que possible » pourrait sans doute être raccourci en « utilisera autant que possible », avec le même effet.

> Pour la génération d'un TCH relative aux surveillances fonctionnelles, on commencera au préalable par lister la liste des surveillances fonctionnelles (infos dans BDS) afin que l'opérateur effectue sa sélection. [...]

Ici, « on commencera au préalable par lister la liste des surveillances fonctionnelles » semble doublement redondant ; « on listera au préalable les surveillances fonctionnelles » ou « on commencera par lister les surveillances fonctionnelles » seraient probablement plus adéquats et permettent d'économiser quelques mots. Il est cependant impossible de tirer de ces observations des conclusions suffisamment générales pour proposer des règles de rédaction efficaces. Qui plus est, le bénéfice réel obtenu en réduisant la longueur de l'exigence (pour rester sous une limite de mots imposée) n'est pas franchement évident ; sans compter que la redondance (omniprésente dans les langues naturelles) n'est pas toujours une mauvaise chose.

Dans la deuxième catégorie, on retrouve de nombreuses phrases de plus de vingt-cinq mots dont la réduction se ferait au prix de modifications compliquées ou artificielles :

> Sur réception de cette TC, le LVC exécute la procédure de mise ON des réchauffeurs des lits catalytiques des tuyères sélectionnées pour la manoeuvre, par l'envoi de commandes sur le bus 1553 ICB.

> La direction de poussée moyenne est définie par un simple vecteur donnant les cosinus directeurs de la direction de poussée moyenne dans le repère à piloter (= repère de visée Rv). [...]

> [...] Cette prédiction est utile à la gestion de la contrainte sur le moment cinétique perturbateur induit par la poussée, dans les cas où cette contrainte pénaliserait trop la capacité de poussée maximale (i.e. configurations dégradées à 3 AG ou 2 tuyères uniquement). [...]

> Afin de corriger a bord les erreurs le long de la trace entre le modèle d'orbite sol et la position orbitale fournie par DORIS, on enverra un ordre MDP de recalage. [...]

3.5.3.3 *Synthèse*

Bien qu'elles soient absentes du GWR, les règles liées à la longueur des phrases sont très populaires dans les langues contrôlées. Il faut sans doute y voir un héritage des formules de lisibilité, où longueur rime presque toujours avec complexité, au point que certaines de ces formules soient en grande partie basées sur ce principe. Il semble bien exister une corrélation (que l'on peut percevoir intuitivement) entre longueur et complexité, mais elle n'est évidemment pas parfaite : une phrase longue peut très bien être plus « simple » qu'une phrase plus courte (pour des facteurs lexicaux, syntaxiques ou autres).

Nous pensons, comme beaucoup d'études l'ont montré, que la longueur peut effectivement être un bon indicateur de la complexité, à condition que ce ne soit pas le seul élément pris en compte. À ce titre, il nous semble qu'elle pourrait être naturellement rapprochée des règles concernant les conjonctions (en particulier « et ») : si une phrase est longue et qu'elle contient des conjonctions reliant des éléments qui pourraient tout aussi bien se retrouver dans des phrases distinctes (typiquement des propositions indépendantes), il y a probablement tout intérêt à la scinder. De nouveau, donc, une recommandation aux rédacteurs – mais certainement pas une obligation stricte comme celle d'ASD-STE, qui n'a selon nous aucun sens si elle les amène à retirer de l'information pourtant utile pour diminuer le nombre de mots[83] – les encourageant à envisager la possibilité de scinder leur phrase si elle est longue ou d'utiliser une liste à puces quand c'est possible nous paraît pertinente. Pour les y aider, signaler les « et » semble une fois de plus être

83. ↑« Some researchers have pointed out, however, that breaking up sentences can not only produce a choppy, monotonous style but also interfere with understanding by removing relationships between clauses [...]. [...] various word-processing tools will point out sentences over a certain length [...]. Such tools give the impression that any sentence over a certain length (such as 25 words) should be changed, and yet the writer may well decide that some long sentences are indeed understandable and necessary. [...] Under pressure to reduce the number of words in a sentence, writers sometimes cut out little words such as prepositions, *that* introducing a clause, and parts of verbs. Research has since identified some of these words as semantic cues and shown that they make a difference in how easily translators and nonnative speakers of English understand a sentence (Kohl, 1999). » (Hargis, 2000 : 126-127).

un moyen simple et sans doute efficace et, de ce point de vue, l'exemple donné par ASD-STE pour scinder une phrase au niveau d'un « connecting word » nous semble également utile.

Là où nous entrons en désaccord avec ASD-STE, c'est sur la façon dont les mots doivent être comptés. Pour nous, il va de soi que cette tâche doit être réalisée avant tout par un outil automatique (ce qui ne pose pas vraiment de problème pour le traitement automatique du langage, de nombreux logiciels courants proposant déjà cette fonctionnalité) et sur la base de règles simples à maîtriser et à implémenter. Même si les concepts de « mot » et de « phrase » sont difficiles à définir exactement pour les linguistes, on peut sans trop de problème établir des critères (typographiques) opératoires permettant d'en établir les limites. Comme nous l'avons dit, dépasser un certain nombre de mots ne devrait pas entraîner le rejet de la phrase, mais un simple avertissement : à charge pour le rédacteur de décider s'il doit en tenir compte ou pas. Il resterait à définir cette limite, qui ne peut être qu'arbitraire. Pour cela, nous disposons d'un indicateur objectif : le nombre moyen de mots par phrase (quinze, la médiane étant de douze) dans les exigences du corpus. Plutôt qu'une dichotomie (« acceptable » ou « trop longue »), une échelle un peu plus fine pourrait être mise en place : « normale », « plutôt longue » (par exemple, à partir de vingt-cinq mots), « longue » (à partir de trente ou trente-cinq mots), « très longue » (au-delà de quarante mots). Cela offrirait une meilleure indication aux rédacteurs, éventuellement en temps réel (directement pendant la rédaction).

3.5.4 *Temps verbaux*

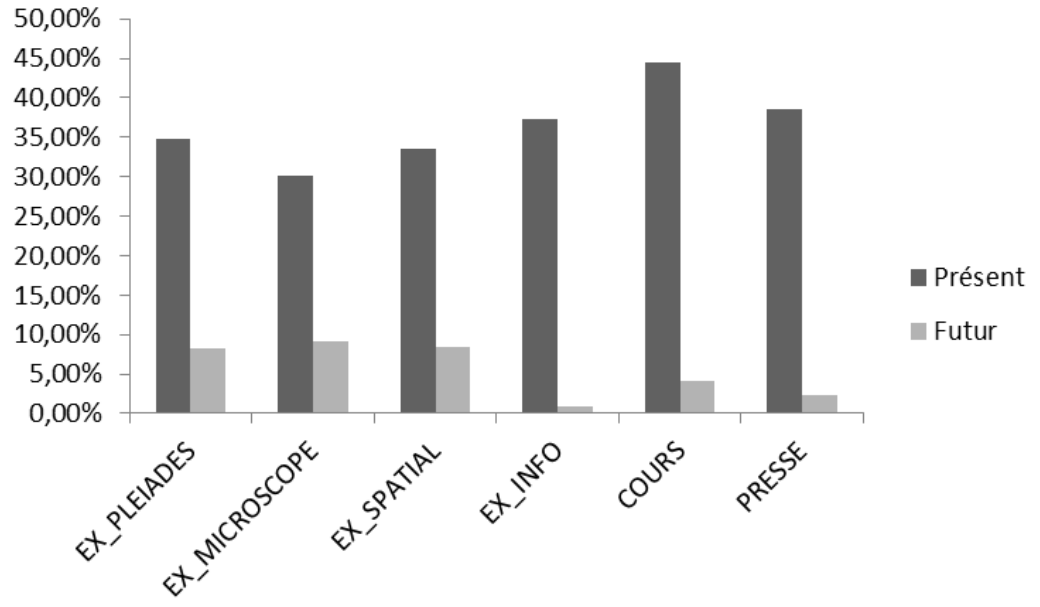
3.5.4.1 *Analyse quantitative*

Hypothèse : la fréquence relative des verbes conjugués au présent et/ou au futur simple est plus grande dans le corpus d'exigences que dans les autres textes.

TABLEAU 9 – Fréquence de certains modes et temps (par rapport au nombre total de formes verbales)

	Verbes	Présent	Futur	Conditionnel	Impératif	Infinitif
EX_PLEIADES	15516	5384 (34,70 %)	1271 (8,19 %)	35 (0,23 %)	0 (0,00 %)	3173 (20,45 %)
EX_MICRO.	5528	1664 (30,10 %)	507 (9,17 %)	8 (0,14 %)	0 (0,00 %)	1190 (21,53 %)
EX_SPATIAL	21044	7048 (33,49 %)	1778 (8,45 %)	43 (0,20 %)	0 (0,00 %)	4363 (20,73 %)
EX_INFO	8759	3261 (37,23 %)	83 (0,95 %)	12 (0,14 %)	0 (0,00 %)	2835 (32,37 %)
COURS	15830	7028 (44,40 %)	651 (4,11 %)	66 (0,42 %)	0 (0,00 %)	2747 (17,35 %)
PRESSE	22054	8508 (38,58 %)	509 (2,31 %)	494 (2,24 %)	2 (< 0,01 %)	3887 (17,62 %)

FIGURE 7 – Fréquence du présent et du futur (par rapport au nombre total de formes verbales)



Pour interpréter correctement ces résultats (TABLEAU 9 et FIGURE 7), il faut tout d'abord garder à l'esprit que les verbes peuvent servir d'auxiliaires pour la formation des temps composés : les chiffres correspondant aux verbes conjugués au présent de l'indicatif englobent donc ceux au passé composé (« il *a* rencontré »), ce qui peut expliquer qu'ils soient surreprésentés. Cette remarque vaut surtout pour les textes journalistiques, qui ont ceci de particulier (par rapport aux autres genres représentés dans cette étude) qu'ils sont les seuls à relater des faits passés et à annoncer des événements futurs.

Quant aux exigences, on a déjà expliqué la situation particulière et la « tension » qui les caractérisent : elles consistent à décrire un objet qui n'existe pas encore, mais en faisant comme s'il existait déjà et répondait aux propriétés qui lui sont attribuées. Cela peut y justifier à la fois l'usage du présent (qui ne peut être que non déictique) et du futur (qui lui, au contraire, est bien repéré par rapport au moment de l'énonciation).

En effet, on constate qu'une forme verbale sur trois environ est conjuguée au présent de l'indicatif dans les exigences spatiales (Pléiades et Microscope étant une fois de plus assez proches), ce qui est considérable, mais que cette proportion reste sensiblement plus faible que dans les autres types de textes (en particulier le cours). Leur vraie spécificité, toutefois, tient en l'usage du futur, beaucoup plus répandu qu'ailleurs et qui trahit leur fonction « programmatique » (un peu à la manière d'un manifeste). C'est aussi une différence importante avec les exigences informatiques, qui l'utilisent pour leur part significativement moins. Cet écart ne peut, selon nous, s'expliquer que par une différence dans les habitudes rédactionnelles et montre bien que les deux options (présent ou futur) sont possibles, qu'elles tendent à se mêler au sein des spécifications et qu'elles peuvent chacune s'exprimer plus ou moins fortement selon les (communautés de) rédacteurs. Ces résultats justifient *a posteriori* d'avoir traité les exigences informatiques à part du corpus. Il resterait à déterminer si l'habitude des rédacteurs d'exigences de projets spatiaux d'employer le futur est totalement spontanée, ou si elle est partiellement apprise, par calque de l'anglais « shall ».

Le faible nombre de verbes au futur dans les exigences informatiques est compensé par une utilisation légèrement plus grande du présent, mais surtout des infinitifs (une forme verbale sur trois). Bien que cela soit nettement moins marqué, on retrouve aussi un peu plus d'infinitifs dans les exigences spatiales que dans les deux autres textes. Cette tendance générale peut s'expliquer par le recours plus fréquent aux verbes modaux (« pouvoir » et surtout « devoir »), qui sont nécessairement suivis d'un ou plusieurs infinitifs exprimant l'action autorisée ou imposée.

Concernant les autres temps ou modes repris dans le tableau (conditionnel, impératif), ils sont pratiquement anecdotiques, voire totalement absents de certains textes, ce qui est logique. Les participes passés (non repris) sont en revanche fréquents, ce qui n'est pas étonnant puisqu'ils participent à la formation des temps composés, mais aussi des voix passives.

3.5.4.2 *Analyse qualitative*

Quelques verbes sont au conditionnel dans les exigences, ce qui peut surprendre dans ce contexte. Il peut s'agir de présenter une éventualité (souvent introduite par « dans le/au cas où ») :

> [...] Elle permet aussi au sol de redémarrer la séquence au cas où la séquence automatique ne *démarrerait* pas - ce qui ne doit pas arriver nominalement – [...]

> [...] Dans le cas ou le satellite *passerait* en configuration dégradée 2 Tuyères, on recalibrera le bras de levier. [...]

> [...] <Numéro_Version> est un numéro qui pourra être appliqué à un fichier (ou un répertoire) particulier ou à un lot de fichiers (ou un lot de répertoires) pour le cas où un traitement *serait* appelé à être exécuté plusieurs fois, soit en cas d'anomalie, soit dans le cas nominal (exemple du fichier " PHRMissionPlanreport " régénéré suite à anomalie) [...]

> Les procédures d'installation devront permettre de fournir sur un support media, une image du système installé ; cette image *pourrait* alors être utilisé lors d'une nouvelle réinstallation du système ou pour des besoins de test.

Ces verbes pourraient être conjugués au présent de l'indicatif. Le choix du conditionnel revient à présenter ces potentialités comme « anormales » (exceptionnelles) ou en tout cas comme moins probables. Parfois, le conditionnel peut correspondre à un irréel visant à simplifier les calculs :

> [...] GOTLIB prédit une transition chaque fois qu'un passage jour/nuit ou nuit/jour a lieu, telle qu'elle se *déroulerait* si le satellite était en NORMAuto sur tout l'horizon de prédiction. GOTLIB ne tient pas compte des cas où l'automatisme de ces transitions est inhibé (passages en NORMTC). [...]

> [...] L'attitude calculée est l'attitude qu'*aurait* le satellite si aucune programmation n'était faite sur le satellite. [...]

Dans ces deux cas, il s'agit de calculer une valeur en considérant seulement le cas le plus simple.

Plus étonnant dans le contexte des exigences, le conditionnel peut servir à présenter une action qui semble seulement facultative ou à suggérer une solution :

> [...] Cette activité de surveillance *pourrait* être typiquement synchronisée avec la génération du bilan mensuel.

> [...] Le mode opératoire *pourrait* être le suivant : [...]

> [...] Ceci se *ferait* par une fonction de type clic-droit -> masquer. [...]

Ces exemples ne semblent pas correspondre exactement à la définition d'exigence, mais plutôt à une suggestion de conception. Autrement dit, ils n'expriment pas tant un besoin qu'un moyen d'y répondre. La forme « pourrait », en particulier, laisse supposer que cela n'est pas obligatoire, et même « se ferait » semble atténuée par rapport au présent de l'indicatif.

Les exemples au présent de l'indicatif (qui peut être vu comme le temps « non marqué » et le plus polyvalent) abondent évidemment dans tout le corpus, qu'il s'agisse de contraintes valables en permanence :

> L'opérateur *peut* se déconnecter d'une station de travail PHR Access Unit à tout moment.

d'une action à réaliser suite à un événement déclencheur :

> Sur réception de cette TC, le LVC *mémorise* la nouvelle destination pour le routage du paquet TM ayant cet APID.

ou à intervalles réguliers :

> Une fois par semaine (le jeudi), le CCC *traite* et *envoie* la programmation mission reçue du CECT 24 heures auparavant.

ou encore de faits déjà établis (il ne s'agit pas alors d'imposer une action, mais de rappeler un fait ou un état qui lui-même peut représenter une contrainte dont il faut tenir compte) :

> Certaines TC décrites dans ce document *sont* classées " interdites en vol ". Il s'*agit* de TC à usage AIT. [...]

Le présent semble donc à même d'exprimer tout type d'exigence. Pourtant, comme on l'a vu, le futur est également employé (dans des proportions différentes selon les rédacteurs) :

> [...] Les trous de mesure longs *seront gérés* manuellement par GAZMIC en mode expertise.

> Le CECT *aura* un accès au S3W pour la consultation de certaines données.

> Suite à la détection des trous de mesure, le CCC *déterminera* les revidages de télémesure nécessaires pour combler les trous détectés. [...]

> La mise à disposition *se fera* nominalement à un jour fixé dans la semaine, mais *pourra* exceptionnellement être décalée pour gagner en réactivité.

En fait, la plupart des exigences au présent pourraient être rédigées au futur, et inversement. Plus qu'un choix ou l'autre, l'important serait surtout d'être cohérent. On constate pourtant que des exigences aux structures proches sont rédigées parfois au présent, parfois au futur. Ainsi dans les deux exigences suivantes (consécutives dans la spécification) :

> IL *doit* être possible d'acquérir des triplets stéréo monopasse. La longueur de l'image *devra* pouvoir atteindre 20 km avec un B/H minimum de 0,2 pour chaque paire consécutive du triplet.

> Il *devra* être possible d'acquérir en monopasse une zone d'au moins 60 km x 60 km, avec une inclinaison inférieure à 20°.

On peut se demander pourquoi la première phrase de la première exigence est au présent, tandis que la deuxième exigence est au futur. Plus encore, on s'étonnera qu'au sein de la première exigence, la première phrase soit au présent et la seconde au futur. Ce type d'incohérence que rien ne semble ici justifier est très fréquent dans le corpus :

> Les opérations de réactivation getter *doivent* être assez courtes pour ne pas impacter sensiblement le planning : leur durée cumulée *devra* être inférieure à 1 semaine (AC) .

Il aurait indéniablement été préférable d'utiliser le même temps (quel qu'il soit) pour les deux verbes, ce changement de temps n'apportant rien, si ce n'est de la confusion.

> * le CCC *récupèrera* de la station les trames HKTM-R après décodage Reed/Solomon,
> * Le CCC *effectuera* l'extraction des paquets TM,
> * le CCC *effectuera* l'archivage de la totalité des paquets TM (plateforme et scientifique) dans la base de données du G3,
> * le CCC *fournira* au CECT l'ensemble des paquets TM qui lui sont nécessaires au format brut et sans doublons,
> * le CCC *doit* avoir la capacité de récupérer et traiter 291Mo de TM par jour.

À nouveau, il est difficile de justifier le choix d'employer le présent pour le dernier élément de la liste, alors que tous les précédents sont au futur. Bien entendu, dans certains cas, il peut s'expliquer par d'autres raisons (concordance des temps) :

> Si toutes les sessions d'un même type n'*ont* pas la même durée, la formule de recalage en *tiendra* compte. [...]

Ici, si un verbe est au présent et le suivant au futur, c'est parce que la grammaire française n'autorise tout simplement pas le futur dans la protase (**Si… n'auront pas…*).

3.5.4.3 *Synthèse*

Il serait sans doute contre-productif de vouloir régir tous les emplois possibles de tous les temps du français (certains étant d'ailleurs logiquement très rares dans le corpus). En revanche, on a constaté que le conditionnel (qui reste occasionnel) pouvait être un symptôme d'une phrase qui n'a pas sa place dans une exigence, parce qu'elle décrit non pas un besoin, mais une solution. Signaler les occurrences de verbes (surtout « pouvoir ») au conditionnel aux rédacteurs (en précisant par exemple : « Une exigence indique ce que le système doit faire, pas comment il *devrait* le faire. ») pourrait donc les aider à traquer ce type de mauvaises pratiques.

Mais le véritable problème réside dans l'alternance d'exigences au présent et au futur ou, pire encore, de phrases au présent et au futur dans une même exigence. Cette alternance ne présente généralement aucun intérêt (la plupart des verbes au futur pouvant être transposés au présent) et risque d'amener de la confusion chez le lecteur, qui peut être tenté de l'interpréter comme un choix motivé. Il n'en est pourtant visiblement rien et elle résulte sans doute d'un manque de rigueur ou de concertation dans le chef des rédacteurs. On peut l'expliquer sans trop de problème par la simplicité d'écrire au présent (temps plus fréquent et plus polyvalent), d'un côté, et par la tentation naturelle de recourir au futur pour décrire ce qui n'existe pas encore ou exprimer ce qui n'a pas encore lieu, de l'autre. À cela s'ajoutent évidemment des habitudes de rédaction différentes selon les rédacteurs : par exemple, certains utilisent intentionnellement le futur comme calque de l'auxiliaire anglais « shall ».

Présent et futur pourraient tous deux convenir. En faveur du présent, on peut avancer qu'il est d'un usage plus simple (puisqu'il peut être utilisé partout) et qu'il véhicule un sentiment de responsabilité et donc d'obligation plus fort : ce qui est rédigé au présent doit être vrai à tout moment, là où le futur ne correspond qu'à un « plus tard » indéterminé et moins urgent. Employer celui-ci a toutefois pour avantage de permettre de distinguer les contraintes qui nécessitent réellement une intervention pour être vérifiées (écrites au futur) de celles qui le sont déjà (écrites au présent)[84] et permet donc d'identifier plus facilement ce qui devra être spécifiquement mis en œuvre, puis vérifié pour valider l'exigence. Il faut par ailleurs noter que certains verbes ne pourront de toute façon pas être conjugués au futur en raison des règles de concordance des temps en français.

Si le choix n'appartenait qu'à nous, nous pencherions plutôt pour le présent dans tous les cas. Il nous semble en effet plus simple pour tout le monde d'im-

84. ↑Exemple : « Les fichiers de log *devront* être visualisables dans l'application Chainsaw. [= contrainte à respecter, pas encore mise en œuvre] Un schéma phr_log.xsd *est* par ailleurs fourni dans le produit PHR-XCI-P-10. [= fait déjà réputé vrai au moment où l'exigence est rédigée] »

poser systématiquement le présent (déjà nettement plus fréquent) que de devoir imposer le futur dans certains cas et accepter le présent dans certains autres. Nous ne voyons pas d'argument décisif à l'avantage du futur : pour la distinction que nous avons évoquée, le même résultat pourrait être obtenu (peut-être encore plus clairement) en utilisant ou non le modal « devoir » selon l'effet voulu. Les rares cas où le futur semble réellement être la meilleure option sont ceux où il indique l'ultériorité par rapport à une phrase au présent[85]. Néanmoins, il serait utile de consulter les experts du CNES sur la question avant de trancher. C'est ce que nous avons fait à travers l'enquête qui leur a été soumise et dont les résultats seront présentés au chapitre 5.

85. ↑Exemple en circonstancielle : « En cas de serveur SVN indisponible au moment du " commit ", le CECT *doit* poursuivre ses activités et réaliser le dépôt SVN ultérieurement, lorsque le serveur SVN *sera* de nouveau accessible. » Si l'on tient absolument à éviter le futur, d'autres formulations restent généralement possibles, par exemple : « [...] le CECT doit poursuivre ses activités et réaliser le dépôt SVN dès que le serveur SVN redevient accessible ».

3.5.5 *Modalisation*

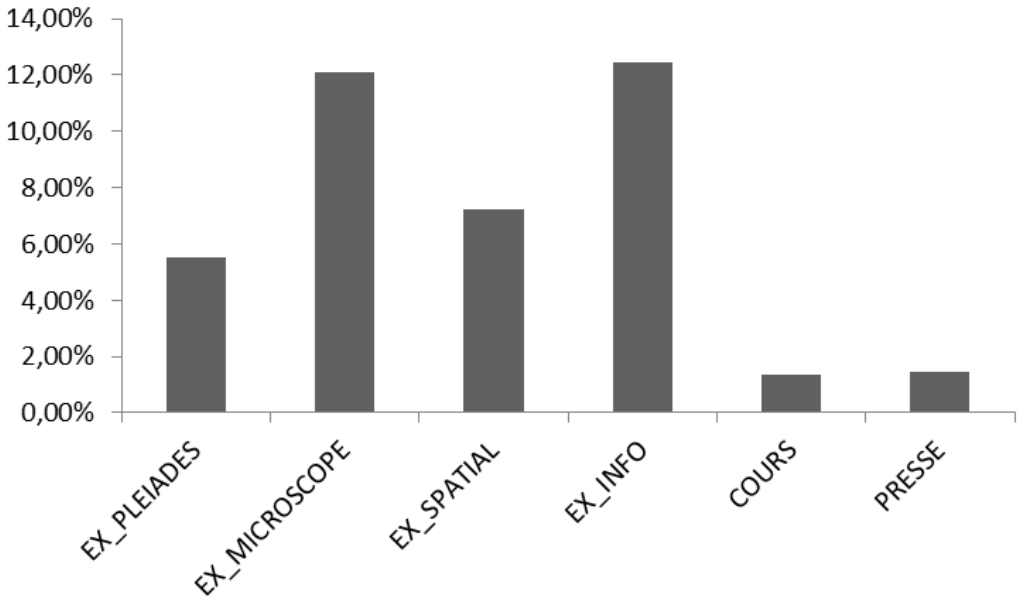
3.5.5.1 *Analyse quantitative*

Hypothèse : la fréquence relative du verbe semi-auxiliaire modal « devoir » est plus grande dans le corpus d'exigences que dans les autres textes.

TABLEAU 10 – Fréquence du modal « devoir » (par rapport au nombre total de verbes)

	« Devoir »
EX_PLEIADES	854 (5,50 %)
EX_MICROSCOPE	669 (12,10 %)
EX_SPATIAL	1523 (7,24 %)
EX_INFO	1091 (12,46 %)
COURS	213 (1,35 %)
PRESSE	317 (1,44 %)

FIGURE 8 – Fréquence du modal « devoir » (par rapport au nombre total de verbes)



Notre hypothèse se voit totalement confirmée par le TABLEAU 10 et la FIGURE 8 (et corrobore en partie ce que nous pensions concernant la fréquence des infinitifs dans les exigences) : le modal « devoir » est beaucoup plus fréquent dans les exi-

gences (spatiales comme informatiques) que dans les deux autres textes. La seule surprise est la forte différence entre Pléiades d'une part et Microscope (et les projets informatiques) d'autre part.

Sachant les résultats précédemment obtenus, on peut supposer que les exigences spatiales (surtout Microscope) utilisent régulièrement les formes « devra / devront », tandis que les exigences informatiques privilégient « doit / doivent ». Dans les deux cas, l'objectif des rédacteurs est de marquer explicitement l'injonction.

3.5.5.2 *Analyse qualitative*

Tout comme pour l'alternance entre présent et futur, l'utilisation non systématique du modal « devoir » ne semble pas répondre à une logique précise. On trouve ainsi des phrases contenant une forme conjuguée (au présent ou au futur) de « devoir » :

> La définition du créneau interdit *doit* prendre en compte une marge de 2.0 secondes [...].

> [...] Le CCC *devra* être capable de générer les TC checksum associées au chargement de ces tables. [...]

Mais on trouve aussi leur équivalent sans :

> La modélisation *prend* en compte les perturbateurs internes et externes : [...]

> [...] Le CCC *sera* capable de générer le checksum et la TC checksum sur les tables de coefficients d'égalisation. [...]

De nouveau, ce changement peut avoir lieu au sein d'une même exigence – avec le même effet de confusion dû au manque d'uniformité (s'agit-il à chaque fois d'une obligation ?) :

> Le sol *doit* connaître la configuration réelle du satellite complet par la télémesure de l'état des relais. Il *connaît* également la configuration vue par le logiciel central. [...]

« Devoir » et « pouvoir » sont régulièrement utilisés ensemble, pour indiquer qu'une action doit être rendue possible ou qu'une possibilité doit exister :

> l'opérateur *doit* pouvoir activer manuellement une tâche d'acquisition de données.

Ce qui explique la multiplication des verbes[86], comme ici (devoir + pouvoir + être + participe passé) :

86. ↑Qui n'est pas toujours justifiée, comme dans le pléonasme suivant : « il doit pouvoir être possible de choisir de récupérer les données sous forme de 5 fichiers journaliers ».

> ZOOMIC *doit* pouvoir être activé de manière automatique à partir du CECT, pour effectuer les traitements d'orbitographie.

Mais, sans surprise, « pouvoir » est aussi utilisé sans « devoir » :

> [...] Un utilisateur peut appartenir à un ou plusieurs groupes. [...]

À une seule occurrence de « devoir » peuvent correspondre plusieurs besoins, exprimés par autant d'infinitifs coordonnés :

> La fonction " Elaboration de statistiques " *doit* exécuter des analyses de tendances et éditer des rapports statistiques [...]

Enfin, comme nous l'avions dit, « ne pas devoir » peut exprimer à la fois l'interdiction et l'absence d'obligation. C'est évidemment la première interprétation qui s'impose dans le corpus :

> [...] Les plans TC ne *doivent* pas se chevaucher. [...]

> [...] L'exécution des retraitements ne *doit* pas perturber l'exécution des traitements opérationnels.

Pour l'absence d'obligation (qui pour certains correspond plus à un commentaire qu'à une exigence au sens strict), d'autres tournures sont préférées :

> [...] il n'est pas obligatoire de l'effectuer en temps réel [...]

> [...] Il n'est pas demandé de fournir des moyens de décommunation de ces paquets. [...]

> Il n'est pas nécessaire de valider la valeur d'un champ si la valeur par défaut convient.

3.5.5.3 *Synthèse*

Notre conclusion est sensiblement la même que pour les temps verbaux : toutes les exigences pourraient être rédigées avec ou sans « devoir », l'important étant surtout d'être cohérent (ce qui est loin d'être toujours le cas en l'état). Même si cela augmenterait légèrement la longueur des exigences, nous pensons que l'utilisation systématique de « devoir » serait avantageuse parce qu'elle renforcerait la valeur injonctive de l'exigence, mais aussi et surtout parce qu'elle permettrait l'identification immédiate de tous les besoins auxquels une réponse doit être apportée. La logique peut être poussée plus loin en imposant la présence d'autant de formes de « devoir » qu'il y a de besoins exprimés, ce qui implique de le répéter avant chaque infinitif (« X doit extraire et analyser les données » devenant « X doit extraire et

doit analyser les données ») pour faciliter la vérification par la suite. Cela lèverait de plus l'ambiguïté entre ce qui doit être fait (« L'emplacement de ces données de support doit être défini dans un fichier de configuration pour chaque station de travail ») et ce qui est déjà fait (« L'arborescence des espaces d'échange avec le CCC est définie dans le document DA2. »), ou entre ce qui est toléré (« La date peut être initialisée par le CMSM, mais il revient au CECT de la mettre à jour. ») et ce qui doit être possible[87] (« Cette date doit pouvoir être forcée manuellement »). Si cette règle est respectée, on peut en conclure qu'une phrase qui ne contient aucun « devoir » n'exprime pas un besoin (ce qui ne signifie pas, selon nous, qu'elle est forcément inutile, mais peut amener le rédacteur à s'interroger sur sa pertinence).

Une interdiction étant en quelque sorte un besoin, nous ne voyons pas d'objection à ce que la forme négative « ne doit pas » soit utilisée pour l'exprimer. Il faut toutefois s'assurer que ce ne soit pas le cas pour exprimer l'absence d'obligation (une règle qui semble instinctivement respectée par tous les rédacteurs, mais qui pourrait malgré tout faire l'objet d'une vérification automatique), pour laquelle une forme unique pourrait être proposée (par exemple, « Il n'est pas obligatoire de/que… »).

87. ↑Certaines phrases dont le verbe est « pouvoir » nous semblent ambiguës : « ORAMIC peut contenir une seconde méthode de restitution d'attitude à des fins d'expertise. Ses sorties sont contenues dans le rapport d'expertise. » S'agit-il de faire en sorte qu'il puisse contenir une seconde méthode (« doit pouvoir »), ou est-ce déjà le cas ?

3.5.6 *Voix passive*

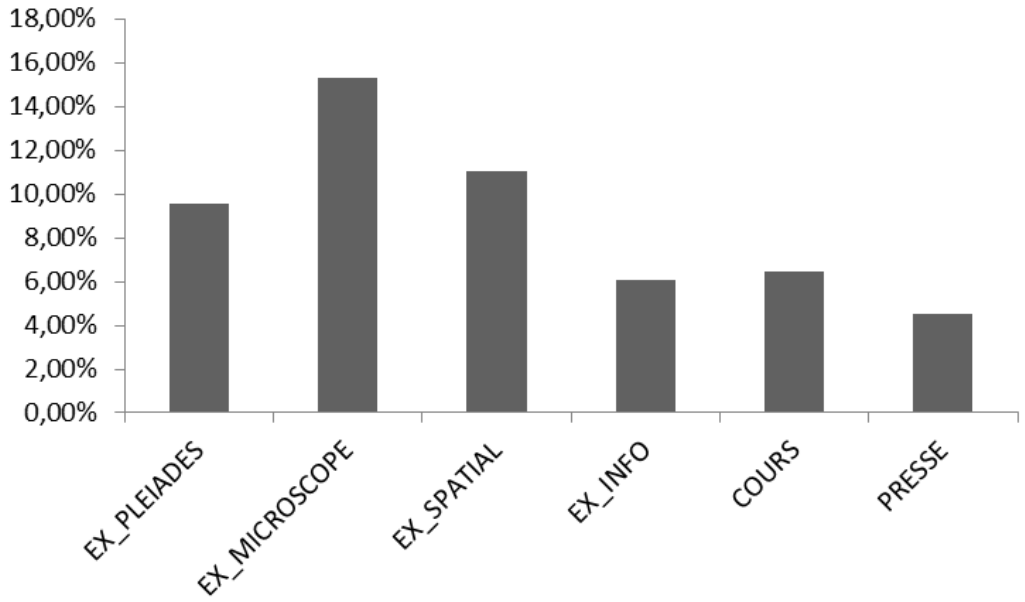
3.5.6.1 *Analyse quantitative*

Hypothèse : la fréquence relative des verbes à la voix passive est plus faible dans le corpus d'exigences que dans les autres textes.

TABLEAU 11 – Fréquence de la voix passive (par rapport au nombre total de verbes, hors participes passés)

	Voix passives
EX_PLEIADES	989 (9,56 %)
EX_MICROSCOPE	541 (15,32 %)
EX_SPATIAL	1530 (11,03 %)
EX_INFO	405 (6,09 %)
COURS	706 (6,44 %)
PRESSE	728 (4,55 %)

FIGURE 9 – Fréquence de la voix passive (par rapport au nombre total de verbes, hors participes passés)



Ces résultats (TABLEAU 11 et FIGURE 9) constituent une réelle surprise, car ils vont complètement à contre-courant des recommandations des langues contrôlées : c'est en effet dans les exigences spatiales (surtout Microscope) que l'on trouve, de loin, le plus de voix passives ; elles y sont même relativement fréquentes (un verbe sur dix environ). C'est la presse qui en fait le moins souvent usage.

Si la voix passive – pourtant quasiment unanimement considérée comme inappropriée à la rédaction technique –, est aussi souvent employée par les rédacteurs d'exigences du CNES, c'est qu'elle doit présenter un intérêt à leurs yeux. Nous tenterons de comprendre lequel et de vérifier si sa (mauvaise) réputation est justifiée.

3.5.6.2 *Analyse qualitative*

Remarquons tout d'abord que le même verbe peut être utilisé parfois à la voix active, parfois à la voix passive :

> En mode NORM_auto, le bord (LVC) gère de manière automatique les transitions GAP vers SUP et SUP vers GAP.

> [...] Les trous de mesure longs *seront gérés* manuellement par ORAMIC en mode expertise en cohérence avec le CMSM.

Le premier exemple pourrait sans problème être rédigé à la voix passive et le second exemple pourrait lui être rédigé à la voix active. On trouve donc des phrases à la voix passive dans lesquelles le complément d'agent est présent :

> L'outil proposera la valeur par défaut du gain de recalage qui devra *être validé* (et pourra *être modifiée*) par les experts avant prise en compte dans le scénario de travail .

Ces phrases sont tout aussi complètes que leur contrepartie à la voix active et ne semblent pas présenter plus de difficultés de compréhension. La principale différence est que l'agent apparaît après le verbe, et non avant. Une fois de plus, on constate que les deux versions peuvent se retrouver dans la même exigence (première phrase à la voix passive, seconde à la voix active) :

> Les données en provenance du CMTG *sont purgées* du serveur FTP par le CECT une fois traitées.
> Le CMTG se charge de purger le serveur d'éventuels plans TC non utilisés.

Dans certains cas, la voix active et la voix passive semblent interchangeables. Dans d'autres, la voix active semble s'imposer :

> Le serveur GIDE trace son activité sur 2 niveaux : [...]

La forme passive, ?*Son activité est tracée sur 2 niveaux par le serveur GIDE*, ne semble pas vraiment acceptable : à moins de chercher un autre représentant que « Le serveur GIDE » pour « son » (ce qui reviendrait à changer complètement le sens de la phrase), l'effet produit est assurément étrange.

Il y a des cas où c'est au contraire la voix passive qui est préférée. Par exemple, afin que le thème (ici, « Les plans de programmation » anaphorisé en « ils ») soit en position de sujet de la phrase :

> Les plans de programmation mis à disposition de GenTC sont conservés sur un horizon glissant de 10 jours configurable. Au-delà de ce délai, ils *sont* automatiquement *supprimés* de l'espace d'import de GenTC par le CECT. [...]

Il peut aussi simplement s'agir de topicaliser le patient, lorsque celui-ci est l'élément principal de l'exigence (ici, « Les produits Nob ») :

> Les produits Nob *sont générés* par le CECT tous les jours, sur les données de la veille de 0h00 à 23h59min59s.

La passivation permet également d'assigner à un même sujet syntaxique plusieurs actions et agents différents :

> [...] La durée des sessions *est entrée* par le CMSM et ne doit pas *être modifiée* par le CECT lors de la mise à jour automatique.

> [...] Les dates des sessions *sont imposées* par le scénario_G2 et non *remaniées* par le GTC. [...]

Si l'on réécrivait ces deux exemples à la voix active, la coordination des propositions par « et » semblerait moins justifiée et l'on serait sans doute contraint d'en faire deux phrases distinctes.

La voix passive peut aussi aider à mettre subtilement en relief l'agent, en particulier lorsqu'il est différent de l'agent de la phrase précédente (« par le CMSM » différent de « par le CECT ») :

> Notons que le scénario de travail *est modifié* et *mis* à disposition du CMSM par le CECT, ce pourquoi il apparait dans ce tableau. Mais la première version *est produite* par le CMSM, d'où le nom avec " CMSM " comme producteur de la donnée. [...]

Pour obtenir le même effet à la voix active, on serait sans doute tenté d'utiliser une clivée (« Mais c'est le CMSM qui produit la première version... »).

Bien d'autres raisons plus ou moins faciles à identifier et à justifier sont encore possibles. Par exemple, ici, on peut émettre l'hypothèse que l'intérêt de la voix passive est qu'elle permet de créer – en apparence en tout cas – un parallélisme avec la proposition qui précède :

> Les valeurs des arguments saisis par le CMSM sont les " libellés pour fiches session ", les valeurs brutes correspondantes *sont déduites* par le CECT au moment de la création (encodage) de la TC.

où « les valeurs brutes correspondantes sont… » ressemble beaucoup plus à « Les valeurs des arguments saisis par le CMSM sont… » que « Le CECT déduit les valeurs brutes correspondantes… ». Au-delà des exemples individuels, il faut surtout remarquer que la voix passive, pourtant réputée plus complexe et moins prototypique que la voix active, est utilisée spontanément à de nombreuses reprises par les rédacteurs là où la voix active serait pourtant possible, ce qui n'est de toute évidence pas le fait du hasard. Dans certains cas, ce choix peut se justifier.

Potentiellement plus problématiques, en revanche, sont les voix passives dans lesquelles le complément d'agent est omis. Il faut d'abord arriver à distinguer les véritables voix passives des structures « être + adjectif » :

> […] La batterie sera rechargée. […]

où « rechargée » peut être vu soit comme un adjectif précisant l'état de la batterie, soit comme un participe passé composant une voix passive.

Le complément d'agent peut parfois être facilement déduit grâce au cotexte :

> Le CECT doit mettre à jour le numéro des séquences à l'état AC, AE. Les numéros des séquences à l'état " E ", " EI ", " EC " ou " C " ne doivent pas *être modifiés*. […]

Ici, « (par) le CECT » est sous-entendu et sa répétition n'aurait sans doute aucun intérêt. C'est encore plus clairement le cas lorsque le complément d'agent apparaît dans une voix passive précédente :

> Les produits Noa *sont générés* automatiquement par le CECT suite à une demande du CMSM, sur un horizon temporel défini par une date de début et une date de fin (dates UTC). […] A l'issue de cet horizon temporel, les produits Noa ne *sont* plus *générés*.

Ajoutons que ne pas préciser l'agent permet de gagner en généralité : « les produits Noa ne sont plus générés » signifie qu'ils ne sont plus générés ni par le CECT, ni par rien d'autre ; de même pour « ne doivent pas être modifiés » : ils ne doivent être modifiés ni par le CECT, ni par quoi que ce soit d'autre qui serait en mesure de le faire.

Il est aussi possible que connaître l'agent ne soit manifestement pas nécessaire :

> […] La contrainte orbitale est mise à jour au CECT même si le champ *avait* déjà *été rempli*.

Dans cet exemple, peu importe par qui ou par quoi le champ a pu être rempli, puisque c'est quoi qu'il en soit sans conséquence sur l'action à effectuer. Toutefois, dans la plupart des cas, l'agent reste implicite alors qu'il est pourtant nécessaire de savoir qui effectue l'action demandée. Seule la connaissance du contexte (fonctionnement du système) ou du reste de la spécification permet de pallier ce manque :

> Ce fichier *est généré* et *mis* à jour : [...]

> Les répertoires d'IMIS devront *être purgés* automatiquement avec une fréquence configurable.

> [...] Dans ce cas, un warning doit *être levé*.

Dans les trois exemples ci-dessus, il est important de savoir ce qui génère et met à jour le fichier, purge les répertoires ou lève le warning, mais aucun élément textuel ne le précise. Il faut espérer que cette information soit suffisamment évidente pour l'expert qui lira l'exigence, sans quoi elle doit être considérée comme incomplète.

Enfin, l'omission de l'agent peut être réellement dommageable lorsqu'aucun acteur ou élément individuel ne peut être identifié comme responsable d'une tâche :

> Un disfonctionnement du MIGS doit *être rétabli* en moins de 12 heures en mode exceptionnel

Nous voyons mal comment il est possible de s'assurer du respect de cette contrainte si rien ni personne n'est chargé d'y veiller : dans le cas où un dysfonctionnement se produit, qui interviendra ? Si le rédacteur n'a pas précisé l'agent, c'est peut-être parce que lui-même n'est pas parvenu à le définir précisément – ce qui est peut-être un signe que l'exigence est trop générale.

3.5.6.3 *Synthèse*

Comme nous l'avons vu, les voix passives sont assez fréquentes dans le corpus d'exigences du CNES. Les interdire toutes nécessiterait de la part des rédacteurs un effort non négligeable pour modifier leurs habitudes, auquel s'ajoute la difficulté indéniable de distinguer les voix passives (habituellement interdites par les langues contrôlées) des structures « être + participe passé employé comme adjectif » (habituellement autorisées).

À supposer qu'il soit envisageable, cet effort en vaudrait-il réellement la peine ? D'un côté, les voix passives pour lesquelles le complément d'agent est donné ne semblent pas poser de problème. Certes, certaines voix passives pourraient sans mal être transformées en voix actives (« La décision de non conservation des données devra être justifiée et approuvée par le CNES. ») et quelques-unes gagneraient sans doute un peu en clarté si elles l'étaient, mais on pourra objecter que d'autres

« sonnent » tout simplement mieux que leur contrepartie active (« Toute intégration d'une nouvelle version d'un traitement qui ne comporte pas d'évolution d'interface peut être réalisé simplement par un opérateur du CECT au moyen d'une IHM [...] »). Dans tous les cas, ces voix passives ne sont en rien moins complètes que les versions actives et l'intérêt de les interdire systématiquement est pour le moins discutable.

D'un autre côté, on trouve également de nombreuses voix passives pour lesquelles l'agent n'est pas précisé, omettant ainsi une information en théorie essentielle dans la phrase. Ces cas sont-ils vraiment problématiques ? Le cotexte immédiat (au sein de l'exigence) rend parfois l'agent évident, même pour un non expert. Lorsque ce n'est pas le cas, c'est généralement le contexte qui s'en charge, mais cette fois à condition de disposer d'une connaissance suffisante du projet. Ainsi, pour évaluer plus objectivement le risque lié aux voix passives dans les exigences, Krisch et Houdek (2015) ont demandé à un expert de juger si les quelque 711 voix passives (dont seulement 126 pour lesquelles le complément d'agent était donné) présentes dans un corpus d'exigences (en anglais) du domaine automobile lui semblaient – ou non – poser problème. Considérées isolément, 236 (33,20 %) d'entre elles sont qualifiées de problématiques par l'évaluateur. Mais lorsque ce dernier a accès à l'ensemble de la spécification dont les phrases font partie, ce nombre chute à deux (0,28 % à peine). Autrement dit, le nombre d'exigences pour lesquelles il n'est pas possible pour un expert, en disposant du contexte constitué par la spécification, de déterminer de façon fiable (sans ambiguïté possible) l'agent de la voix passive est en réalité très faible. Si le risque existe bel et bien (mais uniquement dans les voix passives sans complément d'agent), il semble donc largement surévalué. La nécessité d'interdire les voix passives devrait sans doute être relativisée, le complément d'agent n'étant dans bon nombre de cas pas strictement nécessaire.

Notre conclusion est la suivante : nous pensons que les voix passives avec complément d'agent sont sans risque (ce qui n'interdit évidemment pas aux rédacteurs qui le souhaitent de les réécrire à la voix active) et que les voix passives sans complément d'agent peuvent être, mais ne sont pas nécessairement problématiques. Par conséquent, la meilleure recommandation à faire aux rédacteurs serait, pour nous, de s'assurer, lorsqu'ils utilisent une voix passive, que l'agent puisse être déterminé sans problème par le lecteur – soit parce qu'il est explicité par le complément d'agent, soit parce que le co(n)texte le rend suffisamment évident. Pour les y aider, le plus simple serait de leur signaler automatiquement toutes les phrases à la voix passive sans complément d'agent et de les laisser décider si celui-ci doit ou non être rajouté.

Enfin, il est probable que dans quelques rares cas, ne pas préciser l'agent soit intentionnel – parce que le rédacteur ne peut ou ne veut pas attribuer une responsabilité à quelqu'un ou quelque chose en particulier. Le problème est alors plus profond (et ne relève pas de la formulation) : pour pouvoir être précisé, l'agent doit tout d'abord être identifié clairement. Sans cela, réécrire la phrase à la voix passive ne serait de toute façon pas possible... à moins d'utiliser le pronom « on ».

3.5.7 *Pronom « on »*

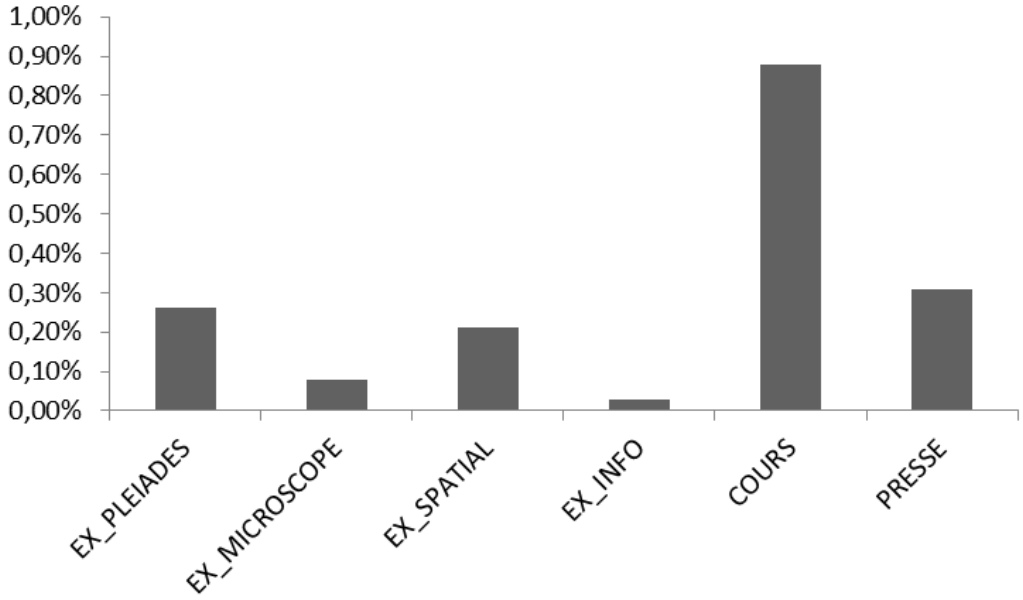
3.5.7.1 *Analyse quantitative*

Hypothèse : la fréquence relative du pronom « on » est plus grande dans le corpus d'exigences que dans les autres textes.

TABLEAU 12 – Fréquence du pronom « on » (par rapport au nombre total de mots)

	Pronom « on »
EX_PLEIADES	310 (0,26 %)
EX_MICROSCOPE	35 (0,08 %)
EX_SPATIAL	345 (0,21 %)
EX_INFO	14 (0,03 %)
COURS	1061 (0,88 %)
PRESSE	502 (0,31 %)

FIGURE 10 – Fréquence du pronom « on » (par rapport au nombre total de mots)



Contrairement à ce que nous pensions initialement, le pronom « on » est moins utilisé dans les exigences que dans les deux autres types de textes (il est même quasiment absent des exigences informatiques et très rare dans Microscope). Cela

ne nous surprend finalement qu'à moitié, car nous l'y supposions plus fréquent pour compenser l'absence de voix passives ; or, comme nous l'avons vu juste avant, elles restent très nombreuses dans les exigences. Remarquons d'ailleurs qu'elles étaient proportionnellement plus nombreuses dans Microscope que dans Pléiades et que la tendance s'inverse ici, ce qui peut peut-être être vu comme un indice allant dans notre sens. Il se pourrait aussi que le pronom « on » ait souffert de sa mauvaise réputation et que les rédacteurs aient cherché à l'éviter, plus que la voix passive. Il est anormalement fréquent dans le cours (souvent par désir de rester général : « on a intérêt… ») et n'est pas rare dans la presse, quand bien même il est souvent qualifié d'oral et de familier.

3.5.7.2 *Analyse qualitative*

Notons tout d'abord que (sans surprise) les pronoms personnels de première et de deuxième personnes sont pratiquement absents des exigences, qui se veulent aussi impersonnelles que possible. On trouve de très rares occurrences de « nous », qui font figure d'exception (et qui pour deux d'entre elles apparaissent conjointement avec « on ») :

> […] SI les tuyères sont redondées *nous* utiliserons l'indice a pour la tuyère nominale et b pour la tuyère redondante. […]

> […] - Si *on* est au niveau ICR, *nous* devons pouvoir choisir de masquer que les ICR, ou les ICR et PR associées, ou les ICR+PR+mailles, ou les ICR+PR+mailles+segments.
> - Si *on* est au niveau PR, *nous* devons pouvoir choisir de masquer que les PR, ou les PR+mailles, ou les PR+mailles+segments, etc

Il arrive cependant que le pronom « on » joue ce rôle. C'est le cas lorsqu'il renvoie implicitement au(x) rédacteur(s) (1re personne) :

> *On* ne listera ici que les TC associées aux familles F_EMETTEUR et F_MANOEUVRES. […]

> […] *On* recommande donc une durée d'observation de l'ordre de 25 à 100 sec, incluse dans le mode MCO considéré. […]

ou bien au(x) lecteur(s) (2e personne) :

> […] En ce qui concerne les notations, *on* se réfèrera au paragraphe 6.2.6.

> […] *On* notera que les FEE en BDS se présente sous forme identité (Id + fee) alors que les calculs du chapitre 6.1 les traitent en FEE pur (proches de zéro). […]

Dans ce dernier exemple, bien que la proposition principale ne soit pas indispensable, elle vise à attirer l'attention du lecteur sur une contrainte particulière. À chaque fois, l'objectif est de dissimuler les déictiques (FLUDERNIK, 1991). Très souvent, le pronom « on » est utilisé pour introduire une donnée arbitraire ou conventionnelle sans paraître trop autoritaire :

> [...] *On* considère que les autres TC sont associées à la famille F_AUTRES. [...]

> [...] *On* prendra comme valeur d'origine du syndrome la valeur oXFFFF (calcul syndrome du premier octet). [...]

> [...] *On* fixe t_1 = to - 1 seconde (avec to date du premier biais de vitesse). [...]

On pourrait avancer que dans ces trois exemples, il aurait été possible d'éviter « on » ; par exemple, dans la première phrase, la proposition principale « On considère que » pourrait être retirée ou remplacée (par exemple, par « Par convention »).

Dans certains cas, on peut considérer que les phrases dont « on » est le sujet relèvent de la fonction métalinguistique de Jakobson, comme dans les énoncés définitoires (REBEYROLLE et TANGUY, 2000) suivants :

> *On* appelle traitement temps réel tout traitement s'effectuant en cours de passage en visibilité, à partir de la télémesure décommutée en temps réel par le CCC. [...]

> Il n'existera aucune valeur numérique codée " en dur " dans GOTLIB. Par codé en dur *on* entend une valeur numérique qui est directement exprimée dans les opérations du code [...]

Dans certains de ses usages, « on », indéfini, ne semble pas renvoyer à quelqu'un ou quelque chose en particulier :

> L'algorithme marche avec autant de points et donc de manœuvres que l'*on* souhaite. [...]

> Les natures d'évènements que l'*on* trouve dans ce fichier sont : [...]

Il est possible aussi que « on » soit employé justement en raison de sa généralité :

> *On* ne peut écrire et lire simultanément le même fichier, y compris un fichier DAS.

> [...] En conséquence *on* ne mixera jamais des clés C et M en parallèle dans une programmation.

Dans les deux exemples ci-dessus, « on » peut signifier « rien ni personne » – l'interdiction apparaît alors comme absolue.

Enfin, on trouve – comme on s'y attendait – de nombreux exemples où « on » est (trop) vague et s'apparente à « quelqu'un » ou « quelque chose » :

> Afin de corriger a bord les erreurs le long de la trace entre le modèle d'orbite sol et la position orbitale fournie par DORIS, *on* enverra un ordre MDP de recalage. [...]

> [...] Pour cela *on* contrôlera les bornes MIN et MAX de la FT BDS (de type identité) associée à ce paramètre.

> [...] *On* vérifiera que les tables sont reçues en intégralité. [...]

À nouveau, seul le contexte permet à un expert de savoir qui ou quoi enverra, contrôlera ou vérifiera. Ces phrases à la voix active paraissent donc tout aussi incomplètes que les voix passives sans agent et la problématique reste la même. Il n'est d'ailleurs pas étonnant que l'on puisse souvent passer d'une formulation à l'autre (la première phrase, par exemple, aurait pu être « ... un ordre MDP de recalage sera envoyé »). Interdire (ou décourager) les voix passives sans interdire le pronom « on », c'est donc prendre le risque que ce dernier soit utilisé pour contourner l'interdiction, sans régler le problème. À vrai dire, « on » nous semble encore plus problématique, car occuper la fonction de sujet lui permet d'omettre non seulement l'agent d'un procès, mais aussi son bénéficiaire :

> *On* devra disposer de ces résultats en ligne depuis le tir.

Dans cette exigence, on ne sait pas à disposition de qui (ou de quoi) les résultats devront être mis.

3.5.7.3 *Synthèse*

Comme nous nous y attendions, la problématique liée au pronom « on » et celle des voix passives (sans complément d'agent) sont semblables : l'un comme l'autre, en omettant l'agent, font courir le risque que celui-ci ne soit pas (clairement) identifiable et donc que l'exigence soit incomplète. Par conséquent, dans une langue contrôlée basée sur le français, pour être efficace, une recommandation concernant la voix passive devrait nécessairement être accompagnée d'une recommandation similaire concernant « on », et vice-versa.

Une fois encore, on constate que plusieurs usages de « on » sont sans risque, tandis que d'autres (y compris quelques-uns qui ne concernaient pas la voix passive) sont potentiellement porteurs de risque, car ils supposent une connaissance du contexte partagée entre rédacteur et lecteur. Notre conclusion est sensiblement la même que pour la voix passive, même si nous pensons que l'utilisation de « on »

devrait être aussi limitée que possible, parce qu'il est rarement utile. Par conséquent, la recommandation pourrait inciter les rédacteurs à s'en passer chaque fois que c'est possible, et à vérifier que l'agent est identifiable dans tous les cas. L'avantage est qu'« on » est très facile à détecter et qu'il est déjà beaucoup moins fréquent que la voix passive dans les exigences du CNES.

3.5.8 *Nominalisations déverbales*

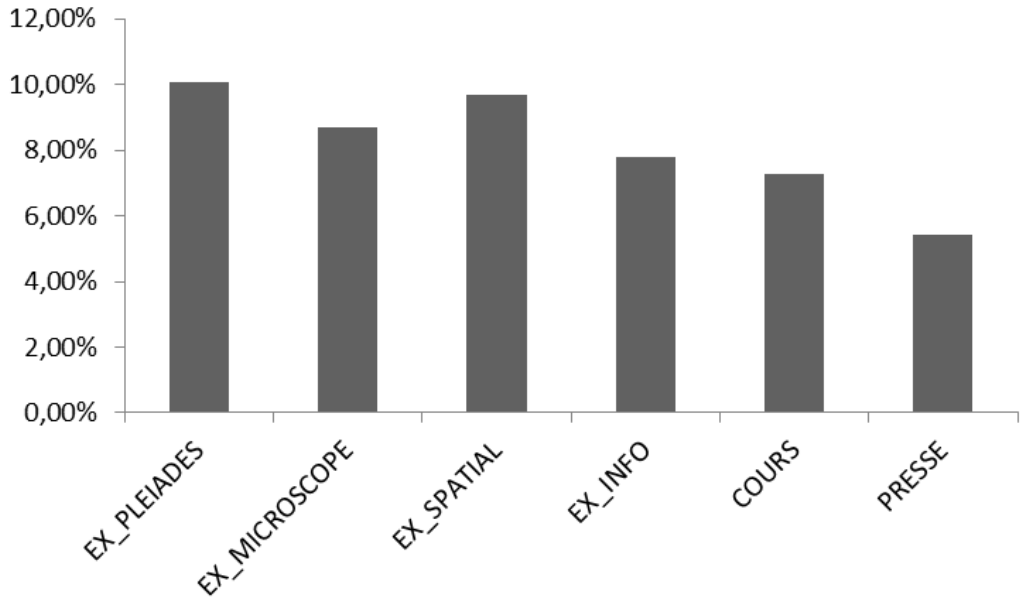
3.5.8.1 *Analyse quantitative*

Hypothèse : la fréquence relative des nominalisations déverbales est plus grande dans le corpus d'exigences que dans les autres textes.

TABLEAU 13 – Fréquence des nominalisations déverbales (par rapport au nombre total de mots)

	Nominalisations déverbales
EX_PLEIADES	11863 (10,05 %)
EX_MICROSCOPE	3771 (8,70 %)
EX_SPATIAL	15634 (9,69 %)
EX_INFO	4337 (7,79 %)
COURS	8715 (7,26 %)
PRESSE	8760 (5,43 %)

FIGURE 11 – Fréquence des nominalisations déverbales (par rapport au nombre total de mots)



Il faut tout d'abord prendre garde au fait que tous les chiffres (TABLEAU 13 et FIGURE 11) sont surestimés, parce que toutes les occurrences de noms pouvant être des nominalisations déverbales ont été comptées, même lorsque ceux-ci sont polysémiques et ne désignent généralement pas un procès (par exemple, « mission » ou « station », fréquents dans les exigences, mais aussi « entreprise » ou « destination », que l'on retrouve davantage dans la presse).

Cela précisé, on ne s'étonnera pas de constater que les nominalisations déverbales sont effectivement plus nombreuses dans les exigences (là aussi, les deux sous-corpus sont comparables), suivies de près du cours (lui aussi technique) et enfin de la presse généraliste. Il nous reste à voir comment ces nominalisations peuvent être utilisées pour remplacer des verbes.

3.5.8.2 *Analyse qualitative*

En raison du grand nombre de résultats et surtout du bruit important, il n'est pas possible de fournir une revue exhaustive des nominalisations déverbales dans le corpus. En effet, beaucoup de ces nominalisations sont polysémiques et n'expriment pas toujours un procès :

> [...] La version mise à jour par le CECT après GPOM, utilisée pour générer la *programmation*.

> Cette TC permet de sélectionner la *configuration* des tranches de COME en définissant les tranches utilisées et les convertisseurs et banques mémoires utilisés par tranche de COME. [...]

Nous nous limiterons à deux brèves observations. La première est qu'une nominalisation déverbale et le verbe dont elle provient (ici, « générer ») peuvent tous deux être utilisés, selon les besoins :

> Les produits Noc_GNSS sont générés par le CECT de façon quotidienne et itérative, pour toutes les sessions à traiter.

> La *génération* des produits répondant à une session de longueur supérieure à la journée se fait par itération successive, chaque jour. [...]

On constate parfois l'utilisation d'une nominalisation déverbale avec un verbe support, là où le verbe d'origine restait possible :

> [...] Cette fonction assure la *gestion* de ce fichier TB/TU de référence.

> Cette TC permet de mettre ON la fonction logicielle qui gère l'ensemble des FOG utilisés pour l'estimation d'attitude. [...]

Ici, les expressions « gérer » et « assurer la gestion » semblent en concurrence.

> [...]
> * Le CCC effectuera l'*extraction* des paquets TM,
> * le CCC effectuera l'*archivage* de la totalité des paquets TM (plateforme et scientifique) dans la base de données du G3, [...]

Dans cet exemple, par deux fois, l'expression « effectuer le + nominalisation déverbale » est préférée aux verbes correspondants (« extraire » et « archiver »).

La deuxième remarque, qui explique pourquoi nous nous y intéressons, est que les nominalisations permettent l'omission à la fois de l'agent et du patient :

> Pour les sessions de *mesure* ou d'*étalonnage*, les informations de position et vitesse du satellite seront transmises au SST pour la correction de l'aberration relativiste. [...]

Dans cet exemple, il n'est précisé ni ce qui est mesuré ou étalonné, ni par quoi ou par qui. Assez souvent, le patient est précisé, mais pas l'agent (comme on le voit aussi pour « correction de l'aberration relativiste ») :

> Le générateur de TCH listera les OBCD associé au PROCESS_ID du destinataire de manière à faciliter la *sélection* des PARAM_ID. [...]

> En cas de passage en mode repli ou survie plate-forme, le bord doit :
> - inhiber l'*exécution* des TCs différées,
> - bloquer la *réception* de TC nominales (par exemple envoyées au moment du repli plate-forme.

Tandis que plus rarement, les deux sont explicités :

> [...] La *programmation* du recalage temps bord par le CECT doit pouvoir être débrayée.

> En mode normal, le délai maximum entre la *réception* des commandes par les électroniques du CGPS et l'application de 63% des poussées commandées doit être inférieur à 250 ms [...]

3.5.8.3 *Synthèse*

Puisqu'elles offrent elles aussi la possibilité d'omettre l'agent (et même le patient), les nominalisations déverbales présentent le même risque d'incomplétude que la voix passive (ou le pronom « on »)[88]. Le même constat s'impose : c'est le plus souvent le contexte ou le cotexte qui permettent de restituer ce qui est laissé implicite par le rédacteur. Cela justifierait donc sans doute que l'on s'y intéresse

88. ↑D'autres structures encore mériteraient que l'on s'y intéresse, notamment certaines tournures impersonnelles : « Il doit être possible [pour qui/quoi ?] de... ».

de plus près et qu'une recommandation en précise l'utilisation. Néanmoins, nous craignons que le grand nombre de nominalisations déverbales présentes dans les exigences et surtout le nombre élevé de faux positifs qui seraient retournés par un outil de vérification automatique constituent une perte de temps rédhibitoire pour les rédacteurs. Pour cette raison, nous préférons nous en passer pour le moment, en attendant que ces nominalisations ne fassent l'objet d'une étude beaucoup plus fine qui permettrait éventuellement d'améliorer la précision[89] (et le rappel, certaines nominalisations techniques étant absentes de Verbaction).

3.6 CONCLUSION

L'objectif de ce chapitre était double. D'une part, nous souhaitions apporter un argument quantitatif en faveur de notre hypothèse concernant l'existence d'un genre propre aux exigences, qui justifie notre démarche d'analyse sur corpus. À cette fin, nous avons comparé les fréquences d'apparition relatives de plusieurs phénomènes dans nos deux sous-corpus et dans d'autres textes, qu'ils soient proches par certains aspects (exigences de projets informatiques, cours sur le domaine spatial) ou au contraire supposés très différents (articles de presse généraliste). Les résultats obtenus semblent certainement aller dans notre sens, puisque l'on constate des différences notables entre les exigences spatiales et les autres genres. Bien sûr, d'un sous-corpus à l'autre (Pléiades et Microscope), les proportions ne sont pas exactement identiques (avec même dans quelques cas des disparités importantes, qui mériteraient que l'on s'y intéresse davantage), mais, à chaque fois, les tendances (plus ou moins marquées) sont similaires : soit un phénomène y est sous-représenté (pronoms, conjonctions, phrases longues, etc.), soit il y est surreprésenté (modal « devoir », voix passives, nominalisations, etc.). Quant aux exigences informatiques, que nous avons préféré traiter à part à ce stade, elles leur ressemblent sur plusieurs points, mais s'en éloignent nettement sur d'autres – ce qui pourrait indiquer que certains traits seraient d'une certaine façon inhérents à cette pratique, tandis que d'autres seraient appris, par habitude, par les rédacteurs appartenant à une même communauté professionnelle.

La proximité entre Microscope et Pléiades peut aussi se constater graphiquement sur la FIGURE 12, issue d'une analyse factorielle des correspondances basée sur quelques étiquettes morphosyntaxiques (conjonctions, pronoms et plusieurs tiroirs verbaux[90]) obtenues grâce à TreeTagger et TXM. Ce sont les spécificités de ce genre que nous nous efforçons d'identifier et d'exploiter (dans ce chapitre et dans le suivant).

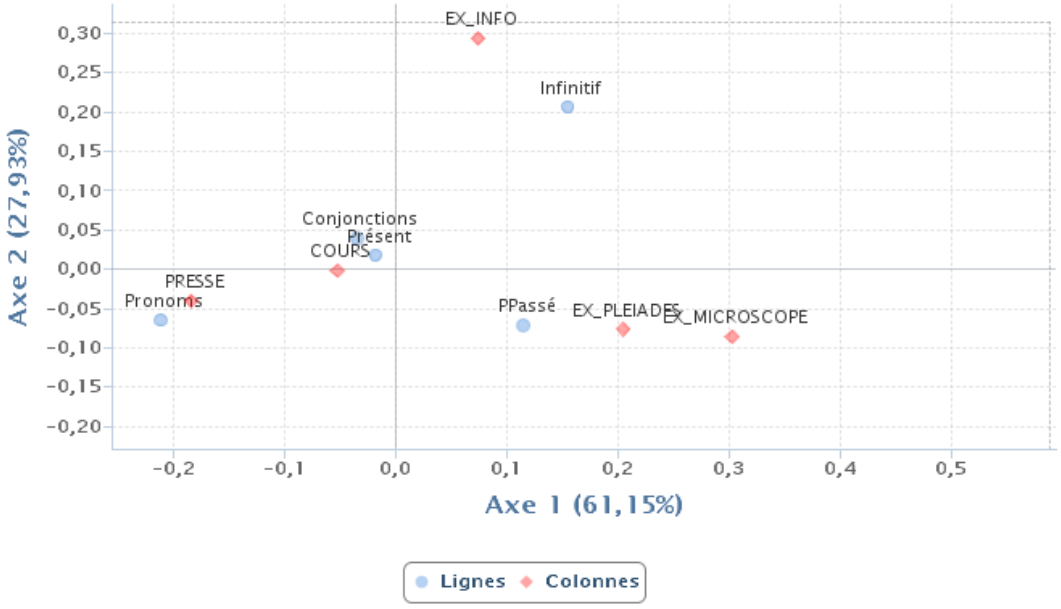
D'autre part, nous avons voulu montrer en quoi une approche *corpus-based* pouvait être utile pour proposer une langue contrôlée en phase avec les usages réels. Nous sommes parti de règles existantes (parfois en les adaptant ou en les extrapolant) et les avons confrontées à la réalité du corpus pour évaluer leur adéquation

89. ↑Des résultats prometteurs ont déjà été obtenus pour l'anglais (LANDHAUSSER et al., 2015).
90. ↑Le futur et le conditionnel n'apparaissent sur cette vue, car ils sont trop éloignés de l'origine.

avec les pratiques des ingénieurs du CNES et leur intérêt réel. Nous avons pu montrer, grâce à des extraits d'exigences authentiques, que plusieurs d'entre elles nécessitaient d'être reformulées et avons proposé des pistes concrètes pour les améliorer. Nous ne prétendons en aucun cas que les suggestions de recommandations que nous avons pu faire sont parfaites – il faudra, de toute façon, qu'elles soient d'abord « éprouvées » en conditions réelles par les rédacteurs, pour que l'on puisse ensuite affirmer qu'elles constituent une solution efficace, le cas échéant, ou pour qu'elles puissent être rectifiées si nécessaire. Mais nous pensons tout au moins avoir démontré la faisabilité et l'intérêt de notre démarche pour proposer des règles mieux adaptées, exemptes d'aberrations majeures et moins arbitraires. Pour chaque recommandation, nous avons cherché un compromis qui la rende efficace – et donc contraignante – tout en s'assurant qu'elle reste suffisamment simple à appréhender et à respecter pour les rédacteurs, qui seront idéalement assistés d'un outil de vérification semi-automatique.

Dans le chapitre suivant, nous verrons comment une approche *corpus-driven* peut enrichir cette réflexion en fournissant des résultats complémentaires.

FIGURE 12 – Analyse factorielle des correspondances (lignes = POS tags, colonnes = textes)



4

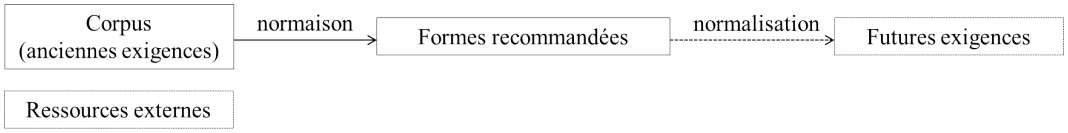
ANALYSE DU CORPUS : APPROCHE CORPUS-DRIVEN

4.1 INTRODUCTION

Dans le chapitre précédent, nous avons voulu montrer comment une approche *corpus-based* permettait de proposer des règles de rédaction visant à contraindre la rédaction en langue naturelle, dans le cadre d'une langue contrôlée basée sur des usages réels. L'inconvénient majeur de cette approche est qu'elle consiste à vérifier des hypothèses formulées *a priori*, inspirées en grande partie de langues contrôlées existantes et dans une moindre mesure de notre connaissance (forcément limitée) du corpus et plus largement de notre connaissance du fonctionnement de la langue. Par conséquent, des propriétés et des structures typiques du corpus et par extension de la grammaire du genre des exigences risquent d'être totalement occultées, pour peu qu'elles ne soient pas envisagées par les hypothèses.

Pour pallier ce problème, nous proposons d'enrichir nos observations de résultats obtenus grâce à une approche *corpus-driven* (voir section 2.2.2), qui ne requiert pas de définir en amont des hypothèses précises (mais ne dispense évidemment pas d'une réflexion sur la façon d'interpréter et d'exploiter les résultats). Au moyen d'outils statistiques (et éventuellement de ressources externes), nous espérons maintenant faire émerger les régularités du corpus et nous en servir ensuite pour suggérer aux rédacteurs des formes langagières recommandées (toujours dans l'optique de proposer une normalisation basée sur la normaison), en particulier sous la forme de *boilerplates* (voir section 2.1.3 et *infra*). L'approche générale est illustrée par la FIGURE 13.

FIGURE 13 – Approche *corpus-driven* pour améliorer la rédaction des exigences



À ce principe général simple peuvent bien entendu correspondre des méthodes multiples, qui sont tributaires des outils et ressources dont on dispose, mais également de la façon dont on souhaite les mettre à profit. Pour en montrer l'intérêt, nous présenterons deux études aux objectifs similaires, mais à la méthodologie légèrement différente. La première (WARNIER et CONDAMINES, 2017b) est la plus simple et la plus intuitive puisqu'elle ne nécessite qu'un seul outil, permettant d'extraire des motifs fréquents du corpus, et un texte de comparaison pour filtrer ces derniers. La seconde (WARNIER et CONDAMINES, 2015), un peu plus complexe, emprunte à la démarche terminologique puisqu'elle permet dans un pre-

mier temps d'extraire les lexèmes représentatifs du genre, qui pourront éventuellement eux-mêmes être utilisés comme valeurs dans les *boilerplates* et qui serviront dans un deuxième temps à filtrer les motifs récurrents.

4.2 PREMIÈRE ÉTUDE

4.2.1 *Méthodologie*

SDMC (Sequential Data Mining under Constraints) (QUINIOU et al., 2012b ; BÉCHET et al., 2013) est un outil en ligne [1] gratuit qui permet l'extraction automatique de motifs séquentiels (définis par ses auteurs comme des suites ordonnées d'*itemsets* – eux-mêmes composés de plusieurs informations sur le mot, telles que son lemme ou sa catégorie grammaticale – ou simplement d'*items*) dans un texte donné. Son interface simple permet son utilisation à la fois par des linguistes (avec la possibilité de revenir au corpus, grâce à un concordancier intégré) et par des spécialistes de la fouille de données [2]. Comme son nom l'indique, SDMC permet l'application d'un certain nombre de contraintes (taille minimale et maximale du motif, présence d'une ou plusieurs catégories grammaticales, etc.) visant à filtrer les motifs extraits. Pour des raisons déjà évoquées de confidentialité de notre corpus, nous n'avons hélas pas pu nous servir de l'interface web de SDMC. En revanche, nous avons pu en utiliser une version hors ligne (qui doit être lancée en ligne de commande), qui nous a été aimablement fournie par ses concepteurs. Bien que celle-ci n'implémente pas l'intégralité des fonctions de son équivalent en ligne [3] (par exemple, la version en ligne permet d'ajouter des contraintes symboliques et syntaxiques, telle que « le motif doit contenir un verbe » (SAHRAOUI et al., 2017), ce qui n'est pas possible dans la version hors ligne), son exécution requiert plusieurs paramètres (langue, type de motif, support minimal, etc.) qui permettent de configurer facilement l'extraction en fonction des besoins de l'analyse souhaitée.

Dans notre corpus d'exigences, nous avons tout d'abord remplacé automatiquement toutes les entités nommées par un code à chaque fois identique, ce qui permet de les traiter comme un même élément lors de l'extraction des motifs [4] et réduit donc partiellement l'hétérogénéité des données.

Nous avons ensuite procédé à l'extraction des motifs indépendamment sur chacun des deux sous-corpus (Pléiades et Microscope). Dans les deux cas, SDMC était

1. ↑`https://sdmc.greyc.fr/`
2. ↑L'application des techniques de *pattern mining* ne se limite pas aux données textuelles : elle trouve également un intérêt certain dans d'autres domaines, notamment la recherche sur les interactions entre gènes (CELLIER et al., 2015).
3. ↑Et qu'elle semble souffrir de certaines limitations ; ainsi, les formes élidées (« n' », « l' », etc.) sont tout simplement ignorées lors de l'extraction des motifs.
4. ↑Ainsi, par exemple, « ORAMIC doit fournir pour chaque session… » et « GAZMIC doit fournir pour chaque session… » deviennent tous deux « NAM doit fournir pour chaque session… ».

configuré pour rechercher des motifs fréquents[5] composés d'au moins deux *items* (sans limite de longueur) et dont la fréquence minimale est de deux, sur la base des formes lemmatisées fournies par TreeTagger[6].

Nonobstant leurs tailles inégales et après plusieurs expérimentations, nous avons empiriquement fixé le seuil de fréquence minimale à deux dans chacun des sous-corpus car, comme le notent Sinclair et Tognini-Bonelli (tous deux cités par Biber, 2009 : 280), il s'agit du nombre minimal d'occurrences pour pouvoir parler de « récurrence » :

> "The distinction between 1 and more than one is also of great importance… [because even two unconnected tokens constitute] the recurrence of a linguistic event […]" (Sinclair 2001: 343-4)

> "It is therefore appropriate to set up as the minimum sufficient condition for a pattern of occurrence to merit a place in the description of the language, that it occurs at least twice, and the occurrences appear to be independent of each other" (Tognini-Bonelli 2001: 89)

Par ailleurs, vu le nombre relativement faible de mots dans chacun des deux sous-corpus (en particulier Microscope), un seuil plus élevé réduirait considérablement le nombre de motifs extraits[7], au risque sans doute d'ignorer des séquences intéressantes. C'est pour la même raison que nous avons préféré travailler sur les lemmes, plutôt que sur les formes fléchies trouvées dans le corpus, et ce même si, ce faisant, nous nous écartons, selon Biber (2009 : 281), d'une approche *corpus-driven* « radicale ». Néanmoins, nous respectons sa vision d'une telle approche « stricte » en ne tenant pas compte du statut grammatical ou syntaxique des mots. Enfin, nous avons préféré ne pas autoriser de *gaps* (« trous ») au sein des motifs, étant donné que ces trous peuvent être remplis par des éléments (pas toujours intéressants) de natures très différentes et que ceux-ci ne sont pas clairement indiqués dans les résultats retournés par SDMC, ce qui rend leur exploitation relativement compliquée. Nos motifs d'*items* ne sont composés que d'éléments fixes (à l'excep-

5. ↑Un motif est dit *fermé* s'il ne peut pas être inclus dans un autre motif (plus long) dont la fréquence d'apparition dans le corpus est identique. (Par exemple, à fréquences égales, « NAM doit fournir pour chaque session » peut être un motif fermé, mais « doit fournir » et « pour chaque session » n'en sont pas.) Cela permet d'éliminer les redondances – et donc de limiter le nombre de motifs. Les motifs *maximaux* n'incluent pas les séquences de longueur inférieure, même si leur fréquence est supérieure. (Par exemple, si « NAM doit fournir pour chaque session » et « NAM doit fournir pour chaque mesure » apparaissent tous deux n fois, ils peuvent être des motifs maximaux ; en revanche, « NAM doit fournir pour chaque » – qui apparait donc quant à lui $2n$ fois – n'en est pas un ; il y a donc potentiellement une perte d'information dans les résultats fournis.) Ces restrictions ne s'appliquent pas dans le cas des motifs dits *fréquents*, et la liste des résultats est donc généralement beaucoup plus longue. Nous les avons néanmoins préférés, car ils sont plus faciles à filtrer et à traiter par la suite.

6. ↑Le format utilisé par Talismane n'est pas supporté par SDMC (qui, sur les fichiers textes bruts, utilise par défaut les résultats de TreeTagger, mais qui accepte aussi les étiquettes du logiciel propriétaire Cordial).

7. ↑Dans le seul cas de Pléiades, fixer le seuil minimal à 3 (au lieu de 2) divise le nombre de résultats obtenus par trois.

tion des entités nommées).

Ces extractions nous ont permis d'obtenir deux listes de motifs fréquents, contenant respectivement 99 349 (Pléiades) et 63 072 (Microscope) motifs – une différence qui n'est pas proportionnelle à la taille des sous-corpus, ce qui indique que l'augmentation du nombre de résultats obtenus en fonction de la taille du corpus n'est pas linéaire. Ces motifs étant trop nombreux pour être analysés manuellement, nous avons procédé à un premier filtrage en ne conservant que ceux qui apparaissent dans chacune des deux listes. Cela a permis de réduire considérablement leur nombre, passant de plus de 25 000 (si l'extraction avait été faite sur le corpus entier, avec un seuil minimal de 4) à seulement 3 854. L'idée sous-jacente est évidemment qu'un motif n'est représentatif des exigences rédigées au CNES que s'il est récurrent dans les deux sous-corpus, c'est-à-dire dans deux projets indépendants. Cela signifie également qu'il doit apparaître au moins quatre fois au total, soit une fois tous les 40 000 mots environ.

De très nombreux motifs ont donc été éliminés à ce stade. Parmi ceux qui étaient spécifiques à Pléiades, on peut citer notamment : « effectuer une demande »[8] (3 occurrences dans Pléiades, aucune dans Microscope), « en temps réel » (16 occurrences dans Pléiades, aucune dans Microscope) et surtout « sur réception de » (pas moins 196 occurrences dans Pléiades, mais une seule dans Microscope[9]). À l'inverse, certains motifs étaient propres à Microscope : « la position du satellite » (2 occurrences dans Microscope, aucune dans Pléiades), « au plan moyen » (4 occurrences dans Microscope, aucune dans Pléiades) ou encore « en configuration de vol nominal » (3 occurrences dans Microscope, aucune dans Pléiades).

Malgré cela, la liste ainsi obtenue comportait encore des motifs qui ne sont en rien spécifiques aux exigences (SDMC ne disposant pas de corpus de référence) et qui ne nous intéressent donc pas vraiment. Afin d'en éliminer le plus possible, nous avons procédé à une troisième extraction, cette fois sur un texte contenant des articles du journal *Le Monde* (déjà présenté au chapitre précédent). Pour que celui-ci soit suffisamment varié, nous l'avons cette fois réduit à une taille égale à exactement dix fois le nombre de mots dans le corpus d'exigences, soit un peu plus de 1 640 000 mots. Pour cette extraction, nous avons fixé le seuil de fréquence minimale à 10 (soit une occurrence tous les 160 000 mots environ, ce qui représente une fréquence relative quatre fois plus petite que pour les exigences)[10]. Le second filtrage a donc consisté à retirer de la liste des motifs trouvés pour les exigences

8. ↑Comme précisé précédemment, nous avons extrait les motifs sur la base des lemmes (« {effectuer}{un}{demande} »). Toutefois, par souci de lisibilité, nous restituons ici les formes trouvées en corpus.

9. ↑Vraisemblablement parce qu'en raison de leurs fonctions différentes, les satellites Pléiades sont très fréquemment amenés à recevoir et traiter des commandes à distance, ce qui n'est pas le cas pour Microscope.

10. ↑Une valeur obtenue en multipliant le seuil cumulé pour le corpus d'exigences (2 + 2 = 4) par la différence de taille entre le corpus d'exigences et le texte de référence (soit un facteur 10), puis en divisant le résultat par quatre. Nous avons donc retiré certains motifs même s'ils avaient une fréquence relative jusqu'à quatre fois plus faible dans la presse.

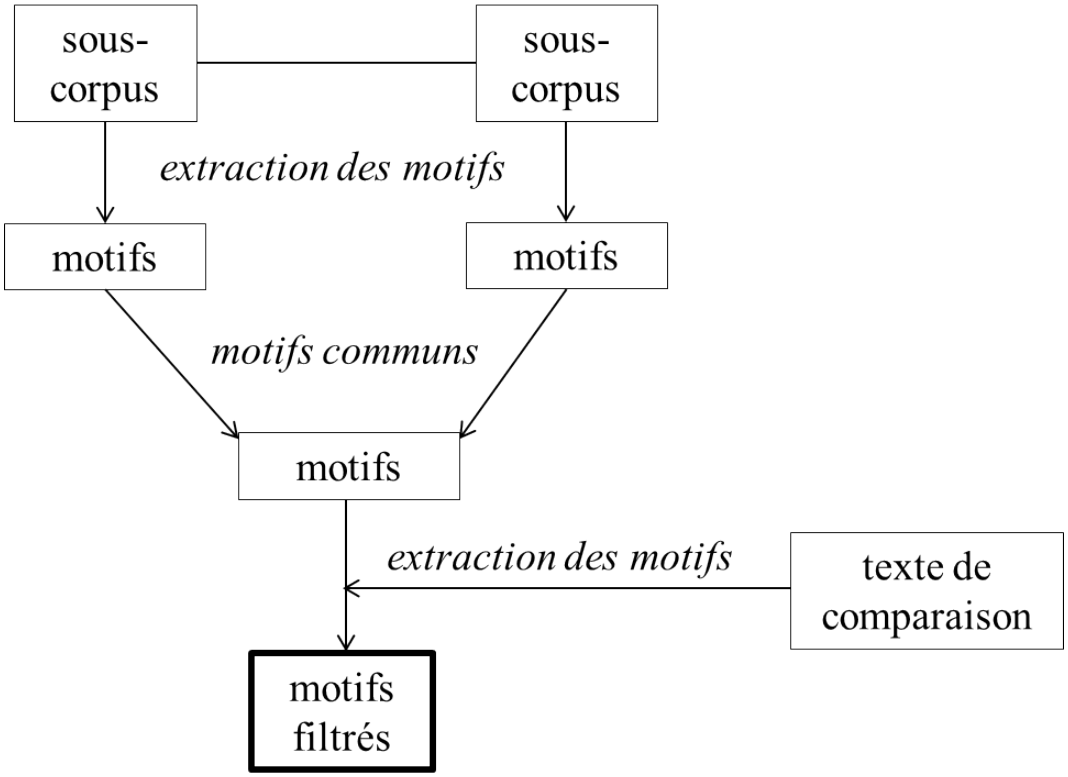
ceux qui étaient aussi obtenus dans la presse généraliste (30 475 motifs au total), lesquels ne sont probablement pas pertinents pour notre étude.

En effet, cette dernière étape a permis d'éliminer 1 413 motifs (pratiquement tous composés de deux ou trois mots) qui sont soit des séquences figées banales (« par rapport à », « en particulier », « à partir de »), soit des combinaisons de mots (pour beaucoup grammaticaux) sans signification particulière (« sur un », « si le », « avec du », « la liste »).

Notre liste finale comporte donc quelque 2 441 motifs, composés en moyenne de 2,8 éléments (les plus longs en comptent sept, par exemple « la date de début et de fin »).

La méthodologie utilisée est rappelée par la FIGURE 14.

FIGURE 14 – Méthodologie (1re étude *corpus-driven*)



4.2.2 *Statut des motifs*

Naturellement, puisque nous ne nous sommes intéressé qu'aux lemmes (sans tenir compte des étiquettes morphosyntaxiques) et que SDMC n'applique aucun traitement particulier, nombre d'entre eux ne présentent à première vue que peu

d'intérêt pour notre analyse [11] ; c'est le cas en particulier des séquences *déterminant + nom* (« l'exigence », « le fichier », « cette information ») ou *nom/verbe + préposition* (« accès au », « correspondant à », « estimé à »). Nous ne nous attarderons donc pas ici sur ces motifs, mais pensons qu'en tout cas certains seraient néanmoins utiles pour d'autres études ou applications (par exemple pour suggérer automatiquement au rédacteur des mots fréquemment associés à ceux qu'il a déjà écrits).

Nous avons repéré manuellement quelques motifs que nous avons jugés réellement intéressants, parce qu'ils semblent correspondre à des habitudes de rédaction des exigences au CNES. Parmi ceux-ci, on peut citer, entre autres, « exigence de performance » (9 occurrences), « en cas d'anomalie » (15 occurrences), « pour le calcul de » (14 occurrences), « risque de collision » (5 occurrences) et « par défaut » (48 occurrences). Comme on le voit, certains semblent relever typiquement des exigences (c'est le cas des deux premiers motifs cités), notamment parce qu'ils intègrent un élément relatif au genre (« exigence », « performance », « anomalie »), tandis que d'autres sont également présents dans d'autres types de texte, mais sont particulièrement fréquents dans notre corpus (ainsi, « par défaut » n'apparaît que cinq fois dans le texte de comparaison *Le Monde*, pourtant dix fois plus large).

Lorsque l'on s'intéresse de plus près à ces motifs, on s'aperçoit sans surprise que certains sont assimilables à des termes, des collocations ou plus généralement à des phrasèmes, ce qui prouve, si cela était nécessaire, que ces notions se rejoignent partiellement (celle de motif subsumant les autres) ; c'est le cas par exemple pour « accélération angulaire », « contrôle thermique », « centre de contrôle », « chaîne fonctionnelle », « charge utile », « composant logiciel », « compte utilisateur », « mettre à disposition », etc. Un constat similaire peut d'ailleurs être posé lorsque l'on utilise un extracteur de termes sur le corpus (ainsi que nous le faisons dans la seconde étude), ce qui montre bien que l'analyse qui est faite du corpus dépend à la fois de sa nature et de l'objectif que l'on se fixe. Notre objectif pratique, ici, est avant tout de repérer ces motifs (comparables à ce que QUINIOU et al. (2012a) nomment *motifs émergents*) qui sont caractéristiques du genre de texte et de voir en quoi ils jouent un rôle dans la transmission de l'information.

Las, très peu de motifs extraits correspondent à des phrases ou même à des propositions (la plupart d'entre eux n'étant composés que de quelques mots à peine), ce qui rend difficile leur exploitation immédiate comme *boilerplates*. On remarque malgré tout la présence de quelques structures répondant à ce format, telles que « Il est à noter que… » (en pratique, le verbe et le pronom impersonnel sont le plus souvent omis : « À noter que… » en début de phrase), « Il doit être possible de… » ou « (Le) NAM doit pouvoir/permettre de… ».

On peut néanmoins s'intéresser aux nombreuses unités de niveau inférieur et de nature hétérogène qui constituent le reste des motifs extraits. Celles-ci peuvent en effet trouver un intérêt pratique, par exemple en les considérant comme des

11. ↑Ce qui ne veut pas dire pour autant qu'ils en soient totalement dénués : par exemple, il peut être intéressant de savoir qu'un nom est utilisé préférentiellement avec un déterminant défini.

éléments non autonomes qui peuvent être combinés entre eux ou rattachés à des *boilerplates* existants (par exemple, « *À tout moment*, il doit être possible de… »), ou bien encore comme des valeurs possibles pour leurs attributs.

4.2.3 *Analyse des motifs extraits*

Comme l'on pouvait s'y attendre, les verbes « être » (utilisé comme copule ou comme auxiliaire du passif) et « devoir » (à valeur déontique) sont parmi les plus courants au sein des motifs.

La structure *être + participe passé*, extrêmement productive [12], permet de repérer facilement des structures passives récurrentes. Il n'est pas anodin que tant de passifs se retrouvent parmi les motifs fréquents dans tout le corpus ; cela confirme l'une des conclusions formulées dans le chapitre précédent, à savoir que ces structures tant décriées sont pourtant habituelles et qu'il est sans doute nécessaire de les prendre en compte.

Il en est de même pour la structure *devoir + infinitif* (elle aussi envisagée dans l'analyse *corpus-based*), où l'on retrouve cette fois les verbes « conserver », « contenir », « couvrir », « disposer », « effectuer », « exécuter », « fournir », « générer », « gérer », « permettre », « pouvoir (être) », « prendre (en compte) », « récupérer » et « respecter ».

Certains verbes sont particuliers, dans la mesure où ils se retrouvent après la négation de « devoir » (*ne devoir pas + infinitif*) : « dépasser », « entraîner » et « excéder », qui permettent de préciser soit des conséquences indésirables (« ne doit pas entraîner de perte de la performance spécifiée »), soit une limite supérieure (« ne doit pas dépasser une minute »). Nous avions déjà signalé que cette formulation est en réalité ambiguë en français, puisqu'elle peut potentiellement signifier à la fois « ne peut en aucun cas… » (interdiction) et « n'a pas l'obligation de… » (absence d'obligation) ; le contexte oriente ici néanmoins très clairement et directement vers la première interprétation.

Que certains verbes se retrouvent fréquemment à la fois après « être » et après « devoir » (en l'occurrence, « couvrir », « effectuer », « fournir », « générer », « gérer », « prendre (en compte) » et « récupérer ») montre bien – si cela était encore nécessaire – que les structures actives et passives, en miroir l'une de l'autre, peuvent être en concurrence pour exprimer un même type de besoin : on peut ainsi comparer « Le Segment Sol de Programmation PHR *doit fournir* des services […] » avec

12. [↑]On y retrouve les verbes « activer (automatiquement/manuellement) », « associer », « calculer », « commander », « composer (de) », « comprendre (entre … et …) », « concevoir », « configurer », « conserver », « couvrir », « décrire », « définir », « déployer », « dater », « dimensionner », « échanger », « effectuer », « élaborer » « fournir », « générer », « gérer », « garantir », « limiter », « lister », « livrer », « mettre (à disposition/à jour/sous tension) », « modifier », « optimiser », « positionner », « prendre (en compte) », « programmer », « réaliser », « récupérer », « répartir », « recharger », « remplir », « renseigner », « sauvegarder », « spécifier », « stocker », « supposer », « transmettre », « utiliser » et « valider ».

« La fonction "Agenda" *est fournie* par le PHR Programming Unit. », ou encore « Le CECT *doit conserver* une copie locale » avec « la donnée *est conservée* en ligne ». Une combinaison des deux (*devoir + être + participe passé*) est évidemment également possible (et loin d'être rare) : « doit être défini », « doit être mis en œuvre ».

Le verbe « être » peut aussi être suivi d'un adjectif, auquel cas le motif indique le plus souvent une qualité attendue de l'objet qui est décrit : « doit être compatible avec », « devra être conforme à », « est disponible », et bien d'autres [13]. De là, il est possible de proposer un *boilerplate* très générique, du type <*sujet*> *devoir être* <*qualité*> (où l'attribut <*qualité*> peut être remplacé par l'un des adjectifs précédemment cités). Pour plus d'efficacité, ce dernier peut être décliné en *boilerplates* plus spécifiques ; par exemple, <*système*> *devoir être* <*qualité*> ou <*propriété*> *devoir être* <*qualité*> (où <*propriété*> peut être remplacé par exemple par « attitude » ou « date », et où <*qualité*> peut être remplacé par exemple par « cohérent », mais normalement pas par « autonome ») [14].

Certaines des expressions identifiées grâce à l'extraction des motifs sont très spécialisées, telles que « gérer en configuration », qui relève de la gestion de configuration, activité qui fait l'objet d'une norme ISO (10007 : 2017) ; d'autres existent dans la langue générale mais semblent être utilisées comme des collocations spécifiques dans notre corpus, telles que « en phase de [routine/qualification] », voire « en phase de recette en vol », applicable uniquement aux satellites en orbite.

Pour finir, pour répondre plus directement à l'objectif d'homogénéisation de l'expression du besoin qui sous-tend l'utilisation de *boilerplates* (et tout en nous appuyant comme toujours sur les usages attestés), nous avons identifié plusieurs cas où des expressions diverses étaient utilisées pour exprimer des idées identiques ou très proches au sein du corpus [15] :

> (devoir) pouvoir [contrôler la cohérence des données] *vs*
> (devoir) avoir la possibilité de [sélectionner le chemin désiré] *vs*
> (devoir) être capable de [délivrer une éphéméride]

Ces trois expressions permettent de spécifier une fonctionnalité attendue du système et sont par conséquent fréquentes dans le corpus. Les douze occurrences de « avoir la possibilité de » (dont neuf précédées de « devoir ») et les vingt-sept occurrences de « être capable(s) de » (dont vingt précédées de « devoir ») pourraient

13. ↑« actif », « activable », « applicable », « autonome », « cohérent », « complet », « configurable », « indépendant », « minimal », « paramétrable », « possible », « responsable », « robuste ».

14. ↑Ainsi on peut imaginer trois *boilerplates* différents : (1) <*sujet*> *devoir être* <*qualité*>, (2) <*système*> *devoir être* <*qualité*> et (3) <*propriété*> *devoir être* <*qualité*>. Le premier est le plus générique et donc le plus permissif, les deux derniers en sont deux déclinaisons différentes (fonctionnant sur le principe de la spécialisation en informatique), plus restrictives, qui devraient lui être préférées chaque fois que cela est possible.

15. ↑Ces expressions concurrentes ne sont pas nécessairement toutes apparues dans les motifs conservés – une fois de plus, lors de l'analyse des résultats, le retour au corpus est inévitable –, mais ces derniers nous ont grandement aidé dans leur repérage (qui sans eux serait difficilement possible).

pourtant toutes être remplacées par « pouvoir » – l'inverse n'étant pas vrai. Cela met en évidence le caractère polysémique de « pouvoir ».

à l'aide de [l'IHM] *vs* grâce à [la fonction] *vs* au moyen de [l'outil] *vs* au travers de [cette fonction]

Tous ces groupes prépositionnels (dont l'expansion est à chaque fois un syntagme nominal) semblent être utilisés indifféremment pour indiquer le moyen.

être décrit *vs* défini *vs* spécifié *vs* détaillé *vs* donné [dans le document]

Ces participes passés, s'ils ne sont pas nécessairement synonymes, ont en tout cas, dans ce contexte, des sens très proches. La nature du document référencé peut varier (annexe, autre chapitre, fichier informatique, etc.) ; dans l'idéal, la référence sera donnée de façon absolue (« … est donné en exigence 20800 »), plutôt qu'en référence relative (« … est donnée au chapitre suivant »), l'ordre des exigences ou des chapitres pouvant changer ultérieurement.

avoir en charge [la réception/d'envoyer au satellite] *vs* être responsable de [la purge des espaces/veiller à l'initialisation] *vs* gérer [la liste des clés]

Les notions de gestion et de responsabilité étant fondamentales dans tous les systèmes (en particulier les systèmes complexes), de nombreuses exigences les font intervenir. On ne s'étonne dès lors pas de la variété de formulations possibles (verbe [16], verbe support, adjectif, suivis au choix d'un infinitif ou de la nominalisation déverbale correspondante, lorsqu'elle existe). D'autres variantes, non attestées dans le corpus, pourraient encore être utilisées : être chargé de, avoir la charge de, avoir la responsabilité de (on rencontre néanmoins « Les satellites sont programmés sous la responsabilité de l'Opérateur Satellite », qui est plus ou moins équivalent à « L'Opérateur Satellite a la responsabilité de la programmation des/programmer les satellites »).

être conforme à [la spécification] *vs* conformément à [la définition] *vs* dans le respect de [la spécification] *vs* respecter [les exigences de sécurité]

Il en est de même pour la notion de conformité, qui par ailleurs était déjà sous-entendue dans les exemples précédents (être défini/spécifié/etc.). À nouveau, les moyens de l'exprimer sont légion : adverbe, verbe, groupe prépositionnel, etc.

16. ↑Le verbe « gérer » pouvant lui-même être nominalisé : « responsable de la *gestion* du PHR ».

de façon à [être directement interprétables] *vs* de manière à [générer les contenus] *vs* de (telle) sorte à [ne pas être à cheval]

Ces trois expressions (qui peuvent aussi être des locutions conjonctives suivies du subjonctif : « de telle façon que ») permettent d'indiquer la conséquence recherchée d'une action.

être (strictement) supérieur (ou égal) à [0] et (strictement) inférieur (ou égal) à [20s] *vs* être compris entre [0.95] et [1.05]

On constate que ces deux tournures peuvent être utilisées afin de préciser les bornes inférieures et supérieures d'un intervalle.

La durée maximale/maximum [avant délivrance] est de [10 minutes] *vs* La durée [des activités] est limitée à *vs* ne doit pas dépasser *vs* ne doit pas excéder *vs* doit être inférieure ou égale à [2 jours]

Ces exemples montrent une nouvelle fois la variété de moyens qui peuvent être employés pour exprimer une même contrainte (en l'occurrence, une durée, mais la même remarque pourrait être faite à propos d'une longueur, d'un poids ou d'une taille [17], notamment). Bien que toutes ces formulations soient acceptables, il pourrait être avantageux de n'en retenir qu'une seule, pour garantir plus de cohérence entre les exigences ; cela faciliterait notamment le repérage d'exigences similaires et leur exploitation (ou leur traduction) automatique. Un *boilerplate* pourrait donc être une solution indiquée ici, si possible avec une phrase affirmative : *Le/la <propriété> <complément> doit être <inférieur(e) | supérieur(e)> [ou égal(e)] à <nombre> <unité>*. Si la limite inférieure et la limite supérieure sont toutes deux précisées (comme dans les exemples précédents), un autre *boilerplate* peut être retenu : *Le/la <propriété> <complément> doit être compris(e) entre <nombre> et <nombre> <unité>*.

Bien d'autres expressions encore sont utilisées en concurrence : passer en mode *vs* dans le mode *vs* vers le mode [MNOF] ; si besoin *vs* si nécessaire ; répondre aux exigences *vs* respecter les exigences ; tenir compte de *vs* prendre en compte, etc. Il convient de décider jusqu'à quel point on souhaite réduire l'hétérogénéité des exigences (ou, au contraire, laisser libre cours à la variation stylistique).

4.2.4 *Élargir la notion de boilerplate*

Le concept de *boilerplate* (ou *template*) pour les exigences, déjà brièvement évoqué en section 2.1.3, est séduisant à plus d'un titre : il est très simple à appréhender et à utiliser (des concepts similaires, tels que les textes à trous et les gabarits,

17. [↑]Comme dans cette phrase, où « maximum » semble redondant : « La taille maximum d'un fichier est limité à TAILLE_MAX_FICHIER secteurs. »

existaient déjà bien avant l'avènement de l'ingénierie des exigences [18]), tout en étant résolument utile lorsque les mêmes types de besoins doivent être répétés un grand nombre de fois, comme c'est souvent le cas dans les projets de grande envergure nécessitant des milliers d'exigences (dont certaines ne contiennent que des variations mineures). C'est d'autant plus vrai avec la généralisation des logiciels de gestion ou de vérification d'exigences (IBM Rational DOORS ou RQA, par exemple) qui permettent de manipuler aisément ces *boilerplates* prédéfinis et de vérifier la conformité des exigences. Certains logiciels ont même été développés spécifiquement à cette fin : DODT (Farfeleder et al., 2011a) permet en outre de vérifier que les termes employés appartiennent à une ontologie du domaine (ou de les y ajouter), tandis que RETA (REquirements Template Analyzer), anciennement RUBRIC (ReqUirements BoileRplate sanIty Checker) (Arora et al., 2015) présente l'avantage d'être utilisable notamment sous la forme d'un *plug-in* pour DOORS.

Les *boilerplates* peuvent être vus comme une solution semi-formelle pour la rédaction des exigences. D'un côté, les exigences élaborées en respectant une collection de *boilerplates* ressemblent en tous points à celles rédigées en langue naturelle (puisqu'en théorie, toutes les phrases sont grammaticalement correctes et peuvent être comprises sans formation spécifique préalable par les locuteurs familiers du domaine) – exception faite évidemment de leur répétitivité, qui est ici souhaitée. De l'autre, elles sont plus formelles, parce qu'elles autorisent moins d'expressivité et sont censées contenir moins d'ambiguïté (Johannessen, 2012). À tout le moins, on peut raisonnablement penser que les *boilerplates* doivent permettre d'augmenter la cohérence des spécifications : bien que la variation soit inévitable en langue naturelle (orale comme écrite), parce que la même idée peut être exprimée par des mots et des phrases divers, les exigences sont pour la majorité d'entre elles des textes écrits plutôt courts, qui, de surcroît, devraient pouvoir être lus indépendamment et dans lesquels les répétitions stylistiques sont non seulement acceptables, mais recommandées (quand bien même cela devrait les rendre moins agréables à lire). Il n'est dès lors pas surprenant qu'une recommandation du GWR (*UniformityLanguage/Similarity*) aille clairement dans ce sens :

> Use the same grammatical constructs to express similar kinds of requirement.

Son développement évoque même explicitement la possibilité d'utiliser des *boilerplates* (simplement dénommés « templates ») :

> For a deep level of similarity in the expression of requirements, the same grammtical [sic] construct can be used every time a similar requirement is being expressed.
> One way of achieving this is to have a pallet of requirement statement templates, for instance of the form:
> 'The <user type> shall be able to <capability> at a rate of <frequency> per <unit>.'

18. ↑D'une certaine façon, la fonction copier-coller que l'on retrouve sur tous les ordinateurs modernes permet d'obtenir un résultat similaire.

Il n'est toutefois pas précisé comment une telle solution peut concrètement être mise en place, puis maintenue. La difficulté cruciale et pourtant souvent éludée réside évidemment dans l'élaboration de l'ensemble de *boilerplates* qui sera proposé (ou éventuellement imposé) aux rédacteurs et qui devrait idéalement correspondre à leurs besoins.

Outre les conseils et recommandations formulés par Rupp et Joppich (2010) ou par les fondateurs de l'approche EARS (Easy Approach to Requirements Syntax) (Mavin et al., 2009), utiles mais plutôt généraux, les *boilerplates* [19] suggérés par le projet CESAR (Cost-efficient methods and processes for safety relevant embedded systems) (Rajan et Wahl, 2013) apparaissent depuis quelques années comme une référence pour beaucoup en ingénierie des exigences [20]. Parmi les exemples donnés par CESAR, on peut citer :

(exemple 1) *The <user> shall be able to <capability>.* [21]
(exemple 2) *... at a minimum rate of <number> times per <unit>*
(exemple 3) *While <operational condition>...* [22]

Ils peuvent être combinés si nécessaire : *While <operational condition>, the <user> shall be able to <capability> at a minimum rate of <number> times per <unit>*. Comme on le voit, certains (ici les exemples 2 et 3) ne peuvent d'ailleurs pas être utilisés isolément.

Ils constituent un modèle utile pour exprimer des exigences (fonctionnelles ou non fonctionnelles) et sont plutôt flexibles. Cependant, si l'exemple 2 ne demande de spécifier qu'un nombre (positif) et une unité (de temps), et n'est donc sujet qu'à de légères variations, il n'en va pas de même pour les exemples 1 et 3, où « user », « capability » et surtout « operational condition » pourraient être remplacés par une quasi-infinité de valeurs très différentes [23] : si l'attribut « user » est relativement restrictif (particulièrement si ces *boilerplates* sont utilisés conjointement avec une ontologie du domaine, dans laquelle la classe « user » serait correctement définie), ce n'est pas le cas de « operational condition », qui est une catégorie ouverte très large. Le rédacteur n'ayant pas vraiment d'indication sur la façon de remplir le

19. ↑Les *boilerplates* y sont définis comme suit : « semi-complete requirements that are parameterized to suit a particular context ». Le « Boilerplate-based RSL » est l'un des « semi-formal Requirement Specification Languages (RSLs) » et est considéré comme plus formel que le « Guided Natural Language », mais moins que le « pattern-based RSL » (un exemple de « pattern » est donné par l'exigence suivante : « diagnosisTaskActivate occurs each 10ms with jitter 1ms »).

20. ↑La liste en ligne de *boilerplates* de Jeremy Dick est également souvent citée – y compris par les auteurs de CESAR –, mais n'est malheureusement plus disponible depuis longtemps (http://www.requirementsengineering.info/).

21. ↑On trouve, dans le corpus, quelques occurrences de phrases correspondant à cette structure : « sera/seront capable(s) de » (deux fois) et « devra/devront être capable(s) de » (quatre fois).

22. ↑L'AFIS (2012) donne quelques exemples de boilerplates en français, pour lesquels la ressemblance est évidente : « *Quand <condition_opérationnelle>...* » ; « *le <système> doit <action>* ».

23. ↑Dans le document CESAR, « order entry clerk », « raise an invoice » et « (While) disconnected from a source of power » sont respectivement donnés comme valeurs possibles.

« trou », il est finalement libre de le faire comme il l'entend [24]. D'un autre côté, tel qu'il est, ce *boilerplate* ne lui permet pas de remplacer « While » par « Even while » ou toute autre conjonction peut-être plus adaptée dans ce contexte. Au-delà des limitations que nous venons d'évoquer, certains des *boilerplates* proposés par CESAR nous semblent très proches, par exemple « ... at least <number> times per <unit> » et l'exemple 2, pour lesquels l'unique différence est que « at a minimum rate of » est remplacé par « at least ».

C'est pourquoi ces *boilerplates* sont probablement de bons exemples, utiles mais révélateurs des principales limites de tels formats : il peut être difficile pour l'utilisateur d'en choisir un plutôt qu'un autre ; ils sont parfois excessivement restrictifs ; mais surtout, ils sont souvent trop permissifs, essentiellement parce qu'ils sont trop génériques (et que leur nombre ne doit pas devenir exagérément élevé). Ils aident certainement à réduire la variation inter-exigences, mais pourraient probablement être raffinés en prenant davantage en compte l'environnement particulier dans lequel ils seront utilisés, ce afin qu'ils soient plus spécifiques et adaptés. Selon nous, un tel objectif ne peut être atteint qu'en analysant préalablement d'authentiques spécifications à l'aide d'outils de traitement automatique du langage ou de *text mining*, ainsi que nous l'avons fait dans cette première étude. Aucun *boilerplate* n'est actuellement utilisé au CNES, ce qui se manifeste par une grande hétérogénéité dans le corpus, ainsi que nous l'avons déjà montré. Cependant, nous ne doutons pas de leur intérêt, notamment parce que l'utilisation du copier-coller pour certaines exigences (au sein d'un même projet) est indéniable – en témoignent les 193 exigences Pléiades commençant par « Sur réception de cette TC ».

Notre proposition est de mettre à profit les motifs extraits et de dépasser le concept canonique de *boilerplate* en combinant un nombre limité de structures relativement génériques (du même type que celles proposées plus haut) avec des éléments (semi-)fixes [25] plus ou moins spécifiques (tels que « être conforme à » ou « en prenant en compte »), de taille et de nature variables, qui peuvent remplir les trous ou compléter les *boilerplates* de façon flexible. Ces éléments seraient évidemment beaucoup plus nombreux, à un point tel qu'il ne serait pas possible pour un rédacteur de les mémoriser tous, mais ils pourraient être intégrés dans un logiciel d'aide à la rédaction des exigences, avec deux bénéfices majeurs. Premièrement, lorsqu'un rédacteur entre les premiers mots d'un élément fréquent, la suite pourrait lui être automatiquement suggérée (de façon bien entendu totalement facultative), à la manière du complètement automatique présent dans certains logiciels ou moteurs de recherche : cela pourrait l'aider dans la difficile tâche de rédaction, lui faire gagner du temps et l'inciter à préférer des tournures déjà attestées. Les motifs extraits pourraient ainsi constituer une « mémoire de rédaction », périodiquement mise à jour. Deuxièmement, si une expression est reconnue, mais qu'il

24. ↑Par exemple, on aurait pu préférer une version plus explicite, dans laquelle le sujet est mentionné dans la condition : « While *the photographer is* disconnected from a source of power ».

25. ↑À rapprocher des *collocational frameworks* de Renouf et Sinclair (1991) – séquences discontinues de mots grammaticaux à l'intérieur desquelles s'insèrent des mots lexicaux – ou simplement des segments répétés (Lafon et Salem, 1983) et des *lexical bundles* (Biber, 2006).

a été décidé antérieurement qu'une autre devrait lui être préférée (par exemple, « devoir être inférieur ou égal à » plutôt que « ne devoir pas excéder », pour éviter une phrase négative), elle pourrait être automatiquement mise en évidence, et la forme préférée suggérée (à la façon des correcteurs orthographiques et grammaticaux intégrés à la plupart des logiciels de traitement de texte) – laissant le choix au rédacteur d'accepter ou non la modification. Ceci constituerait une étape supplémentaire dans la réduction de la variation, sans pour autant réduire l'expressivité.

4.2.5 *Conclusion*

À l'aide de moyens réduits (les deux sous-corpus et un seul outil), il est possible d'extraire de très nombreux motifs du corpus. L'utilisation d'une ressource supplémentaire, facilement accessible, permet en outre de les filtrer pour ne conserver que les plus spécifiques. En fonction des paramètres choisis pour l'extraction, on peut ainsi arriver à un nombre de motifs suffisamment limité (de l'ordre de quelques milliers) pour permettre ensuite une révision manuelle. Bien sûr, d'autres raffinements seraient possibles pour filtrer davantage les résultats (par exemple, éliminer tous les motifs qui se terminent par un déterminant).

Le premier intérêt de cette étude est de repérer des structures typiques du genre ou du domaine : « en cours de + nominalisation déverbale (constitution, transfert, passage, réalisation, traitement, création, téléchargement, récupération, etc.) » ou « en configuration + adjectif (nominale, dégradée, opérationnelle, etc.) », pour n'en citer que deux. Dans une perspective applicative visant à réduire la variation intralinguistique au sein des spécifications, les motifs peuvent aussi servir de modèles (ou du moins d'inspiration) pour des *boilerplates* plus spécifiques et plus efficaces. De ce point de vue, on regrettera que très peu de structures phrastiques (pouvant fournir des *boilerplates* « prêts à l'emploi ») aient émergé ; cela tient probablement à la fois à la nature du corpus (où, en dehors des exigences copiées-collées (spécifiques à un projet), les phrases suffisamment semblables sont rares – ce qui n'a rien de surprenant en langue naturelle et au vu de sa taille modeste) et à la méthodologie, qui pourrait être adaptée. En particulier, la possibilité de laisser des *gaps* d'un ou plusieurs *items* dans les motifs est une piste prometteuse, mais il faudrait, pour pouvoir l'exploiter, que la partie variable du motif soit indiquée dans les résultats retournés par SDMC [26].

En complément des *boilerplates* eux-mêmes, nous proposons d'opter pour une granularité plus fine en réutilisant certains des motifs fréquents (assimilables à des segments répétés) comme valeurs possibles des attributs ou comme compléments des *boilerplates* (ces derniers ne pouvant couvrir tous les cas possibles). Nous croyons que cette suggestion permettrait de faciliter la rédaction des exigences et de réduire encore un peu plus la variation (si possible au profit des tournures considérées comme plus claires) en encourageant la réutilisation de formulations

26. ↑Exemple fictif : {le}{serveur}{mettre}{à}{[X]}{le}{donnée} (avec « [X] » pour « jour » ou « disposition »), plutôt que {le}{serveur}{mettre}{à}{le}{donnée}.

fréquentes au sein des spécifications. Elle requiert, pour être efficace, que celles-ci soient suggérées automatiquement lors de la rédaction et de la vérification des exigences.

Le choix des formulations à promouvoir (au détriment d'autres) ne nous appartient pas, et nous ne partons pas non plus du principe que celles qui sont les plus fréquentes sont forcément les meilleures : il peut aussi s'agir de structures à éviter et donc à déconseiller. Il nécessite de recueillir l'avis des experts, idéalement « en contexte » (c'est-à-dire dans des exemples d'exigences possibles). Cette évaluation peut se faire sous la forme d'un questionnaire semblable à celui que nous présenterons dans le prochain chapitre.

4.3 DEUXIÈME ÉTUDE

4.3.1 *Introduction*

Toujours dans l'objectif de construire la grammaire du genre textuel des exigences, nous essayons cette fois de voir si l'utilisation de termes candidats extraits automatiquement permet d'accéder plus rapidement aux structures (motifs) propres au genre. Nous supposons en effet que ces dernières ont plus de chances de contenir des lexèmes du genre « exigences » (tels que « fonction » ou « défaillance », par exemple). La difficulté, pour les repérer, consiste dès lors à filtrer semi-automatiquement les candidats-termes pour ne conserver que ceux qui relèvent du genre, et non – comme c'est le plus souvent le cas – du domaine.

Même si la tentation existe, nous préférons nous garder de les appeler « termes » (du genre) et nous nous en tiendrons à « lexèmes » (ou « phrasèmes » pour les unités multilexémiques (MEL'ČUK, 2013)) du genre, car il n'est pas certain que toutes les unités auxquelles nous nous intéressons dans cette étude répondent strictement à la définition de *terme* (qui reste cependant variable d'un auteur et d'une tradition à l'autre[27] (HUMBLEY, 2001)) : si RASTIER (1995) a pu dire du terme qu'il est « réputé exprimer univoquement un concept », qu'il définit à son tour comme un « signifié normé », ajoutant que « les "concepts" à traiter sont des signifiés normés par les disciplines, reconnus par la terminologie [...] », Gaudin précisera par la suite que les termes « ne doivent plus être conçus comme des étiquettes de concepts, mais resitués dans le cadre des échanges langagiers au sein desquels ils apparaissent et se maintiennent » (GAUDIN, 2005 : 86-87) : « ils sont utilisés de façon précise dans les discours experts et possèdent une signification normalisée, mais ne sont pas pour autant indépendants des tensions entre groupes d'énonciateurs, des luttes d'influences, des choix sociopolitiques, etc. » (GAUDIN, 2005 : 88). Élever ces unités au rang de termes, ce serait donc considérer qu'elles expriment chacune

27. ↑La distinction entre terme et phrasème n'est elle-même pas aussi claire qu'il n'y paraît, et certains auteurs proposent d'y voir plutôt un continuum (GAUTIER, 2002).

un concept (lui-même en relation avec d'autres [28]) clairement défini au sein d'une discipline ou d'un domaine de connaissance [29] ; or, rien ne nous permet de l'affirmer et nous ne souhaitons pas nous enfermer dans une vision trop restrictive, qui risquerait de nous priver de résultats pourtant utiles dans la perspective de caractériser un genre textuel (indépendamment du ou des domaines concernés).

À l'instar par exemple de TUTIN (2007a), DROUIN (2007) et HATIER (2013), qui se servent d'outils et de corpus différents (représentant diverses disciplines : la linguistique, l'économie et la médecine pour Tutin ; neuf domaines, dont la psychologie, la physique et l'informatique pour Drouin ; dix disciplines de sciences humaines et sociales pour Hatier) afin d'établir un lexique scientifique transdisciplinaire, nous souhaiterions pouvoir extraire de notre corpus les lexèmes appartenant au *genre* des exigences, indépendamment du *domaine* dont dépendent ces exigences (en l'occurrence, celui des systèmes spatiaux de façon générale, et secondairement l'optique pour Pléiades et les mesures physiques pour Microscope).

En effet, la grammaire que nous essayons de construire devrait idéalement être suffisamment générale pour pouvoir concerner les exigences de tous les projets du CNES, quels que soient la nature et l'objectif de l'objet décrit dans les spécifications – qu'il s'agisse d'un satellite, d'un logiciel informatique ou de tout autre système. Dans les faits, toutefois, la distinction entre genre et domaine n'est pas toujours simple, ni forcément claire (voir section 1.4), et c'est particulièrement le cas s'agissant des exigences.

La grammaire du genre « exigences » devrait en théorie être indépendante du domaine (industries aéronautique et aérospatiale, informatique, etc.), mais cela ne va donc pas sans poser quelques difficultés. Si, comme on l'a montré, certaines caractéristiques semblent en effet inhérentes à ces textes (par nature à la fois descriptifs et injonctifs, en ce qu'ils décrivent un objet ou un système qui n'existe pas encore, mais devra exister et se conformer à ces caractéristiques), d'autres, en revanche, peuvent être davantage liées à la nature de l'« objet » qui est spécifié et au domaine auquel il appartient.

Par conséquent, on peut espérer y retrouver des lexèmes relevant du type de texte (ce sont ceux-là qui nous intéressent), mais aussi des termes relevant du domaine auquel appartient le futur objet qui est décrit. Nous supposons raisonna-

28. ↑« Alan Melby (1991 : 22) opère une distinction [...] lorsqu'il parle d'unité lexicale et/ou terminologique. L'unité devient terminologique lorsqu'elle s'insère dans un réseau domainier, par opposition à un réseau langagier [...]. » (CABRÉ, 1998 : 151).

29. ↑Pour TUTIN (2007b), c'est bien le cas du lexique transdisciplinaire des écrits scientifiques (cf. *infra*) : « Ce lexique peut être considéré comme un lexique de genre, n'intégrant pas la terminologie du domaine, mais renvoyant aux concepts mis en œuvre dans l'activité scientifique (*examiner*, *prouver*, *réfuter*, *concluant*, *hypothèse*, *examen*, *encourageant*...). » DROUIN (2007 : 45) fait remarquer que « [l]e lexique scientifique transdisciplinaire n'est pas saillant dans les textes scientifiques dans la mesure où, contrairement à la terminologie, il se rencontre également dans la langue commune. » – une difficulté à laquelle nous sommes également confronté avec les exigences. PECMAN (2007) choisit de parler de « phraséolexique » de la langue scientifique générale.

blement que le niveau lexical est le plus affecté par le domaine – puisqu'il réfère plus directement à l'objet en question –, mais il n'est pas non plus à exclure que les structures syntaxiques puissent elles aussi différer d'un domaine à l'autre.

Que les outils terminologiques existants visent avant tout à faciliter le repérage des termes d'un domaine n'est absolument pas surprenant. Ceci s'explique par les besoins (notamment en entreprise) auxquels ils répondent lorsqu'il est nécessaire d'établir des terminologies ou des ontologies d'un domaine[30], ou encore d'indexer des documents textuels (comme c'est le cas au CNES). Pour toutes ces applications, en effet, ce que nous appelons les lexèmes du genre (qu'il s'agisse de spécifications, de textes procéduraux ou autres) ne trouvent que rarement un intérêt direct (soit parce que le genre du document – et donc les formes langagières qui s'y rapportent directement – n'importe pas tant, soit parce que celui-ci peut de toute façon être connu par d'autres moyens, typiquement des métadonnées). Ils tendent donc à être ignorés, à l'égal entre autres des mots grammaticaux. Si de tels outils retourneront vraisemblablement (en vertu du critère statistique) des termes candidats relevant du genre, ces derniers seront de ce fait le plus souvent considérés par les utilisateurs au mieux comme accessoires, au pire comme du bruit (au même titre que d'autres régularités purement langagières qui ne peuvent pas être associées à un concept). Cela explique, par corollaire, pourquoi des outils destinés spécifiquement à leur extraction n'aient, à notre connaissance, pas (encore ?) été développés. Cela ne signifie pas pour autant que la question du genre soit négligée en terminologie ; mais la prise en compte du ou des genres des textes lors de la constitution du corpus spécialisé doit avant tout permettre un meilleur repérage des termes du domaine (Rogers, 2000)[31]. Enfin, les traits (au sens large) caractéristiques d'un genre sont exploités par plusieurs études dont l'objectif est la classification automatique de textes (Stamatatos, Fakotakis et Kokkinakis, 2000 ; Forsbom, 2005, parmi d'autres), mais il s'agit le plus souvent d'approches purement statistiques assez éloignées de ce que nous cherchons à faire.

Dans cette étude, comme nous l'avons dit, nous nous intéressons justement à ces mots propres au genre des exigences (au détriment des termes du domaine, jugés trop spécifiques puisque généralement liés à un projet en particulier), c'est-à-dire plus fréquents dans ce genre textuel que dans d'autres, et donc représentatifs (au moins au niveau lexical) d'un mode de rédaction particulier. Puisque le développement d'un outil *ad hoc* n'est pas envisageable à ce stade, que, toujours dans une approche *corpus-driven*, nous souhaitons ici analyser les données du corpus sans avoir à injecter de règles ou connaissances linguistiques dans un outil d'extraction existant et qu'une révision entièrement manuelle s'avérerait trop coûteuse en temps, la méthode que nous avons expérimentée consiste à partir des résul-

30. ↑La possibilité de filtrer par domaine les termes présents dans les bases terminologiques est d'ailleurs un besoin identifié par leurs utilisateurs (Blancafort et al., 2011).

31. ↑On sait que les extracteurs de termes obtiennent généralement de meilleurs résultats sur des corpus plus « didactiques », dans lesquels les liens entre les termes (et les concepts auxquels ils renvoient) sont plus explicites. Plus généralement, les performances des outils de traitement automatique du langage varient selon le corpus et en particulier selon le genre textuel (Jacques et Aussenac-Gilles, 2006).

tats fournis par un outil existant et fiable, puis à les filtrer (c'est-dire à éliminer un maximum de candidats-termes que l'on sait appartenir au domaine, probablement beaucoup plus nombreux) à l'aide d'une ressource disponible au CNES pour ne conserver autant que possible que les candidats les plus susceptibles d'être finalement retenus par un validateur humain comme caractéristique du genre « exigences ».

Par la suite, nous pensons que ces lexèmes du genre doivent nous permettre d'accéder plus rapidement à des segments (plus larges) relevant également du genre, en jouant à leur tour le rôle de filtres pour réduire le nombre de résultats retournés par des outils tels que SDMC, dans la mesure où il est permis de penser que les motifs les plus typiques du genre sont basés sur (ou du moins contiennent) des lexèmes de ce même genre.

Les exigences et spécifications représentent donc de ce point de vue des corpus spécialisés et inhabituels (mais pas isolés !) qui posent de nouvelles difficultés – et défis – pour l'extraction terminologique et les outils automatiques qui y sont dédiés.

4.3.2 *Méthodologie*

Toutes les opérations décrites dans cette section ont été réalisées sur nos deux sous-corpus, Pléiades et Microscope. Rappelons que s'ils sont tous deux représentatifs des mêmes niveaux de spécification pour les deux projets, ces derniers diffèrent nettement de par leur ampleur et surtout leurs objectifs. En d'autres termes, les (sous-)domaines auxquels ils se rapportent sont au moins en partie différents.

4.3.2.1 *Termes candidats*

Tout d'abord, les candidats-termes ont été extraits indépendamment sur les deux sous-corpus à l'aide de l'extracteur terminologique développé pour Talismane : se basant sur une analyse syntaxique préalable, il extrait uniquement des syntagmes nominaux (contigus)[32]. Avec un seuil de fréquence minimum fixé à cinq, deux listes de termes candidats ont ainsi été obtenues : 1 551 résultats pour Pléiade, 716 pour Microscope.

La première étape visant à éliminer les candidats-termes relevant du domaine a consisté à retirer de chacune de ces deux listes tous ceux qui appartiennent également à une liste de termes utilisée par le CNES pour indexer les documents dans leur base de connaissance, dont les entrées ont donc constitué pour nous des *stop words*. Cette liste présente l'avantage d'être déjà disponible au CNES, même si son usage est ici détourné puisqu'il n'avait pas été prévu lors de sa conception. Elle a été enrichie pendant plusieurs années (et continue à l'être) à l'aide de documents internes de différents types ; son contenu a de plus été soigneusement validé

32. ↑À l'heure actuelle, il n'est utilisable que sur des corpus en français, mais il devrait à terme devenir plus générique.

par des experts du domaine. Nous pouvons dès lors raisonnablement considérer que les termes qu'elle contient sont représentatifs des différents projets menés au CNES lors de ces dernières années. En outre, il semble exclu qu'elle contienne elle-même des entrées relevant du genre « exigences », étant donné que ceux-ci, en raison de leur trop grande généralité, n'auraient pas vraiment d'intérêt pour l'indexation. Cette étape a permis d'éliminer près de deux cents résultats (12,64 %) pour Pléiades (1 551 avant filtrage, 1 355 après) et une centaine (16,48 %) pour Microscope (716 avant filtrage, 598 après) : « propulsion » (dans les deux), « vidage » (dans Pléiades) et « lanceur » (dans Microscope), par exemple.

La seconde étape[33], quant à elle, a consisté à ne conserver, parmi les candidats-termes ainsi filtrés, que ceux qui apparaissaient dans chacune des deux listes (Pléiades et Microscope) et dont l'usage n'est de ce fait pas limité à un seul domaine. La liste d'entrées communes qui en résulte est bien plus réduite, puisqu'elle ne contient plus que 300 entrées (ce qui signifie que 1 055 entrées étaient exclusives à Pléiades et 298 autres étaient exclusives à Microscope). Ainsi, tandis que la première étape a permis d'éliminer des candidats relatifs au domaine à un niveau plus général (« satellite », par exemple), ici certains candidats n'ont pas été retenus parce qu'ils étaient vraisemblablement plus dépendants d'un des deux projets (et donc plus spécialisés) : « magnétomètre » (Microscope) ou « produit image » (Pléiades), par exemple[34].

Finalement, nous avons rapidement parcouru les termes candidats restants pour éliminer quelques entrées qui constituaient du bruit évident. Notre liste finale est constituée de 263 entrées (données en annexe B) ; sans les étapes intermédiaires, cette liste aurait été composée de plus de 1 850 candidats, ou même de près de 2 000 si l'extraction avait été faite sur l'intégralité du corpus en même temps.

4.3.2.2 *Motifs textuels*

Bien entendu, une grammaire de genre ne devrait pas se limiter au lexique et nous entendons donc nous servir des résultats de l'extraction terminologique pour identifier des structures syntaxiques récurrentes ou, tout au moins, des motifs textuels fréquents contenant ces lexèmes ou phrasèmes du genre. Ceux-ci étant tous des noms ou des syntagmes nominaux, on peut faire l'hypothèse qu'ils vont préférentiellement apparaître en position d'arguments dans des structures caractéristiques du genre.

33. ↑Nous avons procédé dans cet ordre à des fins d'analyse (pouvoir visualiser les termes retirés par la *stop list* sous-corpus par sous-corpus), mais ces deux étapes pourraient avantageusement être inversées (ce qui nécessiterait moins d'opérations) – ou l'on pourrait également souhaiter n'en appliquer qu'une des deux : la seconde si les ressources nécessaires pour la première ne sont pas disponibles, la première si le rappel est jugé plus important que la précision.

34. ↑Cependant, il faut garder à l'esprit la différence de taille entre les deux sous-corpus : des candidats-termes tels que « priorité » auraient peut-être également pu apparaître dans la liste de Microscope si plus d'exigences avaient été récoltées.

De nouveau, nous avons fait usage de SDMC pour extraire des motifs de lemmes dans le corpus d'exigences (avec d'autres paramètres que dans la première étude : motifs fermés, seuil de fréquence minimale fixé à trois), obtenant ainsi quelque 14 000 résultats – ce qui exclut évidemment une révision manuelle par des experts. Cette fois, nous avons décidé de ne conserver (à l'aide d'un script) que les motifs contenant au moins un des 258 lexèmes identifiés à l'étape précédente (par souci de simplicité, les syntagmes nominaux ont été réduits à leur tête) : en effet, nous partons du principe que les structures basées sur des lexèmes appartenant au genre étudié ont plus de chance d'être elles-mêmes typiques de ce même genre. Cette restriction a permis de limiter le nombre de motifs à environ 6 000 (sans avoir validé manuellement les lexèmes au préalable (cf. *infra*), ce qui aurait permis de faire baisser encore ce nombre), parmi lesquels « être connaître avec un [précision] meilleur que (nombre) »[35], « être conforme au [format] »[36] et « devoir respecter le [contrainte] ». Le nombre de motifs restant malgré tout élevé, nous avons décidé de nous concentrer sur ceux qui contiennent un verbe. De cette façon, nous nous plaçons à un niveau intermédiaire entre le lexique et le discours.

La méthodologie proposée est illustrée par la FIGURE 15.

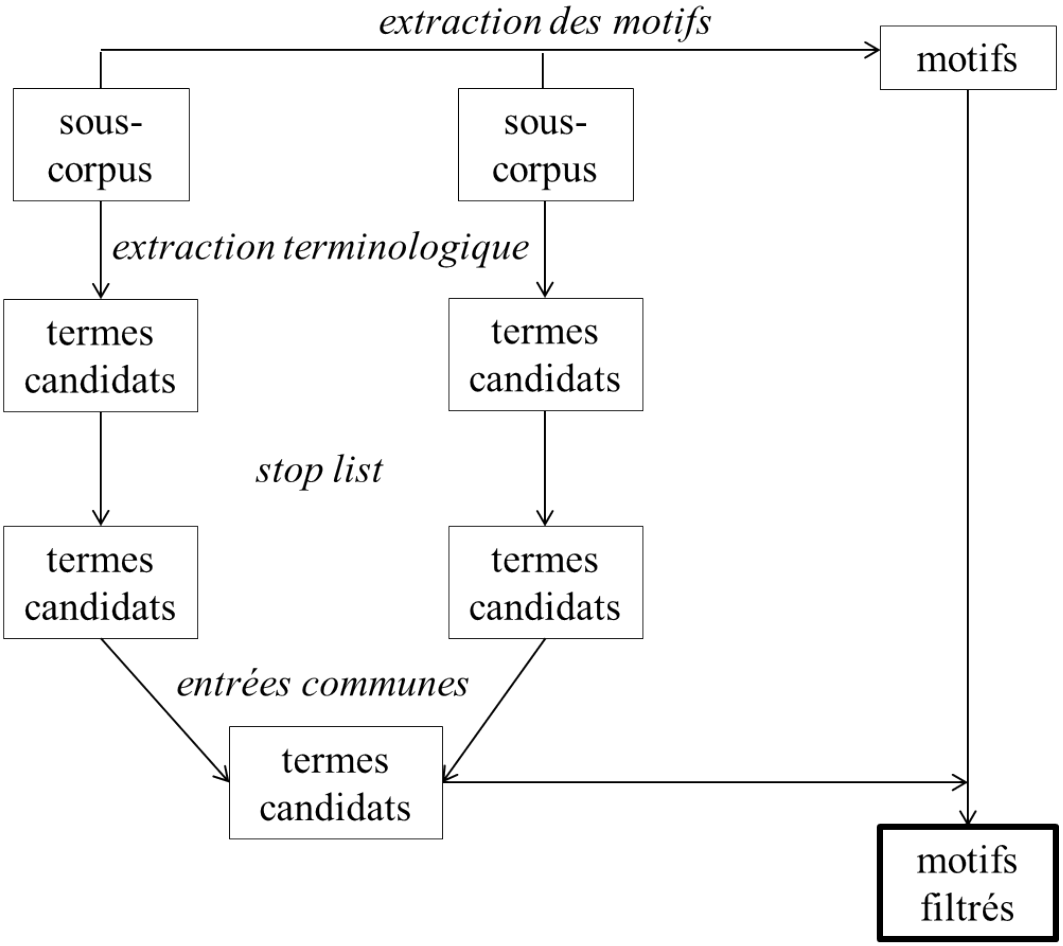
4.3.3 *Aperçu des résultats*

Les résultats ainsi obtenus sont de natures diverses. Concernant les lexèmes, tout d'abord, il faut remarquer que, malgré les deux étapes de filtrage, certains (candidats-)termes relevant plutôt du domaine subsistent : « tuyère », « lune », « éclipse », par exemple, ou encore plusieurs sigles (« HKTM », « BDS », « SST », etc.). D'autres, néanmoins, se révèlent beaucoup plus intéressants pour notre étude, parce qu'ils semblent caractéristiques du genre. Ils peuvent définir un besoin (« *besoin* de test/programmation/restitution ») ou les propriétés de l'objet décrit (« *taille* du buffer temporaire/du paquet TM », « *durée* de désaturation/la manœuvre ») ; ils peuvent spécifier des fonctions attendues (« *fonction* de gestion/filtrage ») ; ils peuvent aussi être relatifs à la gestion du projet : problèmes potentiels (« défaillance », « défaut » – mais pas « anomalie », présent pour sa part dans la liste du CNES), documentation nécessaire (« *rapport* d'avancement/d'expertise »), validation (« acceptation », « confirmation », « autorisation »). D'autres encore peuvent relever tantôt plutôt du genre (« *date* de début du produit »), tantôt plutôt du domaine (« *dates* de début et de fin de vidage TM »), en particulier en fonction de leurs modifieurs : comme déjà dit, la distinction entre les deux n'est pas toujours si nette et le recours au contexte peut être nécessaire ; quoi qu'il en soit, et comme toujours, l'interprétation et l'exploitation des résultats sont conditionnées par les objectifs poursuivis. Enfin, il est intéressant de remarquer que les lexèmes retenus concernent à la fois les exigences fonctionnelles (« fonctionnalité », « réception », etc.) et non fonctionnelles (« disponibilité », « cohérence », etc.) ; cette différence

35. ↑Les lexèmes qui ont permis au motif d'être retenu sont indiqués entre crochets. Pour rappel, ces motifs sont constitués uniquement de lemmes.

36. ↑« conforme » aurait aussi pu constituer un lexème intéressant, mais l'extracteur de termes de Talismane n'extrait malheureusement pas les adjectifs (dans sa version actuelle du moins).

FIGURE 15 – Méthodologie (2e étude *corpus-driven*)



pourrait être prise en compte pour raffiner le lexique et la grammaire du genre.

Parmi les candidats-termes initiaux qui n'ont pas été retenus, quelques-uns nous informent malgré tout sur le fonctionnement des syntagmes nominaux au sein du corpus spécialisé. Remarquons tout d'abord que certains semblent très proches : « secteurs défaillants » et « secteurs défectueux », par exemple, peuvent sans doute renvoyer à une même réalité ; si c'est le cas, il s'agit de synonymes, pourtant proscrits à la fois par la terminologie classique et les langues contrôlées. On constate également des variations formelles, qui peuvent se traduire par l'omission (non systématique) de prépositions et de déterminants [37] : c'est le cas, entre autres, pour « charge batterie » (*vs* « charge *de la* batterie »), « fonction contrôle (des tranches/du déploiement) » (*vs* « fonction *de* contrôle »), « calibration FEE » (*vs* « calibration *de* FEE ») et « calcul durée de convergence » (*vs* « calcul *de la* durée »). Un exemple plus marqué de construction paratactique de ce type nous est donné par l'expression (qui semble figée) « fonction de gestion bord instrument », dans laquelle trois

37. [↑]Ces termes candidats fonctionnent donc tantôt comme des syntagmes classiques, tantôt sur un mode qualifié de « déviant » (CONDAMINES et PICTON, 2014), typique des corpus spécialisés.

noms sont ainsi juxtaposés.

Concernant les motifs filtrés (c'est-à-dire contenant au moins un des lexèmes retenus), enfin, les verbes qui y apparaissent le plus souvent sont les suivants : « être », « devoir », « mettre », « prendre (en compte) », « fournir », « pouvoir », « définir », « passer (en mode/dans l'état) », « contenir », « donner », « utiliser », « gérer », « sélectionner », « rejeter », « traiter », « correspondre », « générer », « décrire », « tenir (compte) », « exécuter », « vérifier » et « calculer ». Certaines structures basées sur ces verbes sont effectivement typiques du corpus [38] :

SN permettre (de SV+SNdéverbal)
« Le DUPC permettra de modifier localement les paramètres du calcul »

SN fournir SN2 [à SN3] / SN fournir [à SN2] SN3 [39]
« Cette interface fournit les positions navigateur de l'instrument »
« Le système de navigation fournira au système informatique central une référence de temps »

SN utiliser SN2 [pour SV]
« Le système GIDE utilisera le protocole FTP pour effectuer les transferts »

Sur réception de cette TC, le LVC exécuter la procédure de mise (ON+OFF) de SN[, par l'envoi de commandes (sur+vers+à) SN3]
« Sur réception de cette TC, le LVC exécute la procédure de mise ON de la carte IOT sélectionnée, par l'envoi de commandes discrètes sur l'OBMU »

SNdéverbal devoir s'exécuter [conditions]
« La consolidation du scénario de travail au CECT doit s'exécuter en moins de 15 secondes »

SN devoir ((avoir la capacité de+être capable de)+(avoir l'autorisation de+être autorisé à)) traiter SN2
« Le CCC doit avoir la capacité de récupérer et traiter 291 Mo de TM par jour »

Ces structures récurrentes font donc partie, avec d'autres, de la grammaire du genre « exigences » au CNES.

4.3.4 *Conclusion*

Les spécifications et les exigences représentent un type particulier de discours spécialisé, dans lequel les termes du domaine et les lexèmes du genre sont souvent liés, ce qui complique leur identification automatique. Les outils classiques sont en effet avant tout destinés à l'extraction des termes d'un domaine donné, le genre

38. ↑SN : syntagme nominal ; SV : syntagme verbal ; [...] : éléments facultatifs ; (a+b+c) : alternative.

39. ↑Variation dans l'ordre des compléments du verbe.

des textes composant le corpus n'étant pour eux qu'une propriété accessoire qui n'est généralement pas prise en compte. Notre étude et des études similaires sur d'autres genres de texte (notamment les articles scientifiques) montrent pourtant qu'il existe un intérêt et un besoin de pouvoir identifier des segments textuels récurrents à travers différents documents du même type ou du même genre, indépendamment des sujets traités ou des domaines concernés.

En attendant que des logiciels spécifiquement destinés à cet usage soient mis librement à disposition de la communauté, l'approche simple que nous avons utilisée ici consiste à exploiter les résultats fournis par un extracteur de termes s'appuyant sur des critères syntaxiques et statistiques, puis à filtrer ceux-ci en éliminant automatiquement tous les termes candidats susceptibles de relever avant tout du domaine. Pour ce faire, deux techniques (qui peuvent utilement se compléter) ont été expérimentées. La première nécessite de disposer d'une liste des termes du domaine auxquels appartiennent les documents, pour les retirer de la liste des candidats-termes. Une telle liste peut être obtenue soit en partant de terminologies déjà existantes (comme c'est le cas pour un nombre sans cesse croissant de domaines), soit en adaptant des ressources créées au sein de l'entreprise à d'autres fins (dans notre cas, une liste de mots-clés utilisés pour l'indexation de divers documents au sein d'une base de connaissances). La seconde repose sur la comparaison de plusieurs (sous-)corpus représentant des domaines suffisamment différents et part du postulat que si ces (sous-)corpus spécialisés ont en commun le genre des textes, mais se distinguent par les domaines traités, alors les éléments qui se retrouvent dans chacun d'eux ont toutes les chances de dépendre du genre [40]. Ces deux étapes permettent ainsi de réduire plus ou moins fortement le nombre de termes candidats à réviser.

Certains des résultats que nous avons fait émerger de notre corpus sans connaissances *a priori* nous paraissent très intéressants : « tâche », « support », « maintenance » ou encore « contrainte » semblent en effet renvoyer à des concepts fondamentaux de l'activité de spécification (et ce, indépendamment d'un projet ou même d'un domaine particulier). Ceci nous conforte dans l'hypothèse de l'existence de « lexèmes du genre » dont l'identification (quels que soient la méthode ou les outils qui sont utilisés à cette fin) présente un intérêt théorique et applicatif.

Aussi nous sommes-nous servi de ces lexèmes comme d'un autre critère pour filtrer les motifs proposés par SDMC, en partant cette fois du principe que les motifs contenant un (potentiel) lexème du genre étaient probablement les plus susceptibles d'être représentatifs de la grammaire de ce genre. Cela nous a permis d'obtenir des résultats différents de ceux de la première étude, dont certains nous semblent assurément dignes d'intérêt (par exemple, « le [système] assurera la [fonction] de (+ nominalisation déverbale) »).

40. ↑Certains des lexèmes retenus sont également fréquents dans les exigences de projets informatiques : « besoin », « taille », « durée », « fonction », « défaut », « rapport », etc.

4.4 CONCLUSION

Les résultats obtenus grâce à ces deux études montrent bien que les approches *corpus-based* et *corpus-driven* ne sont pas opposées, mais complémentaires. Nous avons cette fois voulu montrer comment les régularités du corpus (souvent difficiles à détecter sans l'aide d'outils spécifiques) pouvaient servir de modèles pour des formes recommandées aux rédacteurs. Ces dernières peuvent entre autres consister en une série de *boilerplates*, c'est-à-dire de (morceaux de) phrases ou d'exigences types à paramétrer selon les besoins. L'idée n'est pas de remplacer entièrement le texte libre par des *boilerplates*, mais simplement d'encourager leur utilisation lorsque cela s'y prête : certains besoins très proches se répètent en effet souvent au sein d'une même spécification et il est alors avantageux de les exprimer par des exigences formellement semblables. L'intérêt des *boilerplates* est généralement souligné en ingénierie des exigences, sans qu'il soit précisé comment les constituer ; nous proposons donc une méthode outillée pour faciliter cette tâche.

Ces résultats ne sont évidemment pas « parfaits », puisqu'ils nécessitent (inévitablement) une révision humaine, idéalement suivie d'une validation par des experts du domaine. Les deux études avaient avant tout pour objet de démontrer la faisabilité d'une telle démarche. De fait, beaucoup d'améliorations seraient sans doute possibles, que ce soit au niveau de la méthode suivie ou des outils adoptés et de leurs réglages. Comme nous l'avons vu, des opérations et des paramètres différents produisent des résultats différents, plus ou moins intéressants et plus ou moins facilement exploitables. Nous voulions avant toute chose montrer ce qu'il est possible de faire en se servant de ce qui est à notre disposition (le corpus d'exigences[41] et d'autres ressources disponibles au CNES ou en dehors) et les pistes qui ont été dégagées nous semblent prometteuses.

Les approches basées sur l'analyse distributionnelle automatique (testées par Tutin (2007b) dans le cadre du lexique transdisciplinaire des écrits scientifiques, pour des résultats qu'elle ne dit que partiellement concluants) mériteraient également assurément que l'on s'y intéresse, notamment en vue de la détection des mots qui peuvent être utilisés comme attributs pour les *boilerplates*, pour laquelle elles présentent un intérêt évident. La sémantique distributionnelle (qui repose sur l'hypothèse distributionnelle, qui veut qu'une similarité de signification soit corrélée à une similarité de distribution) a en effet pour objet de calculer la similarité sémantique entre mots (notion très large, pour ne pas dire floue, puisqu'elle concerne des relations aussi variées que la synonymie, l'antonymie, l'hyperonymie, la méronymie, etc.) et ne nécessite théoriquement, pour cela, ni connaissances *a priori*, ni ressources autres que le corpus sur laquelle l'analyse est effectuée. Il est intéressant de noter, comme l'ont rappelé entre autres Fabre et Lenci (2015) et Fabre et al. (2014), que l'analyse distributionnelle harrissienne avait originellement pour vocation d'être appliquée à des (petits) corpus spécialisés, seul moyen de proposer

41. †Lui-même limité, rappelons-le. Disposer de plus de deux sous-corpus permettrait d'adopter des conditions plus souples : par exemple, on pourrait dire d'une structure qu'elle est récurrente si elle apparaît dans au moins deux projets sur trois, ou trois projets sur six, etc.

des catégories sémantiques bien délimitées.

Le principal problème auquel sont confrontés les chercheurs est la dispersion des données, qui explique logiquement que ceux-ci se soient tournés vers les corpus toujours plus grands qui sont mis à leur disposition ou qui peuvent être facilement constitués, et qui peuvent parfois allègrement dépasser le milliard de mots – à tel point que FABRE et al. (2014) et TANGUY, SAJOUS et HATHOUT (2015) disent de leur corpus spécialisé d'environ deux millions de mots qu'il est « modeste » ou « de taille réduite ». Ces derniers, comme d'autres, compensent le manque de mots par un enrichissement linguistique (en l'occurrence, grâce à l'analyse syntaxique en dépendance des phrases du corpus fournie par Talismane et représentée sous la forme de triplets) et montrent que des paramétrages différents retournent des résultats différents. D'autres auteurs se sont intéressés à des corpus spécialisés de très petite taille (d'un ordre de grandeur similaire au nôtre) : BERTELS et SPEELMAN (2016) obtiennent ainsi des résultats sur un corpus d'environ 360 000 mots (dans le domaine des machines-outils), au prix d'un paramétrage fin et de certaines concessions (utilisation des lemmes plutôt que des formes fléchies, ce qui n'est pas sans rappeler les choix que nous avons faits dans nos études). PÉRINET et HAMON (2015) vont encore plus loin dans l'abstraction des contextes en utilisant la relation d'hyperonymie pour représenter, par exemple, « chaise », « fauteuil » et « tabouret » par « siège » dans les contextes distributionnels. Enfin, en lien direct avec les exigences et les spécifications techniques, MERDY (2016) met à profit l'analyse distributionnelle automatique pour identifier l'ambiguïté et en particulier les termes flous ou sous-spécifiés (MERDY, KANG et TANGUY, 2016).

Dans tous les cas, il ne faut pas oublier que, même dans l'approche *corpus-driven* la plus stricte, si les outils fournissent des résultats, c'est bien au chercheur ou à l'utilisateur qu'il incombe de décider comment il les interprète et les exploite.

5

VALIDATION DES PROPOSITIONS : ENQUÊTE

5.1 INTRODUCTION

Afin d'évaluer plus objectivement la pertinence de certaines des règles linguistiques proposées de façon récurrente par les langues contrôlées, nous avons souhaité, par le biais d'une enquête en ligne, recueillir l'appréciation de locuteurs du français (quel que soit leur niveau d'expertise dans le domaine spatial et la rédaction technique) sur plusieurs formulations possibles d'une même exigence, certaines tenant compte des recommandations et d'autres non (Warnier et Condamines, 2017a) [1]. Notre objectif principal, au travers de cette expérimentation (qui est à rapprocher de ce que Ganier (2002) appelle « recueil de données subjectives » [2], parmi les méthodes d'évaluation des documents techniques centrées sur l'utilisateur), était de déterminer si les exigences rédigées en respectant les règles des langues contrôlées étaient systématiquement et significativement préférées (c'est-à-dire jugées comme « meilleures » et plus claires) à des exigences qui expriment le même besoin, mais qui sont formulées de façon différente (en s'écartant par conséquent des principes prônés par ces mêmes langues contrôlées). Nous nous sommes concentré sur quelques phénomènes parmi ceux présentés au chapitre 3, mais la méthode est facilement reproductible et pourrait donc être appliquée à d'autres phénomènes, ainsi qu'aux suggestions émergeant du corpus, notamment en vue de proposer des *boilerplates* (voir chapitre 4). Nous avons également cherché à savoir si les avis des sujets catégorisés comme « experts » de la rédaction technique – possiblement influencés à la fois par les règles en question et par l'expérience qu'ils ont acquise – divergent significativement sur ce point de ceux des locuteurs qui n'ont aucune ou qui n'ont que très peu d'expérience de la lecture et de la rédaction de spécifications [3].

1. [↑]Nous tenons à remercier très sincèrement les personnes qui nous ont aimablement aidé dans la réalisation de cette enquête : Pierre-Vincent Paubel (ingénieur en informatique dans le laboratoire CLLE) pour ses conseils et son assistance technique sur Qualtrics ; Aurélie Mouneyrac (doctorante en psychologie dans l'équipe CLLE-LTC) pour nous avoir guidé dans le choix des tests statistiques à effectuer afin d'exploiter les résultats obtenus et pour ses explications ; Daniel Galarreta, Jean-François Gory et Nicolas Deslandres (CNES) pour nous avoir apporté leur indispensable expertise dans la validation des questions et pour avoir bien voulu assurer la promotion de l'enquête auprès de leurs collègues ; les nombreux répondants anonymes qui ont consacré un peu de leur temps pour notre travail à titre gracieux.

2. [↑]À la différence que l'auteur n'envisage que l'évaluation du document produit (et pas des variantes possibles), donc sans la dimension comparative qui nous intéresse fondamentalement ici.

3. [↑]Évidemment, de nombreuses autres comparaisons seraient possibles (et sans doute pertinentes) ; par exemple, entre spécialistes et non-spécialistes du domaine spatial – la connaissance du domaine technique facilitant probablement la compréhension.

5.2 DISPOSITIF

Après avoir comparé plusieurs plateformes de sondage en ligne (SurveyMonkey, Google Forms) et réalisé un prototype avec le logiciel libre LimeSurvey, nous avons finalement retenu la solution commerciale Qualtrics pour implémenter notre questionnaire. En effet, celle-ci nous offrait davantage d'options de personnalisation que ses concurrentes et s'impose comme une référence pour les enquêtes réalisées dans un cadre académique ou professionnel. Notre enquête, dans sa version finale, prend la forme d'un questionnaire sur Internet[4] composé de plusieurs pages successives. Il est nécessaire que l'utilisateur complète et valide le questionnaire dans son intégralité pour que les résultats puissent être exploités correctement.

Avant toute chose, il est demandé à l'utilisateur de donner son accord, notamment pour que les données recueillies (anonymement) puissent faire l'objet d'une publication scientifique, en cliquant sur un bouton au libellé explicite (voir l'annexe C pour le texte de consentement libre et éclairé). Une fois son consentement recueilli, l'utilisateur est invité à lire un texte (également reproduit en annexe C) qui lui précise le contexte de l'étude menée et la nature des énoncés qu'il va avoir à évaluer, c'est-à-dire des exigences. Cette explication liminaire étant destinée avant tout aux locuteurs non familiers des spécifications – les experts n'ayant évidemment pas besoin de ces précisions –, elle se veut la plus simple et la plus générale possible, se contentant de fournir à l'utilisateur les éléments de contexte dont il aura besoin, sans l'orienter vers un type d'interprétation en particulier[5] ; elle insiste sur le fait que les exigences servent à exprimer un besoin, qu'elles revêtent nécessairement un caractère *obligatoire* et qu'il importe donc qu'elles soient correctement interprétées par toutes les parties prenantes.

Les consignes de la tâche qui va suivre sont ensuite données : pour chacune des vingt questions (une par page), entre deux et six formulations différentes (correspondant à autant de variantes possibles d'une exigence dont le contenu reste censément le même) sont affichées et il est demandé à l'utilisateur d'attribuer une note entre 1 et 5 inclus – 5 étant la meilleure note, à décerner aux formulations qui paraissent les plus adaptées (c'est-à-dire les plus claires et les moins ambiguës) et 1 la moins bonne note, à attribuer aux formulations à éviter absolument. Pour ce faire, l'utilisateur est invité à s'imaginer être à la place de la personne en charge de lire ces exigences et de les respecter.

Il est en outre clairement demandé à l'utilisateur d'attribuer la note maximale (5) à *au moins une* des propositions, celle(s) qu'il juge la (les) meilleure(s) (ou, à défaut, la (les) moins mauvaise(s)) sur la page affichée[6], et d'attribuer les autres

4. ↑Accessible à l'adresse : https://toulousepsychology.eu.qualtrics.com/SE/?SID=SV_3wpnmNkQbMDpIm9.

5. ↑C'est pourquoi nous avons renoncé à fournir à ce stade un exemple d'exigence, de peur que celui-ci ne serve de parangon auquel l'utilisateur risquerait de vouloir se référer (consciemment ou non).

6. ↑Dans l'ensemble, cette consigne a été bien respectée : seuls quelques utilisateurs (experts) se sont refusés à attribuer la note maximale pour certaines questions, estimant qu'aucune proposition

notes de façon relative (4 signifiant par exemple « un peu moins bon », et 2 signifiant « beaucoup moins bon »). La raison pour laquelle nous avons imposé cette contrainte est que nous redoutions que les experts de la rédaction d'exigences, que nous savions particulièrement critiques vis-à-vis de ces dernières (comme nous avons pu le constater lorsque nous avons fait valider nos propositions par certains d'entre eux), ne soient tentés de n'attribuer que des notes très faibles (voire minimales) à chacune des propositions, parce qu'ils auraient considéré – pour des raisons tout à fait valables, mais qui ne nous préoccupent pas ici et auraient interféré avec notre expérimentation – qu'aucune d'entre elles n'était acceptable (par exemple parce qu'elles relèvent davantage du commentaire que de l'exigence ou parce qu'un terme est mal employé). Le cas échéant, les résultats obtenus auraient été biaisés (notes trop faibles), voire inutilisables (notes toutes égales à 1), car une comparaison des notes n'aurait plus été possible : en effet, rappelons que notre objectif est de déterminer quelle formulation est la plus appréciée, ce qui implique nécessairement de comparer les scores obtenus par chacune d'entre elles (autrement dit, la note attribuée à une proposition n'est pas intéressante *dans l'absolu*, mais seulement de façon *relative*, par rapport aux notes attribuées aux autres propositions).

Nous avons préféré un système de notation à un système plus simple de classement (mettre en 1re position la formulation jugée la plus acceptable, en 2e position la formulation jugée un peu moins acceptable, et ainsi de suite) pour deux raisons : d'une part parce que cela permet de mesurer l'écart de préférence entre deux propositions (par exemple, une proposition avec la note 4 est jugée un peu moins acceptable qu'une proposition avec la note 5, mais beaucoup plus acceptable qu'une proposition notée 1) et, d'autre part, parce que cela permet de traiter beaucoup plus simplement les cas d'*ex aequo*, possibilité qui nous a paru indispensable (deux propositions doivent pouvoir obtenir le même score si l'utilisateur estime qu'elles se valent).

Une fois qu'il a pris connaissance de ces informations, l'utilisateur peut accéder aux questions portant sur les exigences, c'est-à-dire à la principale partie du questionnaire (voir la FIGURE 16 pour un exemple). Chaque question s'affiche sur une page distincte : la consigne est d'abord rappelée succinctement, puis les différentes variantes sont affichées les unes en dessous des autres (et à côté de chacune d'elles se trouvent cinq boutons correspondant aux cinq notes possibles) ; seule la formulation linguistique étant évaluée ici, nous n'avons pas cherché à reproduire la mise en page d'une spécification du CNES. Aucun indice visuel n'est donné sur les différences entre propositions : c'est à l'utilisateur de les repérer. Pour pouvoir accéder à la page suivante, il est obligé d'attribuer une note à chaque proposition.

n'était acceptable (mais toujours pour des raisons indépendantes du phénomène linguistique que nous visions à travers ces questions) ; cela reste sans conséquence sur la validité et l'intérêt des résultats, l'influence sur la moyenne étant minime. Par ailleurs, il est évident que leur décision n'est pas sans intérêt pour nous, puisqu'elle met en évidence un problème dans la formulation des exigences et dans la façon dont celles-ci sont perçues – mais ce problème se situe vraisemblablement à un autre niveau que celui sur lequel nous nous focalisons dans le questionnaire.

Facultativement, il a la possibilité d'entrer un commentaire libre dans un champ de texte situé en bas de la page ; en l'invitant à indiquer les raisons qui justifient le choix de ses notes, nous espérons obtenir des résultats plus qualitatifs, en plus des résultats quantitatifs fournis par les données chiffrées.

FIGURE 16 – Exemple de question posée dans l'enquête en ligne

- Commencez par attribuer la note "5" à la (ou les) formulation(s) qui vous semble(nt) préférable(s).

- Attribuez ensuite une note sur 5 à chacune des autres formulations (par exemple, 4 = un peu moins bien et 1 = beaucoup moins bien).

	1	2	3	4	5
La configuration doit pouvoir être modifiée par deux événements.	○	○	○	○	○
La configuration est modifiée par deux événements.	○	○	○	○	○
La configuration devra pouvoir être modifiée par deux événements.	○	○	○	○	○
La configuration peut être modifiée par deux événements.	○	○	○	○	○
La configuration sera modifiée par deux événements.	○	○	○	○	○
La configuration pourra être modifiée par deux événements.	○	○	○	○	○

Afin d'éviter les biais dus à l'effet d'ordre (PERREAULT, 1975 ; MARSDEN et WRIGHT, 2010), les questions, de même que les propositions pour chaque question, apparaissent dans un ordre aléatoire (donc théoriquement à chaque fois différent)[7] : si effet d'ordre il y a, il devrait par conséquent être compensé par le grand nombre de participants. Par ailleurs, sans que cela ne soit indiqué, le temps exact passé par l'utilisateur sur chaque page est enregistré (et peut éventuellement servir d'indicateur de la complexité de la question posée), mais il n'est pas limité (l'utilisateur peut passer autant de temps qu'il le souhaite sur une question).

Enfin, une série de questions visant à mieux cerner le profil du participant lui sont posées : outre son âge et son sexe, il lui est demandé s'il travaille ou a travaillé au CNES, s'il a déjà travaillé à la réalisation de systèmes spatiaux, s'il savait déjà

7. ↑ À l'exception d'une question (apparaissant toujours en dixième position), où nous avons intentionnellement inséré un piège : deux formulations identiques sont proposées (respectivement en deuxième et cinquième positions). Nous souhaitions ainsi nous assurer que les participants respectaient bien la consigne (autrement dit, qu'ils attribuent la même note dans les deux cas). À l'analyse des résultats, il s'avère que de nombreux participants ont attribué deux notes légèrement différentes (une différence d'un point pour treize d'entre eux et de deux points pour trois autres) – ce qui illustre bien que l'effet d'ordre est inévitable. Néanmoins, la différence finale est infime (la moyenne de l'une est de 2,83 et celle de l'autre de 2,84 sur 5), ce qui nous conforte dans l'idée que les résultats sont malgré tout fiables.

ce qu'est une exigence technique, s'il est habitué à lire des exigences techniques, s'il est habitué à en écrire, s'il a été formé à la rédaction des exigences, s'il a déjà travaillé sur des exigences Pléiades (qui ont servi de modèles pour nos questions, cf. *infra*), s'il a été formé en linguistique, et si le français est sa langue maternelle. Dans la plupart des cas, si la réponse à l'une de ces questions est oui, il est demandé d'apporter un bref complément d'information (dans quel contexte, depuis combien de temps, etc.). Pour finir, le participant a la possibilité, s'il le souhaite, de formuler ses dernières remarques (sur l'enquête en général) et d'indiquer une adresse de courrier électronique via laquelle il pourra être éventuellement recontacté (les données n'étant alors évidemment plus anonymes).

Étant donné que nous visions deux populations différentes (des experts, habitués à lire des exigences, et des locuteurs « naïfs » dans ce domaine), le lien menant au questionnaire a été diffusé par deux canaux distincts : par courrier électronique (à une liste de 135 destinataires) en interne au CNES d'une part, et sur les réseaux sociaux d'autre part. Nous n'avons pas cherché à savoir par quelle voie les participants ont accédé au questionnaire, mais cette information se déduit aisément de leur réponse à la question leur demandant s'ils travaillent ou ont travaillé au CNES.

5.3 CHOIX DES QUESTIONS

L'étape la plus cruciale dans l'élaboration de ce questionnaire est celle qui a consisté à déterminer les phénomènes à étudier et donc les questions à y inclure. Comme précisé en introduction, nous avons privilégié des phénomènes récurrents dans les langues contrôlées ou dans les guides de bonnes pratiques relatives à l'ingénierie des exigences, à savoir :

— la *modalisation* et les *temps verbaux* (l'injonction doit-elle être marquée par le verbe « devoir » ? le verbe doit-il être conjugué au présent de l'indicatif ou au futur ?),
— le *nombre de phrases* (vaut-il mieux deux propositions coordonnées dans une même phrase, ou deux phrases indépendantes ?),
— l'*anaphore* (faut-il systématiquement répéter les noms en entier, ou est-il parfois préférable d'utiliser par exemple un pronom ?),
— et la question de la diathèse et de l'*agent* (la voix active doit-elle toujours être préférée à la voix passive ? qu'en est-il de « on » ?).

Les vingt questions visent donc toutes à tester l'un de ces quatre phénomènes (nous avons privilégié un petit nombre de phénomènes, ce qui nous permet de tester plusieurs fois chacun d'eux) ; vingt questions étant la limite au-delà de laquelle nous avons considéré que le questionnaire serait trop long, et donc probablement décourageant[8]. Naturellement, il n'est pas précisé au participant quel phénomène

8. [↑]Le temps de réponse moyen est de 32 min 22 s, mais il tombe à 24 min 29 s si l'on exclut les neuf temps de réponse supérieurs à une heure (qui nous paraissent particulièrement élevés et peu

est testé par la question à laquelle il lui est demandé de répondre – bien que celui-ci puisse sans doute se déduire sans trop de difficulté.

Afin que les exigences proposées paraissent aussi crédibles que possible, nous avons soigneusement sélectionné des phrases adéquates directement dans le corpus Pléiades : ces exigences ont donc été rédigées en contexte réel. Néanmoins, nous avons souvent été amené à les adapter (en les raccourcissant ou en les simplifiant [9]), dans l'optique non seulement de simplifier la tâche des participants, mais également de nous assurer que nous ne testions qu'un seul phénomène à la fois. Pour chacune des exigences retenues et adaptées, il nous a aussi fallu produire un certain nombre de variantes pour tester le phénomène envisagé (ainsi, par exemple, si le verbe était conjugué au futur, nous avons fabriqué une formulation identique, mais dans laquelle le verbe est conjugué au présent [10]). Les formulations ainsi obtenues peuvent dès lors être qualifiées de « semi-authentiques », et toutes ont été vérifiées par des experts du CNES, ce qui nous a souvent conduit à les adapter à nouveau. Bien entendu, toutes sont grammaticalement correctes.

Nous présentons ci-après les différentes questions que comporte le questionnaire, classées selon le phénomène testé (l'ordre choisi pour les présenter ici est arbitraire) :

5.3.1 *Modalisation*

Nous essayons ici de savoir comment se marque le mieux le caractère injonctif des exigences. Pour certaines questions, nous proposons six formulations différentes : (1) verbe principal conjugué au présent ; (2) « peut » [11] + infinitif ; (3) « doit pouvoir » + infinitif ; (4) verbe principal conjugué au futur ; (5) « pourra » + infinitif ; (6) « devra pouvoir » + infinitif. Ces exigences précisent qu'une fonctionnalité doit être disponible ou qu'un choix peut être fait ; elles impliquent qu'une possibilité doit exister.

I. L'opérateur a le choix du mode de saisie : il *rentre / peut rentrer / doit pouvoir rentrer / rentrera / pourra rentrer / devra pouvoir rentrer* l'adresse directement

représentatifs – certains approchant même les trois heures –, l'utilisateur ayant pu interrompre sa tâche pendant un long moment avant de la reprendre).

9. ↑Ainsi, dans de nombreux cas, nous n'avons conservé qu'un morceau de l'exigence concernée. En revanche, nous avons pris le parti d'y laisser un certain nombre des sigles et acronymes rencontrés (même si ceux-ci sont très probablement inconnus des participants extérieurs au CNES), car nous pensons que leur présence n'entrave pas la compréhension du sens global de la phrase et qu'ils obligent le lecteur à se concentrer davantage sur les structures syntaxiques. En outre, ils rendent probablement l'exigence plus authentique aux yeux des experts. Les versions originales des exigences que nous avons utilisées sont données en annexe D. Exemple d'exigence dans notre corpus : « [...] Le plan TC devra respecter les contraintes décrites dans [DR20] », rendue plus explicite dans le questionnaire sous la forme : « Le plan de télécommande devra respecter les contraintes décrites dans le document DR20 ».

10. ↑Dans les faits, nous avons aussi bien produit une variante au présent sur la base d'une exigence au futur que l'inverse.

11. ↑Nous avons privilégié les verbes au singulier, bien plus nombreux que les formes correspondantes au pluriel dans le corpus.

ou indirectement en précisant le nom d'un label.
(Note : nous n'avons fait varier le verbe que dans la seconde proposition.)

II. La configuration *est / peut être / doit pouvoir être / sera / pourra être / devra pouvoir être* modifiée par deux événements.

Pour d'autres questions, le nombre de possibilités se réduit à quatre : (1) verbe principal conjugué au présent ; (2) « doit » + infinitif ; (3) verbe principal conjugué au futur ; (4) « devra » + infinitif. La différence avec la sous-catégorie précédente est donc que le verbe « pouvoir » n'apparaît plus.

III. L'ensemble des TC *est / doit être / sera / devra être* saisi en base de données.

IV. Le DUPC *combine / doit combiner / combinera / devra combiner* en une seule séquence de lecture l'ensemble des ordres de programmation des trois canaux.

V. Sur ordre du système, le LVC *déroute / doit dérouter / déroutera / devra dérouter* l'exécution du MDP principal vers le MDP secondaire.

VI. Le plan de télécommande *respecte / doit respecter / respectera / devra respecter* les contraintes décrites dans le document DR20.

Nous partons du principe que, dans ce contexte particulier, pour chaque question, toutes les formulations proposées sont pragmatiquement équivalentes, dans la mesure où elles signifient la même chose et doivent avoir le même effet – rappelons à nouveau qu'une exigence est par nature obligatoire. Cependant, nous supposons que les lecteurs pourraient avoir une préférence soit pour celles où l'injonction est marquée très explicitement par le verbe « devoir », soit, au contraire, pour celles où ce modalisateur, jugé superflu, est absent (parce qu'elles seraient plus courtes et plus faciles à traiter cognitivement, au risque de paraître plus descriptives que normatives). De même, nous aimerions savoir s'ils préfèrent que l'exigence soit rédigée au présent (avec un effet d'immédiateté et donc d'urgence plus grand) ou bien au futur (le système tel qu'il est décrit n'existant pas encore au moment de la rédaction).

5.3.2 *Nombre de phrases*

Ce que nous cherchons à déterminer ici, c'est si les utilisateurs préfèrent qu'une exigence complexe soit rédigée sous la forme d'une seule phrase, plus longue (avec éventuellement plusieurs propositions coordonnées) ou plutôt qu'elle soit scindée en deux phrases distinctes (le plus souvent sans coordination). Les questions qui suivent proposent donc un choix binaire : (1) une phrase ; (2) deux phrases [12].

I. L'automate de la fonction IOS passe à ON et l'ensemble des traitements sur les équipements senseur solaire, magnétomètres, magnéto-coupleurs, roue à

12. [↑]La balise HTML <br> se traduit par un retour à la ligne sur la page web ; les deux phrases apparaissent donc l'une en dessous de l'autre.

réaction est séquencé. / L'automate de la fonction IOS passe à ON.
L'ensemble des traitements sur les équipements senseur solaire, magnétomètres, magnéto-coupleurs, roue à réaction est séquencé.

II. Il est possible d'importer et d'exporter toutes les règles de transfert déclarées. / Il est possible d'importer toutes les règles de transfert déclarées.
Il est possible d'exporter toutes les règles de transfert déclarées.

III. Les champs SM_ID et FM_ID seront extraits à partir de la BDS. / Le champ SM_ID sera extrait à partir de la BDS.
Le champ FM_ID sera extrait à partir de la BDS.

IV. Si le transfert d'un fichier de télémesure échoue après le nombre de tentatives fixées, le fichier sur le destinataire est effacé et l'information est remontée aux IHM. / Si le transfert d'un fichier de télémesure échoue après le nombre de tentatives fixées, le fichier sur le destinataire est effacé.
L'information est remontée aux IHM.

On peut par ailleurs s'interroger sur l'opportunité de scinder une exigence complexe en plusieurs exigences minimales (en vertu du principe d'*atomicité* des exigences), auquel cas le choix deviendrait ternaire : une seule exigence d'une phrase (cas 1), une seule exigence composée de deux phrases (cas 2), ou deux exigences d'une phrase chacune (nouveau cas). Toutefois, cette troisième option aurait demandé des modifications trop importantes (par exemple, dans la dernière question, le segment « L'information est remontée aux IHM » ne peut pas être présenté isolément sans paraître incomplet), qui auraient inutilement compliqué l'expérimentation et rendu le tout plus artificiel. En outre, ce choix aurait sans doute été trop complexe à appréhender pour les utilisateurs non familiers des spécifications et de l'ingénierie des exigences. Nous avons par conséquent préféré éviter de telles considérations à ce stade.

5.3.3 *Anaphore*

À travers les questions suivantes, nous avons cherché à savoir s'il est préférable de répéter un nom lorsqu'il a déjà été mentionné dans la même exigence (et si oui, avec quel déterminant et sous quelle forme) ou si l'usage des pronoms (plus naturel) est mieux indiqué. La plupart admettaient quatre possibilités (anaphore fidèle définie, anaphores fidèles démonstratives (Le Pesant, 2002) et pronominalisation) : (1) déterminant défini + nom (seulement la tête dans les deux premières questions, en entier dans les deux dernières) ; (2) déterminant démonstratif + nom (seulement la tête) ; (3) déterminant démonstratif + nom (groupe en entier) ; (4) pronom personnel.

I. Le paquet cyclique ne sera généré que si *le paquet / ce paquet / ce paquet cyclique / s'il* est activé par le LVC.

II. La configuration du système est stockée à bord. *La configuration / Cette configuration / Cette configuration du système / Elle* ne peut être modifiée que par deux événements : événement 1 et événement 2.

III. En cas de réception d'un fichier vide, *le fichier vide / ce fichier / ce fichier vide / il* n'est pas transmis.

IV. La liste des TCD est définie dans le document C12 ; *la liste des TCD / cette liste / cette liste des TCD / elle* est exhaustive.

Pour l'une des questions, seules trois possibilités existent : (1) déterminant défini + nom ; (2) déterminant démonstratif + nom ; (3) pronom.

V. L'utilisateur doit savoir à tout moment si *l'utilisateur* / si *cet utilisateur* / s'*il* est connecté sur le serveur GIDE nominal ou redondant.

Évidemment, nous n'avons proposé des formulations contenant un pronom que si l'antécédent de ce dernier apparaissait clairement (sans quoi l'exigence deviendrait ambiguë et la formulation serait nécessairement à éviter).

5.3.4 *Agent*

Enfin, ces dernières questions permettent de recueillir l'appréciation des utilisateurs sur des formulations dans lesquelles l'agent est exprimé de différentes façons et sur celles dans lesquelles il est omis. Deux questions laissaient le choix entre quatre possibilités : (1) voix passive avec complément d'agent ; (2) voix passive sans complément d'agent ; (3) voix active ; (4) pronom « on ».

I. L'opération n'est nécessaire que si *la surveillance fonctionnelle est autorisée par le LVC / la surveillance fonctionnelle est autorisée / le LVC autorise la surveillance fonctionnelle / on autorise la surveillance fonctionnelle.*

II. *Le vidage des tables sera contrôlé par le CCC. / Le vidage des tables sera contrôlé. / Le CCC contrôlera le vidage des tables. / On contrôlera le vidage des tables.*

Pour les trois dernières questions, l'agent n'étant pas clairement précisé dans les exigences originales, nous n'avons proposé que deux options : (1) voix passive sans complément d'agent (ou formulation semblable) ; (2) pronom « on ».

III. *La durée d'observation imposée est comprise entre 20 et 100 secondes. / On impose une durée d'observation comprise entre 20 et 100 secondes.*
(Note : la véritable passivation aurait donné « Une durée d'observation comprise entre 20 et 100 secondes est imposée », mais nous avons jugée qu'elle paraissait très peu naturelle et préféré une formule attributive proche.)

IV. *Les TC associées aux familles F_EMETTEUR et F_MANOEUVRES sont listées ici. / On liste ici les TC associées aux familles F_EMETTEUR et F_MANOEUVRES.*

V. *La TC est refusée si la taille du paquet programmé dépasse la taille maximum. / On refuse la TC si la taille du paquet programmé dépasse la taille maximum.*

5.4 RÉSULTATS

5.4.1 *Présentation générale*

Au total, 87 personnes (50 femmes et 37 hommes) ont complété intégralement le questionnaire entre le 18 mai et le 9 août 2016. L'utilisateur le plus jeune avait 18 ans (âge minimum pour participer) et le plus âgé avait 63 ans au moment de répondre ; l'âge moyen était de 38,6 ans. Seules cinq personnes ont déclaré que le français n'était pas leur langue maternelle. Plusieurs questions permettaient d'en apprendre davantage sur leur profil ; les réponses sont synthétisées dans le TABLEAU 14 [13].

TABLEAU 14 – Profil des utilisateurs ayant répondu à l'enquête

	oui	non
Avez-vous déjà travaillé ou travaillez-vous au CNES ?	41 (47,13 %)	46 (52,87 %)
Avez-vous déjà travaillé ou travaillez-vous à la réalisation de systèmes spatiaux ?	36 (41,38 %)	51 (58,62 %)
Avant d'ouvrir ce questionnaire, saviez-vous déjà ce qu'est une exigence technique ?	60 (68,97 %)	27 (31,03 %)
Êtes-vous habitué à *lire* des exigences techniques ?	48 (55,17 %)	39 (44,83 %)
Êtes-vous habitué à *écrire* des exigences techniques ?	38 (43,68 %)	49 (56,32 %)
Avez-vous déjà été formé à la rédaction des exigences ?	18 (20,69 %)	69 (79,31 %)
Avez-vous déjà travaillé sur des exigences Pléiades ?	8 (9,20 %)	79 (90,80 %)
Avez-vous été formé en linguistique ?	20 (22,99 %)	67 (77,01 %)

Ainsi, on constate qu'il y a presque autant de réponses provenant d'employés du CNES que de personnes extérieures, bien qu'ils ne soient que 36 à avoir déjà travaillé directement à la réalisation de systèmes spatiaux. Huit personnes seulement ont déclaré avoir déjà travaillé sur des exigences Pléiades (dont sont tirés nos exemples). Plus de deux tiers des répondants ont déclaré savoir ce qu'est une exigence technique (une proportion plus élevée que ce à quoi nous nous attendions), mais seulement 48 disent avoir l'habitude d'en lire (ce qui représente tout de même plus de la moitié des réponses), 38 l'habitude d'en écrire et 18 avoir été formés à la rédaction.

13. †Pour les besoins de l'analyse statistique, nous avons ici réduit les réponses à un simple « oui » ou « non », mais elles contiennent en réalité parfois davantage d'informations : fonction exercée, formation suivie, etc.

Il est intéressant de remarquer que le nombre de personnes habituées à lire des exigences est plus élevé que le nombre d'employés du CNES ayant répondu à l'enquête, ce qui signifie que certaines d'entre elles y ont été confrontées dans une autre entreprise, voire dans un autre domaine que le spatial (l'informatique, par exemple). C'est la réponse à cette question (habitude de lire des exigences techniques) que nous avons retenue pour décider de catégoriser l'utilisateur comme « expert » ou non. En effet, bien que ce choix soit certainement discutable pour plusieurs raisons (en particulier parce que la compétence réelle n'est pas testée et parce que la distinction n'est pas fine), il reste un moyen simple d'obtenir deux populations comparables de tailles relativement proches (48 experts / 39 non experts). Nous estimons en effet que lire des exigences techniques de façon régulière depuis suffisamment longtemps apporte nécessairement une certaine compétence, puisqu'il est nécessaire de les interpréter correctement ; à tout le moins, cela signifie que ces utilisateurs sont familiers de ce type d'écrits, au contraire de ceux qui n'en ont jamais (ou rarement) lu, et qu'ils sont donc dans une position privilégiée pour répondre à la tâche demandée dans cette enquête, à savoir se placer du point de vue de l'ingénieur chargé de lire et de respecter les exigences. Ces utilisateurs, désormais appelés experts, ont mis en moyenne un peu plus de temps pour répondre au questionnaire : 25 min 10 s, contre 23 min 45 s pour les non experts (toujours en excluant les temps de réponse supérieurs à une heure).

5.4.2 *Choix des analyses*

Nous avons cherché à savoir si des tendances se dégageaient des résultats obtenus grâce à cette enquête. Pour ce faire, nous pouvons nous baser, d'une part, sur les notes attribuées de façon relative à chaque proposition et, d'autre part, sur les commentaires laissés à propos de certaines questions (au total, ceux-ci représentent plus de 7 000 mots). Les premières constituent donc des données chiffrées qu'il convient d'analyser et d'interpréter à l'aide de tests statistiques – lesquels nous ont été recommandés par une psychologue formée en statistique – afin de proposer des conclusions générales, tandis que les seconds doivent être dépouillés manuellement, mais peuvent apporter des distinctions plus fines et surtout éclairer ou justifier le choix de ces mêmes notes. Nous nous appuierons donc sur certains de ces commentaires pour tenter d'expliquer les résultats des analyses statistiques.

Aussi avons-nous voulu déterminer si, pour chaque question :

(a) la distribution des réponses est identique ou si, au contraire, certaines propositions sont significativement préférées (et donc s'il y a un ordre permettant de les classer). Nous nous appuyons pour cela sur les résultats du test non paramétrique de Friedman. Les résultats sont significatifs pour toutes les questions, sauf une [14] ;

14. [†]Celle qui demande de comparer « L'automate de la fonction IOS passe à ON et l'ensemble des traitements sur les équipements senseur solaire, magnétomètres, magnéto-coupleurs, roue à réaction est séquencé. » avec sa variante en deux phrases.

(b) la distribution des réponses pour une même proposition est identique pour les deux groupes (experts / non experts) ou si, au contraire, l'un des deux groupes lui attribue une note significativement plus basse ou plus élevée que l'autre. Nous nous appuyons pour cela sur les résultats du test non paramétrique de Mann-Whitney [15].

Cette distinction experts *vs* non experts nous semble intéressante pour l'exploitation de nos résultats. En effet, une étude précédente (De Jong et Lentz, 1996) a montré, pour des brochures d'information publique sur le droit aux allocations de logement, que les jugements des experts et les retours des utilisateurs à qui elles sont destinées pouvaient fortement diverger. Dans notre cas, c'est-à-dire dans une communication d'experts (rédacteurs) à experts (lecteurs), nous nous attendons également à relever des différences de jugement, dues au contexte particulier de la communication technique : les experts peuvent avoir une appréciation sur la meilleure façon de transmettre une information dans le contexte particulier de la spécification technique, compte tenu à la fois de leur expérience et de certaines croyances qu'ils ont pu acquérir sous l'influence des règles des langues contrôlées, là où les non experts devraient logiquement se référer davantage (voire exclusivement) à leur connaissance du fonctionnement de la langue générale, et donc constituer un groupe « neutre ». Cette opposition justifie dès lors que nous nous intéressions aux deux groupes séparément, même si, *in fine*, ce sont bien les experts qui sont avant tout visés.

En outre, nous avons également regroupé les moyennes des notes des questions appartenant à une même catégorie (par exemple, une phrase *vs* deux phrases) pour détecter des tendances au niveau des phénomènes envisagés. Nous avons à nouveau effectué un test de Friedman, suivi d'un test de Wilcoxon afin de vérifier que la différence entre la variante avec la moyenne la plus élevée et les autres variantes est significative (autrement dit, qu'il y a non seulement un ordre, mais qu'en outre, l'écart entre le premier élément et les suivants est suffisant). C'est donc cette dernière étape qui nous permettra de tirer les conclusions les plus générales quant à la « meilleure » façon de rédiger une exigence (ou, du moins, celle qui semble être préférée par les lecteurs).

Par résultat statistiquement *significatif*, il faut entendre qu'il est raisonnablement très peu probable que celui-ci soit le fait du hasard (autrement dit, que si l'hypothèse nulle était vraie, il serait très peu probable d'obtenir ce résultat) ; conventionnellement, pour de nombreux tests, le seuil de significativité (valeur α) est fixé à

15. ↑La variable correspondant à la note attribuée pour chaque formulation est en réalité ordinale (et non numérique), la distance entre chaque item n'étant pas forcément toujours identique. Elle correspond cependant à une échelle de Likert, pour laquelle il est admis que la moyenne peut être calculée. Par prudence, nous nous en tiendrons aux tests non paramétriques (plus robustes), même si certains auteurs avancent que les tests paramétriques (plus puissants) peuvent aussi être utilisés (De Winter et Dodou, 2010 ; Sullivan et Artino, 2013).

0,05 [16] : le résultat obtenu, s'il est significatif [17], a donc moins de 5 % de chances d'être dû au hasard.

Les résultats détaillés sont donnés dans l'annexe E. Pour chaque formulation, nous y précisons combien de personnes lui ont attribué la note maximale de 5/5 (pour rappel, la même note peut être attribuée à plusieurs propositions différentes), la moyenne générale obtenue auprès de tous les répondants, la moyenne des notes données par les experts et la moyenne des notes données par les non experts. Pour ces deux dernières moyennes, nous indiquons également si la différence est significative entre les deux groupes. Pour chaque question, les propositions sont classées par moyenne générale décroissante : celles qui ont été plébiscitées apparaissent donc en premier.

Nous adopterons ici une version beaucoup plus synthétique : pour chaque question, nous préciserons quelle variante serait préconisée par la plupart des langues contrôlées (marquée par L), laquelle a été préférée par les experts (marquée par E), laquelle a été préférée par les non experts (marquée par N) et laquelle est la plus proche de ce qui a été trouvé en corpus (marquée par C), donc celle qui a servi d'inspiration pour la question (les autres ayant été forgées d'après elle). Nous proposerons ensuite une synthèse pour chaque sous-catégorie (c'est-à-dire pour chaque groupe de questions proposant des formulations semblables pour un même phénomène).

5.4.3 *Modalisation*

I. L'opérateur a le choix du mode de saisie : il rentre / peut rentrer (N) / doit pouvoir rentrer (L-E) / rentrera / pourra rentrer (C) / devra pouvoir rentrer l'adresse directement ou indirectement en précisant le nom d'un label.

Langues contrôlées : « doit pouvoir rentrer »
Experts : « doit pouvoir rentrer » (4,17)
Non experts : « peut rentrer » (3,92)
Corpus : « pourra rentrer »

Ces variantes se distinguent principalement par le nombre de verbes utilisés : un, deux ou trois, le premier pouvant à chaque fois être conjugué au présent ou au futur. Il n'y a rien d'étonnant à ce que les formes les plus courtes, qui ne contiennent pas le verbe « pouvoir », ne soient pas retenues : elles n'indiquent pas qu'une possibilité doit exister et sont donc beaucoup moins claires, voire trompeuses. Les langues contrôlées recommandent généralement d'utiliser une forme de « devoir » dans la phrase principale de l'exigence (quitte à ce que la phrase soit

16. ↑Pour les tests de Wilcoxon, nous avons appliqué la correction de Bonferroni, qui nécessite d'abaisser ce seuil en le divisant par le nombre de comparaisons effectuées.

17. ↑S'il ne l'est pas, la prudence recommande de ne rien en conclure ; la certitude n'est pas assez grande pour rejeter l'hypothèse nulle (autrement dit, il peut ou non exister un effet, mais rien ne permet de l'affirmer).

plus longue) pour expliciter l'injonction ; les experts rejoignent clairement cette recommandation, en privilégiant le présent : il y a donc ici convergence. En revanche, les non experts optent pour la variante sans « devoir », sans doute parce qu'elle leur paraît un peu moins surchargée et donc plus élégante. On ne trouvait du reste ni l'une ni l'autre dans le corpus, puisque c'est le verbe « pouvoir » conjugué au futur qui avait été utilisé par le rédacteur. Il est vrai cependant que le verbe au présent dans la première proposition pouvait ici orienter le choix des répondants vers les formulations au présent.

II. La configuration est / peut être (N-C) / doit pouvoir être (L-E) / sera / pourra être / devra pouvoir être modifiée par deux événements.

Langues contrôlées : « doit pouvoir être »
Experts : « doit pouvoir être » (4,17)
Non experts : « peut être » (3,79)
Corpus : « peut être »

La voix passive augmente encore le nombre de verbes consécutifs (allant jusqu'à quatre : « doit/devra pouvoir être modifiée »), mais les résultats restent tout à fait semblables à ceux de la question précédente : les experts continuent à préférer la variante avec « devoir » au présent, plus explicite, les non experts celle au présent, mais sans « devoir », qu'ils n'estiment pas indispensable en contexte. C'est d'ailleurs cette dernière que l'on retrouve dans le corpus – pourtant composé d'exigences écrites par des experts.

Synthèse intermédiaire

TABLEAU 15 – Synthèse : Modalisation

	Moyenne Générale	Moyenne Experts	Moyenne Non Experts	Différence significative
« doit pouvoir » + infinitif	3,69	4,17	3,10	oui
« peut » + infinitif	3,32	2,87	3,86	oui
« pourra » + infinitif	3,10	2,68	3,62	oui
« devra pouvoir » + infinitif	2,96	3,25	2,60	oui
verbe au présent	2,91	2,62	3,27	oui
verbe au futur	2,71	2,47	3,00	oui

Ainsi que le montre le TABLEAU 15, les experts et les non experts s'accordent sur l'utilisation du présent, mais leur préférence diverge significativement lorsqu'il s'agit d'utiliser « devoir » : les premiers préfèrent l'inclure (même si l'exigence devient inévitablement plus longue), les seconds l'omettre. Pour les experts, « devoir » est un véritable marqueur d'exigences ; il les distingue des commentaires ou des remarques. Il y a une différence significative entre la moyenne obtenue pour « doit pouvoir » + infinitif et toutes les autres, sauf « peut » + infinitif. Si l'on ne

prend en compte que les notes attribuées par les experts, « doit pouvoir » + infinitif est significativement préféré à toutes les autres variantes.

III. L'ensemble des TC est / doit être (L-E-N) / sera (C) / devra être saisi en base de données.

Langues contrôlées : « doit être »
Experts : « doit être » (4,38)
Non experts : « doit être » (4,15)
Corpus : « sera »

Dans cette exigence sans « pouvoir », les non experts rejoignent l'avis des experts (et donc aussi les langues contrôlées) en utilisant le modal « devoir » : il permet de marquer l'injonction, sans alourdir trop l'exigence. La convergence est donc très forte ; seul le rédacteur de l'exigence avait préféré une autre formulation, au futur (pourtant peu apprécié à travers l'enquête).

IV. Le DUPC combine (N) / doit combiner (L-E-C) / combinera / devra combiner en une seule séquence de lecture l'ensemble des ordres de programmation des trois canaux.

Langues contrôlées : « doit combiner »
Experts : « doit combiner » (4,40)
Non experts : « combine » (4,08)
Corpus : « doit combiner »

Les non experts préfèrent cette fois omettre « devoir », qui apparait toutefois dans l'exigence originale. La convergence reste donc forte.

V. Sur ordre du système, le LVC déroute (N-C) / doit dérouter (L-E) / déroutera / devra dérouter l'exécution du MDP principal vers le MDP secondaire.

Langues contrôlées : « doit dérouter »
Experts : « doit dérouter » (4,33)
Non experts : « déroute » (3,85)
Corpus : « déroute »

Les résultats pour cette question sont très similaires à ceux de la précédente, si ce n'est que l'exigence originale ne contenait pas le modal « devoir ».

VI. Le plan de télécommande respecte / doit respecter (L-E-N) / respectera / devra respecter (C) les contraintes décrites dans le document DR20.

Langues contrôlées : « doit respecter »
Experts : « doit respecter » (4,63)
Non experts : « doit respecter » (4,23)

Corpus : « devra respecter »

La convergence est ici presque parfaite : à part le rédacteur de l'exigence ayant servi de modèle, qui avait choisi « devra » (qui peut être rapproché du « shall » préconisé par certaines normes anglophones), tous (experts, non experts et langues contrôlées) privilégient « doit » + infinitif.

Synthèse

TABLEAU 16 – Synthèse : Modalisation (2)

	Moyenne Générale	Moyenne Experts	Moyenne Non Experts	Différence significative
« doit » + infinitif	4,19	4,43	3,88	oui
« devra » + infinitif	3,62	3,70	3,53	non
verbe au présent	3,30	2,97	3,70	oui
verbe au futur	3,17	2,96	3,42	oui

Ainsi que le révèle le TABLEAU 16, dans les exigences sans le modal « pouvoir », experts comme non experts sont d'accord pour utiliser préférentiellement « doit » + infinitif (la différence avec les trois autres propositions étant significative), ce qui correspond parfaitement aux recommandations de certaines langues contrôlées (ainsi qu'à la nôtre). Ces résultats apportent donc un argument en faveur de ces dernières et il y aurait sans doute tout intérêt à en tenir compte (surtout en sachant que dans le corpus existant, on retrouve toutes les variantes possibles).

Lorsque l'on s'intéresse aux commentaires laissés par les répondants, diverses tendances se distinguent. Concernant le choix du temps (présent ou futur) tout d'abord, plusieurs font remarquer que ce qui importe, c'est avant tout d'être cohérent (tous les verbes au présent ou tous au futur). Il est toutefois avancé que le présent augmente le sentiment d'injonction, le futur pouvant laisser croire que quelqu'un d'autre ou une autre partie du système s'en chargera plus tard. Un expert reconnaît néanmoins être influencé par l'utilisation de « shall » dans les exigences de l'ECSS et donc préférer le futur en français. Quant à l'injonction, les experts semblent unanimes sur le fait qu'elle doit être marquée par le modal « devoir » ; les phrases qui n'en contiennent pas pouvant être assimilées à de simples informations.

5.4.4 *Nombre de phrases*

I. L'automate de la fonction IOS passe à ON et l'ensemble des traitements sur les équipements senseur solaire, magnétomètres, magnéto-coupleurs, roue à réaction est séquencé. (E-C) / L'automate de la fonction IOS passe à ON.
L'ensemble des traitements sur les équipements senseur solaire, magnétomètres, magnéto-

coupleurs, roue à réaction est séquencé. (L)

Langues contrôlées : deux phrases
Experts : une phrase (3,88)
Non experts : pas de préférence (4,18)
Corpus : une phrase

Cette question est la seule pour laquelle la différence entre les résultats obtenus pour les propositions n'est pas significative ; on ne peut donc rien en conclure de façon certaine. Les valeurs des deux variantes sont en effet extrêmement proches, avec une très légère préférence pour celle qui ne contient qu'une phrase (et donc une conjonction), peut-être parce que les deux actions sont ressenties comme étant liées (en effet, il est probable qu'une ne doive pas être exécutée sans l'autre). Dans tous les cas, on ne sait pas si elles doivent être exécutées séquentiellement ou en même temps.

II. Il est possible d'importer et d'exporter toutes les règles de transfert déclarées. (N-C) / Il est possible d'importer toutes les règles de transfert déclarées.
Il est possible d'exporter toutes les règles de transfert déclarées. (L-E)

Langues contrôlées : deux phrases
Experts : deux phrases (4,13)
Non experts : une phrase (5,00)
Corpus : une phrase

Les non experts choisissent unanimement la variante en une seule phrase, qui évite la répétition de nombreux mots, tandis que les experts ont une infime (0,01 point) préférence pour la variante en deux phrases, sans doute car ils y voient deux fonctionnalités distinctes. Cette exigence reste d'assez haut niveau, puisqu'il n'est pas dit comment l'importation et l'exportation doivent être réalisées.

III. Les champs SM_ID et FM_ID seront extraits à partir de la BDS. (E-N-C) / Le champ SM_ID sera extrait à partir de la BDS.
Le champ FM_ID sera extrait à partir de la BDS. (L)

Langues contrôlées : deux phrases
Experts : une phrase (4,33)
Non experts : une phrase (4,97)
Corpus : une phrase

Les non experts continuent à préférer la phrase unique, ce qui est cette fois aussi le cas des experts (même si l'autre composante de l'alternative est également bien notée). Cela signifierait que les deux opérations peuvent raisonnablement être exprimées au sein de la même phrase (et donc de la même exigence).

IV. Si le transfert d'un fichier de télémesure échoue après le nombre de tentatives fixées, le fichier sur le destinataire est effacé et l'information est remontée aux IHM. (L-E-N-C) / Si le transfert d'un fichier de télémesure échoue après le nombre de tentatives fixées, le fichier sur le destinataire est effacé.
L'information est remontée aux IHM.

Langues contrôlées : une phrase
Experts : une phrase (4,10)
Non experts : une phrase (4,62)
Corpus : une phrase

La conjonction « et » étant sous la portée de la condition, le GWR autorise exceptionnellement son utilisation, sans quoi le résultat ne serait plus exactement le même. Les experts et les non experts optent sans surprise pour cette option.

Synthèse

TABLEAU 17 – Synthèse : Nombre de phrases

	Moyenne Générale	Moyenne Experts	Moyenne Non Experts	Différence significative
une phrase	4,37	4,11	4,69	oui
deux phrases	3,72	3,92	3,48	oui

Il y a une préférence (significative sur la moyenne générale – voir TABLEAU 17) pour les variantes en une phrase, très marquée chez les non experts mais plus faible chez les experts (qui dans certains cas choisissent la variante en deux phrases). Ce résultat va donc globalement à l'encontre de la recommandation du GWR et du principe d'atomicité des exigences poussé à l'extrême, et montre qu'il y a sans doute une place pour les conjonctions (même « et ») dans les exigences – comme nous le recommandions d'ailleurs. Nous prenons toutefois garde de tirer des conclusions définitives, car les cas que nous avons envisagés sont assez différents les uns des autres. D'autres questions sur la même problématique seraient utiles, même si, comme nous l'avions déjà fait remarquer, envisager toutes les possibilités semble difficile.

Dans les commentaires laissés à propos de ces questions, on trouve des arguments en faveur d'une phrase unique : l'exigence est elle-même moins longue et moins redondante, et le risque d'oublier de prendre une phrase en compte serait moindre. Cela dépend surtout s'il y a une relation évidente entre les deux phrases, qui justifierait de les lier, sinon il serait préférable de scinder. Les experts avancent qu'une phrase par fonctionnalité rend l'exigence plus facile à lire, comprendre, réaliser et tester, même s'il y a des cas où les conjonctions sont avantageuses, voire nécessaires : les conditions (« si 1, alors 2 ») et les séquences (« 1, puis 2 »), entre autres. Malheureusement, sans connaître le système décrit, il est assez difficile de

déterminer si deux actions doivent être exécutées en même temps ou successivement.

5.4.5 *Anaphore*

I. Le paquet cyclique ne sera généré que si le paquet / ce paquet / ce paquet cyclique (L) / il (E-N-C) est activé par le LVC.

Langues contrôlées : « ce paquet (cyclique) »
Experts : « il » (4,58)
Non experts : « il » (4,85)
Corpus : « il »

Si l'on s'en tenait strictement à la recommandation du GWR, il serait nécessaire de répéter le syntagme nominal dans la subordonnée. On constate pourtant que les experts et les non experts préfèrent largement le pronominaliser : la phrase qui en résulte est plus courte, plus naturelle et n'est pas pour autant ambiguë. Sans surprise, la variante avec reprise du nom précédé d'un déterminant défini n'est pas du tout appréciée : elle paraît incongrue et peut même sembler ambiguë (la répétition amenant à se demander s'il s'agit bien du même paquet dans les deux cas).

II. La configuration du système est stockée à bord. La configuration / Cette configuration / Cette configuration du système (L) / Elle (E-N-C) ne peut être modifiée que par deux événements : événement 1 et événement 2.

Langues contrôlées : « Cette configuration (du système) »
Experts : « Elle » (4,12)
Non experts : « Elle » (4,38)
Corpus : « Elle »

Les résultats obtenus pour cette question sont semblables à ceux de la précédente, et ce même si l'antécédent du pronom se trouve dans une autre phrase (ce qui n'entrave pas la compréhension).

Synthèse intermédiaire

Contrairement à ce que préconise le GWR, le Tableau 18 révèle que les experts préfèrent (significativement) utiliser un pronom personnel sujet non ambigu plutôt que de répéter le syntagme en entier, même si ce dernier est précédé d'un déterminant démonstratif.

III. En cas de réception d'un fichier vide, le fichier vide / ce fichier (E-N-C) / ce fichier vide (L) / il n'est pas transmis.

TABLEAU 18 – Synthèse : Anaphore

	Moyenne Générale	Moyenne Experts	Moyenne Non Experts	Différence significative
pronom	4,47	4,35	4,62	non
démonstratif + tête	3,72	3,83	3,59	non
démonstratif + groupe	3,18	3,64	2,63	oui
défini + tête	2,60	2,54	2,68	non

Langues contrôlées : « ce fichier (vide) »
Experts : « ce fichier » (4,10)
Non experts : « ce fichier » (4,36)
Corpus : « celui-ci »

Cette fois, les experts et même les non experts préfèrent répéter une partie du syntagme plutôt que d'utiliser le pronom « il », pourtant non ambigu. Il faut toutefois remarquer que dans l'exigence originale, c'était le pronom démonstratif « celui-ci » qui était utilisé, ce qui se justifie probablement par le fait que la fonction syntaxique n'est pas la même dans les deux cas (complément du nom / sujet). Nous avions toutefois préféré le remplacer par « il » pour garder la même structure dans toutes les questions. Plusieurs personnes l'ont d'ailleurs suggéré en commentaire et il est probable qu'il aurait obtenu un meilleur score. L'une d'entre elles a aussi indiqué que la tournure « En cas de réception d'un fichier vide… » (nominalisation déverbale) lui paraissait inutilement compliquée (et « laide »), et suggère de la remplacer par exemple par « Si le fichier reçu est vide… ».

IV. La liste des TCD est définie dans le document C12 ; la liste des TCD / cette liste (E-N) / cette liste des TCD (L) / elle (C) est exhaustive.

Langues contrôlées : « cette liste (des TCD) »
Experts : « cette liste » (4,10)
Non experts : « cette liste » (4,51)
Corpus : « elle »

De nouveau, la plupart des répondants (experts ou non) préfèrent répéter la tête du syntagme. Nous ne savons pas exactement pourquoi les résultats divergent de ceux des deux premières questions ; peut-être est-ce simplement parce que la répétition permet ici de donner une longueur similaire aux deux propositions et ainsi éviter un déséquilibre trop important (dans l'exigence originale, la seconde phrase était plus longue : « Elle est donnée ici à titre informatif »).

Synthèse intermédiaire

Dans ces deux questions, c'est la variante avec un déterminant démonstratif suivi du nom qui est significativement préférée (TABLEAU 19). La répétition du

Tableau 19 – Synthèse : Anaphore (2)

	Moyenne Générale	Moyenne Experts	Moyenne Non Experts	Différence significative
démonstratif + tête	4,25	4,10	4,44	non
pronom	3,63	3,60	3,65	non
démonstratif + groupe	3,50	3,81	3,12	oui
défini + groupe	2,96	3,21	2,65	oui

syntagme en entier reste assez appréciée des experts, moins des non experts.

V. L'utilisateur doit savoir à tout moment si l'utilisateur / si cet utilisateur (L) / s'il (E-N-C) est connecté sur le serveur GIDE nominal ou redondant.

Langues contrôlées : « cet utilisateur »
Experts : « il » (4,96)
Non experts : « il » (4,95)
Corpus : « il »

Dans cet exemple où le nom « utilisateur » n'a pas d'expansion, le pronom obtient une note presque parfaite à la fois des experts et des non experts, les deux autres propositions (moins naturelles) étant beaucoup moins bien notées.

Synthèse

Dans la majorité des cas, le pronom (à condition évidemment qu'il ne crée pas d'ambiguïté) est préféré à la répétition du nom ou du syntagme nominal. Cela confirme que proposer une règle imposant sans nuance de répéter tous les noms n'est sans doute pas la solution la plus pertinente, et que plus de souplesse serait nécessaire (par exemple en indiquant dans quels cas le pronom est autorisé ou au contraire à éviter, comme nous l'avons fait précédemment).

Les commentaires reçus à propos de ces questions font remarquer que les répétitions créent de la lourdeur et que les variantes plus courtes et plus simples sont aussi plus claires. Ils ne manquent pas non plus de signaler qu'en l'absence d'un déterminant démonstratif, la répétition du nom crée une possible ambiguïté.

5.4.6 *Agent*

I. L'opération n'est nécessaire que si la surveillance fonctionnelle est autorisée par le LVC (E-N-C) / la surveillance fonctionnelle est autorisée / le LVC autorise la surveillance fonctionnelle (L) / on autorise la surveillance fonctionnelle.

Langues contrôlées : « le LVC autorise… »
Experts : « … est autorisée par le LVC » (4,42)

Non experts : « … est autorisée par le LVC » (4,51)
Corpus : « … est autorisée par le LVC »

Ce résultat constitue une surprise, car il va à l'encontre de ce que recommandent la plupart des langues contrôlées : toujours utiliser la voix active. C'est pourtant la voix passive (que l'on retrouve aussi dans le corpus) qui est préférée, tant par les experts que par les non experts. Les deux variantes sont possibles, mais la voix passive permet peut-être d'insister davantage sur la surveillance fonctionnelle (placée en premier), plus importante que l'agent ici. Évidemment, les deux variantes moins complètes (voix passive sans agent et pronom « on ») sont beaucoup moins bien évaluées.

II. Le vidage des tables sera contrôlé par le CCC. (N-C) / Le vidage des tables sera contrôlé. / Le CCC contrôlera le vidage des tables. (L-E) / On contrôlera le vidage des tables.

Langues contrôlées : « le CCC contrôlera… »
Experts : « le CCC contrôlera… » (4,56)
Non experts : « … sera contrôlé par le CCC » (4,62)
Corpus : « … sera contrôlé par le CCC »

Sur cette question, les avis des experts et des non experts divergent. Les experts se rangent ici du côté des langues contrôlées, en préférant la voix active, plus classique (sujet – verbe – complément). Les non experts, quant à eux, ont une préférence pour la voix passive, que l'on retrouvait une fois de plus dans le corpus – peut-être parce qu'elle permet une focalisation différente.

Synthèse intermédiaire

TABLEAU 20 – Synthèse : Agent

	Moyenne Générale	Moyenne Experts	Moyenne Non Experts	Différence significative
passif avec agent	4,45	4,35	4,56	non
voix active	4,26	4,32	4,18	non
passif sans agent	2,60	2,45	2,78	non
« on »	1,86	1,64	2,14	oui

Comme on le voit (TABLEAU 20), la voix passive avec agent est légèrement préférée à la voix active, mais cette différence n'est pas significative. On peut par conséquent légitimement s'interroger sur l'intérêt réel d'imposer systématiquement la voix active dans les exigences (et d'autres textes techniques) : il semblerait que dans certains cas, les experts la choisissent et que dans d'autres, ce soit la voix passive qui remporte les suffrages, sans doute en fonction de ce qui doit être mis en avant. En réalité, comme nous l'avions déjà souligné, ce n'est pas la voix passive

elle-même qui risque de poser problème, mais la voix passive *sans agent* clairement identifié. Là encore, ces résultats semblent aller dans notre sens et appellent à une révision des règles existantes.

On trouve d'ailleurs des commentaires opposés à propos de ces deux questions : certains recommandent de placer l'agent en début de phrase, d'autres de commencer par l'« objet principal de l'exigence » (en l'occurrence, « le vidage des tables »), un autre encore dit ne pas voir de différence entre la voix active et la voix passive. Tous s'accordent en revanche sur le fait que l'agent doit être précisé, et « on » est toujours qualifié de plus mauvais (même quand la formulation alternative n'a pas non plus d'agent). Les questions suivantes confortent du reste clairement cette dernière observation.

III. La durée d'observation imposée est comprise entre 20 et 100 secondes. (E-N) / On impose une durée d'observation comprise entre 20 et 100 secondes. (C)

Experts : « La durée d'observation imposée… » (4,75)
Non experts : « La durée d'observation imposée… » (4,92)
Corpus : « On impose… »

Les trois dernières exigences ne contenaient pas d'agent (et sont donc forcément « mauvaises » du point de vue des langues contrôlées), si ce n'est le pronom « on ». Ce dernier est à chaque fois le plus mal noté (en particulier par les experts), et ce même s'il est possible de l'identifier, par exemple, au rédacteur, et même si c'était la variante trouvée en corpus.

IV. Les TC associées aux familles F_EMETTEUR et F_MANOEUVRES sont listées ici. (E-N) / On liste ici les TC associées aux familles F_EMETTEUR et F_MANOEUVRES. (C)

Experts : « Les TC associées… » (4,87)
Non experts : « Les TC associées… » (4,62)
Corpus : « On liste… »

V. La TC est refusée si la taille du paquet programmé dépasse la taille maximum. (E-N-C) / On refuse la TC si la taille du paquet programmé dépasse la taille maximum.

Experts : « La TC est refusée » (4,73)
Non experts : « La TC est refusée » (4,77)
Corpus : « La TC est refusée »

TABLEAU 21 – Synthèse : Agent (2)

	Moyenne Générale	Moyenne Experts	Moyenne Non Experts	Différence significative
passif sans agent	4,78	4,78	4,77	non
« on »	3,16	2,92	3,45	oui

Synthèse

Comme nous nous y attendions, le passif sans agent est significativement préféré au pronom « on » (TABLEAU 21) ; ce dernier n'étant apprécié ni des experts, ni des non experts (ce qui corrobore les observations de HUPET et COSTERMANS (1976)). On en trouve pourtant plusieurs occurrences dans le corpus, mais il vaudrait sans doute mieux les limiter autant que possible.

5.5 CONCLUSION

Dans ce dernier chapitre, nous avons rendu compte des résultats d'une enquête en ligne visant à tester la pertinence des règles trouvées dans les langues contrôlées que nous avions décrites au chapitre 3. La méthode étant facilement reproductible (et ne demandant pas un temps excessif au personnel du CNES), elle pourra également servir à tester d'autres règles, ainsi que des structures identifiées grâce à l'approche statistique. L'objectif est de combler un manque de preuve empirique dans les « bonnes pratiques » de rédaction des exigences et de voir si celles-ci nécessitent, comme nous le pensions, d'être partiellement redéfinies (et comment). Cette étape de validation, trop souvent négligée, nous paraît pourtant indispensable, sous cette forme ou sous une autre.

Nous avons donc demandé à des experts et à des « profanes » de la rédaction des exigences de noter différentes variantes d'une même exigence, certaines rédigées en suivant les recommandations et d'autres non, et avons analysé les résultats pour déterminer si les premières étaient réellement préférées aux secondes. Comme toujours, nous avons pris soin de partir du corpus d'exigences pour nous assurer du caractère (semi-)authentique de nos exemples (et par la même occasion de leur plausibilité), qui ont cependant dû être partiellement simplifiés par rapport aux exigences réellement rédigées et validées au CNES.

Il nous a en effet semblé préférable de demander aux volontaires ayant accepté de répondre à l'enquête d'évaluer des exemples concrets inspirés des règles, plutôt que ces mêmes règles directement, afin de limiter en partie les biais dus aux préjugés, en particulier chez les experts (d'après nos échanges avec eux, il y a fort à parier que si nous leur avions simplement demandé s'il était préférable de rédiger une phrase à la voix active ou à la voix passive, ils auraient sans hésiter répondu que seule la voix active devait être employée – et pourtant, ils ont globalement préféré une subordonnée à la voix passive dans une des questions), et dans une moindre mesure chez les non experts (chez qui les jugements épilinguistiques

sont une réalité). Cela nous a également permis de leur présenter des exemples variés, pour éviter qu'ils se focalisent sur un cas particulier et en négligent d'autres (par exemple, un pronom peut effectivement représenter un risque dans un certain contexte, mais être parfaitement acceptable dans d'autres). Bien entendu, nous privilégions l'avis des experts, puisqu'ils seront, au final, les utilisateurs de la langue contrôlée que nous essayons de définir ; il nous a néanmoins paru intéressant de comparer leur opinion avec celle des non experts, pour voir si leur habitude du genre modifie leur perception de la façon dont les exigences devraient être rédigées (et en particulier s'ils sont, comme on pourrait s'y attendre, plus en phase avec les recommandations traditionnelles – ce qui, de façon générale, s'est avéré).

Nous avons de ce fait systématiquement comparé trois avis (et l'occurrence trouvée en corpus) – celui des langues contrôlées, celui des experts (du genre et accessoirement du domaine) et celui des non experts – afin de voir dans quels cas ils convergent et dans quels cas ils divergent. Si l'on se concentre sur les deux premiers, trois cas de figure principaux peuvent se présenter : si l'avis des experts et la recommandation vont dans le même sens, cela accrédite en partie cette dernière ; si, au contraire, ils s'opposent, cela indique peut-être qu'elle nécessite d'être revue, de préférence en concertation avec les experts du CNES ou même avec les concepteurs de la langue contrôlée ; enfin, si aucune variante n'est significativement préférée par rapport aux autres, il convient de se demander si la règle est réellement nécessaire ou si elle ne représente pas juste une perte de temps. Au premier cas correspond par exemple l'utilisation du modal « devoir », qui plaît aux experts comme aux non experts ; au deuxième correspond notamment la pronominalisation, qui semble souvent préférable à la reprise d'un syntagme nominal si l'identification de l'antécédent du pronom ne pose aucun problème ; au troisième correspond entre autres la voix passive, qui ne semble ni toujours meilleure, ni toujours moins bonne que la voix active. Pour un même phénomène, ces résultats ne sont toujours identiques d'une question à l'autre, mais des tendances assez claires se dessinent malgré tout. Cela confirme, de plus, que les règles se doivent de rester suffisamment flexibles.

Cette expérimentation souffre de quelques limitations dont nous sommes bien conscient. Tout d'abord et surtout, il faut se rappeler que nous n'avons pas évalué la compréhension effective (ce qui aurait nécessité un dispositif plus lourd [18]), mais seulement recueilli le jugement d'un nombre limité de répondants (sans certitude que les deux soient corrélés). Comme déjà dit, on sait, par expérience, que ces jugements peuvent être biaisés par des idées préconçues ou l'influence des langues contrôlées (il est d'ailleurs symptomatique que la proposition plébiscitée par les experts ne soit que rarement celle qui est trouvée en corpus !). Nous avons tenté de

18. [↑] On peut s'inspirer par exemple de celui mis en place par Jahchan dans le cadre de sa thèse (à venir, 2019) intitulée « To What Extent Does Text Simplification Improve Human Comprehension ? Cognitive Evaluations for the Optimization of the Airbus Cockpit Controlled Language for Future Aircraft » pour tester la compréhension de messages formulés conformément à une langue contrôlée utilisée dans les cockpits d'avion Airbus (qui prévoit l'omission de nombreux éléments syntaxiques) et les comparer avec des messages plus naturels.

limiter ce biais en proposant des exemples contextualisés, mais cela ne correspond évidemment pas à des conditions réelles de rédaction et de lecture. Les résultats obtenus nous paraissent malgré tout très intéressants et assez cohérents.

Ensuite, nous n'avons pas testé l'expertise réelle des répondants, nous contentant de les catégoriser comme « experts » ou « non experts » sur la seule base de leur réponse (anonyme) à une question, ce qui est commode mais forcément critiquable (pour les personnes travaillant sur des projets orbitaux du CNES, toutefois, nous avons de bonnes raisons de penser qu'ils sont effectivement habitués à lire des spécifications). À l'avenir, nous pourrons nous focaliser sur l'avis des experts du CNES et, idéalement, nous assurer à la fois du profil des participants et des conditions de passation (pas d'interruption, etc.).

Enfin, à travers cette expérience, nous avons voulu comparer les scores obtenus par différentes propositions inspirées de règles trouvées dans des langues contrôlées. Mais de meilleures règles (et donc de meilleures propositions) que celles que nous avons testées peuvent très bien exister ; nous pouvons tout au plus asserter qu'une proposition est la meilleure parmi celles que nous avons testées, mais pas dans l'absolu (ce qui serait de toute façon difficilement prouvable). Il faut en particulier noter que dans cette étude de faisabilité, nous avons volontairement testé des cas plutôt simples ; des situations plus complexes pourraient nécessiter plus d'évaluations.

Quoi qu'il en soit, cette contribution méthodologique nous a permis d'en apprendre davantage sur les préférences des utilisateurs et d'évaluer objectivement plusieurs règles proposées dans des langues contrôlées, parfois sans la moindre justification. À tout le moins, elle doit permettre d'éviter les recommandations les moins fondées, qui risqueraient d'être contre-productives à l'usage, et de trouver des pistes d'amélioration avant qu'elles ne soient imposées aux rédacteurs. D'autres analyses seraient encore possibles grâce aux résultats obtenus, par exemple pour comparer l'avis des experts habitués à lire et rédiger des exigences avec l'avis des experts qui en lisent régulièrement, mais n'en écrivent pas eux-mêmes.

CONCLUSION ET PERSPECTIVES

À travers ce travail, nous avons présenté une méthodologie visant à améliorer la constitution des langues contrôlées pour la rédaction technique. Ces propositions répondent selon nous à une triple nécessité : celle d'adapter finement les recommandations à la situation à laquelle elles s'appliquent, de leur offrir une base objective et de les évaluer, si possible en amont de leur utilisation effective en contexte réel.

L'intérêt des langues contrôlées est bien perçu et leur succès n'est probablement pas le fait du hasard : elles permettent en effet de prévenir (ou du moins de limiter) le risque langagier, dont les conséquences, dans certaines circonstances, peuvent être graves. Malheureusement, la multiplication des langues contrôlées à laquelle nous avons assisté lors des dernières décennies ne s'est pas (ou pas assez) accompagnée d'une réflexion sur la manière dont elles devraient être élaborées ou améliorées. Beaucoup de recommandations se transmettent d'une langue contrôlée à une autre, sans nécessairement avoir jamais été remises en question et testées. Certaines reposent sur l'héritage des formules de lisibilité, beaucoup mieux étudiées mais qui n'ont, pour la plupart, pas été conçues pour la rédaction technique. D'autres, enfin, semblent procéder d'intuitions, souvent pertinentes mais pas toujours bien exploitées.

Une analyse détaillée de quelques règles trouvées dans deux langues contrôlées connues nous a permis de montrer leurs multiples limites – mais aussi leur intérêt. Nous n'en remettons bien entendu pas l'existence en cause ; au contraire, elles servent de point de départ à nos propres analyses. Nous pensons cependant qu'elles mériteraient d'être mises à l'épreuve des usages et des utilisateurs, et qu'il est regrettable que l'expertise des linguistes leur fasse souvent défaut. Des études psycholinguistiques (favorisées notamment par l'essor du *plain language*) fournissent par ailleurs des résultats très intéressants sur la façon dont sont traités cognitivement des phénomènes linguistiques tels que la voix passive, mais notre compréhension reste malgré tout limitée.

C'est pourquoi nous suggérons que la linguistique de corpus peut également apporter sa contribution, au travers de l'analyse de productions existantes, et ce même si aucune langue contrôlée n'est imposée aux rédacteurs (comme c'est le cas au CNES). En effet, même en l'absence d'une norme imposée, une norme peut se constituer spontanément et servir de base pour la formulation de règles. Aussi l'existence d'un genre ou d'un sous-langage est-elle une caractéristique qui peut être mise à profit, après en avoir déterminé la grammaire, pour passer des régularités aux règles.

Pour répondre à la demande initiale du CNES (que nous avons mise en lien avec une problématique (socio)linguistique plus générale), nous avons appliqué notre méthode à un corpus d'exigences authentiques constitué à partir des spécifications de projets spatiaux qui nous ont été fournies. Nous en avons ainsi montré la faisabilité et avons obtenu quelques résultats que nous jugeons intéressants.

Dans le cadre de l'approche *corpus-based*, beaucoup de nos hypothèses (librement inspirées des langues contrôlées existantes) se sont vues confirmées, mais nous avons aussi observé des propriétés inattendues, dont la fréquence élevée de tournures passives au sein des exigences. Ces passifs nombreux, sans danger lorsque l'agent est précisé, invitent à repenser la règle qui en interdit catégoriquement l'utilisation. Grâce à l'approche *corpus-driven*, nous avons repéré des lexèmes et des motifs fréquents qui pourraient être suggérés aux rédacteurs pour augmenter la cohérence des spécifications et leur faire gagner du temps. Enfin, les résultats obtenus grâce l'enquête nous permettent de mieux connaître le sentiment des utilisateurs (experts) sur la formulation idéale des exigences et montrent, eux aussi, que les règles ne sont pas toujours parfaitement adaptées.

Les résultats que nous avons obtenus doivent à leur tour faire l'objet d'une évaluation, à deux niveaux. Il faut tout d'abord évaluer l'adéquation de la solution que nous proposons par rapport à la demande qui nous a été formulée ; c'est donc au CNES, qui en est à l'origine (en tant que « porteur du besoin », pourrait-on dire), de les juger de ce point de vue. Deux aspects (en lien avec le concept d'utilisabilité), en particulier, devraient être pris en compte : la facilité de mise en œuvre de nos règles (rappelons qu'aucune langue contrôlée n'est utilisée au CNES à l'heure actuelle et que nous étions donc libre du choix de la solution à proposer ; nous avons cherché à rester aussi proche que possible des usages réels), qui pourrait être évaluée par un retour à l'utilisation des ingénieurs, et leur efficacité (c'est-à-dire leur capacité à résoudre les problèmes visés sans en créer d'autres). Dans tous les cas, toute proposition de notre part devra être validée par les experts (ce qui a déjà partiellement été fait à travers l'enquête), car leur connaissance du domaine et du métier est évidemment tout aussi précieuse que notre connaissance de la langue.

L'autre évaluation doit porter sur l'intérêt des résultats pour la linguistique, que nous avons cherché à mettre en avant tout au long de cette étude, et sur la pertinence de notre approche (qui doit bien sûr encore être améliorée). Elle doit se faire par les pairs et nécessite de rendre compte des résultats et de leur analyse, ainsi que nous l'avons fait.

Signalons les difficultés particulières que nous avons rencontrées : le corpus (dont le contenu est assez technique) peut poser quelques problèmes d'analyse aux outils de TAL classiques, qui sont rarement entraînés pour les corpus spécialisés, mais c'est surtout la confidentialité des données qui a constitué une contrainte forte.

Rappelons en outre qu'il ne s'agissait que d'une étude exploratoire ; si celle-ci est jugée concluante et que la méthode (dont nous avons déjà souligné l'apport et la reproductibilité) est réutilisée, il pourrait être avantageux d'en automatiser certaines étapes.

Pour finir, deux étapes supplémentaires semblent s'imposer comme perspectives évidentes pour donner suite à notre étude. La première serait d'implémenter les propositions de règles que nous avons faites dans un outil de vérification automatique (tel qu'il en existe déjà plusieurs). Nous avons d'ailleurs tenu compte de cette finalité lors de nos analyses. Elle présente un intérêt indéniable pour les ingénieurs (puisqu'elle simplifierait la rédaction), mais peut aussi nous apporter de précieuses informations. Si l'on analyse les décisions prises par les rédacteurs d'accepter ou non les suggestions qui leur sont faites, il devrait être possible à la fois de déterminer le taux de bruit (et de voir s'il est acceptable) et d'identifier des cas problématiques ou particuliers, non pris en compte par la règle générale. Ces cas particuliers pourraient ensuite être détectés par le logiciel, sans que le rédacteur ait à s'en soucier.

La seconde perspective tient à la composition de notre corpus et à sa représentativité. Comme nous l'avons dit, celui-ci peut sans doute être considéré comme représentatif des exigences de projets spatiaux en français au CNES, mais il serait peut-être possible de gagner en généralité en intégrant des exigences provenant d'autres institutions ou entreprises et/ou d'autres types de projet. Un corpus plus large et intégrant d'autres domaines permettrait en outre d'étudier la dispersion des unités et leur répartition dans les sous-corpus. Plus encore, la même approche pourrait être appliquée à d'autres types de textes (ainsi bien sûr qu'à d'autres langues). Le besoin de clarté ne se cantonne pas aux exigences et se fait également ressentir, entre autres, dans les textes juridiques et administratifs [19]. Sur ce point, les objectifs du mouvement pour le *plain language* et des langues contrôlées pour la rédaction technique convergent clairement.

19. [↑] À titre d'exemple, nous avons été contacté à propos de l'interprétation à donner à une phrase du Bulletin Officiel des Impôts, qui précise que « Lorsque les dépenses concernent des éléments multiples [...], celles-ci doivent, par ailleurs, porter sur une partie significative des éléments installés. », l'interprétation à donner à « partie significative » pouvant diverger.

ANNEXES

A

GLOSSAIRE DES SPÉCIFICATIONS CNES

AC	A Confirmer
ACU	A Compléter Ultérieurement
AD	A Définir
ADC	Analytical to Digital Converter
ADU	A Définir Ultérieurement
AG	Actionneur Gyroscopique
AIT	Assemblage Intégration et Tests
AIV	Assembly Integration and Verification
AMDEC	Analyse des Modes de Défaillances de leurs Effets et de leur Criticité
AOS	Acquisition Of Signal
AP	Assurance Produit
API	Application Program Interface
APID	Application Process Identifier
ASH	Acquisition Safe Hold
ASM	Acceleration Sensor Mode
AUS-2Ghz	Station du réseau 2Ghz CNES située à AUSSAGUEL
AUS-TTCET	Station bande S Myriade située à AUSSAGUEL
BCP	Bureau de Coordination des Passages
BCU	Bloc Charge Utile
BDL	Bloc De Ligne
BDMS	Base de Données Mission Système
BDP	Boundary Protection Device
BDS	Base de Données Système *ou* Base de Données Satellite
BOL	Beginning of Life
BOSS	Banc Opérationnel Système Satellite
BPD	Boundary Protection Device
BVSS	Bancs de Validation Système et SCAO
C	Civil *ou* Clos
C/C	Commande Contrôle
CADU	Channel Access Data Unit
CAP	Custom Accurate Pointing
CCC	Centre de Commande Contrôle
CCSDS	Consultative Committee for Space Data Systems
CdG	Centre de Gravité
CdM	Centre de Masse
CECT	Centre d'Expertise de la Compensation de Traînée
CEM	Compatibilité Electro-Magnetique
CGPS	Cold Gas Propulsion System
CGPSP	Cold Gas Propulsion System Panel
CINESOOM	Bibliothèque cinématique du simulateur mission d'études Pléiades HR

CLCW	Command Link Control Word
CLTU	Command Link Transmission (ou Transfer) Unit
CMC	Commanding and Monitoring Component
CMG	Control Moment Gyro
CMS	Centre de Mécanique Spatiale *ou* Centre de Mission Scientifique
CMS(-)M	Centre de Mission Scientifique (Microscope)
CMTG	Centre de Mission Technologique GNSS
CNES	Centre National d'Etudes Spatiales
CO-I	CO - Investigateur
COCPIT	GSD management tool
COME	COmpresseur MEmoire
COO	Centre d'Orbitographie Opérationnelle
COP	Command Operation Procedure
COR	Centre d'Opérations Réseau
COTS	Commercial Off The Shelf
Cpr	Coefficient de pression de radiation solaire
CSG	Centre Spatial Guyanais
CSS	Cascading Style Sheet
CU	Charge Utile
CU-REF	Charge Utile de REFérence
CVI	Contôle visuel immédiat
DA	Document Applicable
DAS	Données Auxiliaires Satellite
DCPS	Département Conception et évaluation des Performances des Systèmes
DCU	Deciphering Ciphering Unit
Ddl	Degré de liberté
DdV	Début de vie
DE	Demande d'Evolution
DF	Drag-Free
DMPH	Département Mesures Physiques
DNS	Domain Name System
DoD	Depth of Discharge
DR	Document de Référence
DRPPC	Data Remote Processing PC
DSP	Densité Spectrale de Puissance
DSTM	DORIS Scientific TeleMetry
DT	Direct Tasking
DTD	Data Type Definition
DUPC	Dual Programming Center
ECM	Electronic Control Module
EGSE	Electrical Ground Support Equipment
EOL	End of Life
EP	Equivalence Principle
EPC	Electronic Power Conditioner
F-CLTU	Forward CLTU
fo	Fréquence orbitale

FAR	Filtre Anti Repliement
FARM	Frame Acceptance and Reporting Mechanism
FAT	File Allocation Table
FCMUGC	French Civil Main User Ground Center
FCT	Domaine de FonCtionnemenT hors mission
FDIR	Failure Detection Isolation and Recovery
FDMUGC	French Defense Main User Ground Center
FDTM	Failure Diagnostic Telemetry
FdV	Fin de vie
FEEU	Front End Electronic Unit
Fep(i/s)	Fréquence du principe d'équivalence (inertiel/spinné)
fep	Fréquence du signal EP
fepi	Fréquence du signal EP en mode inertiel
FOG	Fiber Optic Gyrometer
Forb	Fréquence orbitale
Forge	Système de gestion de développement collaboratif de logiciels
FOS	Frontal de l'Opérateur Satellite
FRM	Full Range Mode *ou* Full Resolution Mode
FTP	File Transfer Protocol
G2	Groupe 2 du CCC
GAM	Getter Activation Mode
GAP	Geocentric Accurate Pointing
GAZMIC	Outil de gestion du gaz de propulsion pour Microscope
GCO	Groupe de Coordination des Opérations
GIDE	Generic Interface for Data Exchange
GN2	Azote sous forme gazeux
GOTLIB	Bibliothèque de guidage opérationnelle du projet Pléiades HR
GPE	Groupe Performance en Exploitation
GPOM	Groupe Pour les Opérations de la Mission
GS	Générateur Solaire
GSD	Ground Segment Database
GTC	Groupe TéléCommande du CCC
HKTM	House Keeping TeleMetry
HKTM-P	House Keeping TeleMetry — Pass
HKTM-R	House Keeping TeleMetry — Recorded
HR	Haute Résolution
HRM	High Resolution Mode
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
I/F	Interface
IADC	Inter Agency space Debris Coordination committee
ICC	Image Calibration Center
ICD	Interface Control Document
ICU	Image Calibration Unit *ou* Interface Control Unit
ICUME	ICU Mechanical Ensemble
IEA	Instrument et Equipement Aérospatiaux

IERS	International Earth Rotation Service
IHM	Interface Homme Machine
INTEC	INTerface Exchange Component
IOS	IO for Servicing equipment
IP	Internet Protocol
IPU	Image Processing Unit
I_CECT	Intervenant du CECT
I_CMS	Intervenant du CMS
I_CMSM	Intervenant du CMSM
KDC	Key Distribution Center
KMJ	Kerberos Module for Java
KRN-2Ghz	Station du réseau 2Ghz CNES située à KIRUNA
KRN-TTCET	Station bande S Myriade située à KIRUNA
LDAP	Lightweight Directory Access Protocol
LDV	Ligne De Visée
LEOP	Launch Early OPerations
LNTHD	Ligne Numérique Très Haut Débit
LOC	Données de Localisation satellite
LOS	Loss Of Signal
LTTM	Long Term TeleMetry
LVC	Logiciel de Vol Central
M	Militaire
MAG	Magnetometer
MAN	MANoeuvre
MAP	MAintien à Poste
MCA	Mode de Contrôle de Accélérations
MCA3 & MCA6	sous modes du MCA pour la transition vers les sous modes mission
MCAc	sous-mode du MCA optimisé pour la Calibration de l'instrument
MCAi	sous-mode du MCA optimisé pour le pointage Inertiel
MCAs	sous-mode du MCA optimisé pour le pointage Spinné
MCA_cp	sous mode du MCA pour la calibration des propulseurs
MCI	Masse Centrage et Inertie
MCO	Manœuvre de Correction Orbitale
MDP	Message De Programmation
ME	Masse d'Epreuve
MIC	Microscope
MICROSCOPE	MICROSatellite à traînée Compensée pour l'Observation du Principe d'Equivalence
MIGS	MIcrosatellite Ground Segment
MIT	Massachussetts Institute of Technology
MLI	Multi Layer Insulation
MOI	Maître d'Oeuvre Intégrateur
MOSL	Maître d'Oeuvre Satellite
MPE	Micro Propulsion Electronics
MSLIB	Bibliothèque de mécanique Spatiale
MSP	Mode Stellaire Propulsif

MT	Micro Thruster
MTB	Magneto Torquer
MTM	Magnetometer
MTQ	Magneto TorQuer
MV	Modèle de Vol
NA	Non Applicable
NTP	Network Time Protocol
OBC	On Board Computer
OBMU	On Board Management Unit
OBS	On Board Software
OCP	Outil Central de Planification
OEF	Orbit Event File
ONERA	Office National d'Etudes et de Recherches Aérospatiales
ORAMIC	Outil de Restitution d'Attitude pour MICroscope
OS	Operating System
OT	Opération Technologique
PAN	Panchromatique
PCDU	Power Conditioning and Distribution Unit
PCM	Pulse Code Modulation
PdP	Plan des Passages
PDT	Plan De Travail
PDV	Prise De Vue
PE	Principe d'équivalence
PF ou P/F	Plate-Forme
PFR	Plateforme de référence
PGGS	Proteus Generic Ground Segment
PHR	Pléiades High Resolution
PHRLIB	Pléiades HR orbital Library
PI	Principal Investigateur
PLTM	PayLoad Telemetry
POG	Plan d'Opérations Général
PP	Période de Programmation
PRM	Pressure Regulation Module
PRS	Pressure Regulation Stage
PSM	Position Sensor Mode
PSO	Position Sur Orbite
Qc	Quaternion de Consigne
Qol	Quaternion Orbital Local
RC	commande de ReConfiguration station
RCF	Return Channel Frame
RF	Radio Fréquence
RMS	Root Mean Square
ROC	Réseau Opérationnel CNES
ROCF	Return Operational Control Frame
ROL	Repère Orbital Local
RW	Reaction Wheel

R_CECT	Responsable du CECT
R_CMS	Responsable du CMS
R_CMSM	Responsable du CMSM
S/L	Satellite
SAD	Système d'Analyse de Données
SAG	Structure d'Accueil Générique
SAGE	Space Accelerometer for Gravity Experiment
SAS	Senseur Acquisition Solaire
SCAA	Système de Contrôle d'Attitude et d'Accélération
SCAO	Système de Contrôle d'Attitude et d'Orbite
SCI	Domaine de performance SCIentifique (nominal)
SDD	Système de Diffusion de Données à la communauté scientifique
SDGC	Secure Dual Ground Center
SDM	Service de Détermination des Manœuvres
SED	Sous Ensemble de Détection
SEF	Service d'Echange de Fichiers
SEU	Single Event Upset
SFTP	SSH File Transfer Protocol
SGD	Système de Gestion des Données
SI	Système Informatique
SI-CMS	Système Informatique du CMS
SI-CMSM	Système Informatique du CMSM
SIPM	Satellite Inertial Pointing Mode
SIS	Système d'Information Satellite
SLE	Space Link Extension
SLID	SateLlite IDentifiant
SPD	Système de Production de Données pour l'expertise scientifique
Spec	Spécification
SPG	Scientific Performance Group
SRD	Système de Récupération de Données de l'Instrument et du Satellite
SSH	Secure Shell
SSM	Satellite Spinning Mode
SST	Senseur STellaire
STB	Spécification Technique de Besoin
STR	Star TRacker
SU	Sensor Unit
SU-EP	Sensor Unit pour la mesure de EP
SU-REF	Sensor Unit de référence
SUMI	Sensor Unit Mechanical Interface
SUP	Sun Pointing
SVN	SubVersioN : logiciel permettant le contrôle de versions des logiciels
SWG	Scientific Working Group
SWWW	Serveur WWW
T-SAGE	Twin Space Accelerometer for Gravity Experiment
TAC	Technical Archive Component
TAI	Temps Atomique International

TC	TéléCommande
TCD	TéléCommande Directe
TCH	Téléchargement
TCS	Techniques pour la Conception et la Simulation de systèmes
TDB	Technical Data Bank of TAC
TDI	Time Delay Integration
TGS	Ticket Granting Service
TGT	Ticket Granting Ticket
TM	TéléMesure
TMI	TéléMesure Image
TMR	Test Mass Release
TSAGE	Twin Space Accelerometer for Gravity Experiment
TTCET	Telemetry and TeleCommand Earth Terminal
TU	Temps Universel
TUC	Temps Universel Coordonné
TWT	Travelling Wave Tube
UC	Cas d'utilisation
UML	Unified Modeling Language
URL	Uniform Resource Locator
VCDU	Virtual Channel Data Unit
XML	eXtensible Markup Language
XS	Multispectral
XSD	XML Schema definition
XSLT	eXtensible Stylesheet Language Transformation

B

RÉSULTATS DE LA DEUXIÈME ÉTUDE CORPUS-DRIVEN

accélération
accès
action
activation
activité
aide
allocation
an
analyse
annexe
arborescence
archivage
argument
attribut
bande
base
BDS
besoin
biais
bloc
boucle
canal
capacité
caractéristique
cas
catalogue
cause
centre
centre de contrôle
chaîne
champ
changement
chapitre
charge
classe
client
cohérence
compte
conception
condition
consigne
construction
contenu
contrainte
contrôle
convergence
coordination
correction
correspondant
création
créneau
date
date de début
date de fin
début
décommutation
défaut
délai
demande
démarrage
dépôt
description
destination
développement
différence
disponibilité
disposition
document
donnée
Doppler
durée
écart
échange
échelle
éclipse
effet
élément
émission
ensemble
entrée
envoi
éphéméride
équipement
espace
estimation
étalonnage
état
événement
évolution
exécution
exemple
exigence
exploitation
extension
façon
facteur
fichier
figure
file
fin
flag
flux
fois
fonction
fonctionnalité
fonctionnement
format
forme
fourniture
fréquence
ftp
générateur
génération
gestion
graphe
groupe
heure
HKTM
HKTM-R
horizon
IHM
indisponibilité
information
installation
intégration
intervalle
jour
journal
lien
ligne
liste
logs
lune
machine
maintenance
manière
matrice
maximum
mécanisme
mémoire
méthode
minimum
minute
mise
mise à jour
mise on
mode
mode opérationnel
modification
mois
moment
montée
moyen
niveau
nœud
nom
nombre
nomenclature
numéro
objet

œuvre
opérateur
opération
ordre
orientation
origine
partie
passage
période
perte
phase
place
plage
plan
plan TC
plateforme
point
position
poste
poussée
précision
prédiction
présence
prise
prise en compte
problème
procédure
processus
produit
programmation
programmation mission
radian
ralliement
rapport
réalisation
recalage
réception
recette
recette en vol
référence
réglage
règle
rejeu
repère
répertoire
repli
reprise
requête
réseau
respect
restitution
retour
retrait
rythme
scénario
script
seconde
sécurité
semaine
sens
séquence
serveur ftp
session
soleil
somme
sortie
spécification
SST
stat
suite
suivi
support
survie
synthèse
système
table
tableau
tâche
taille
taux
télécommande
télémesure
température
temps
terme
terre
totalité
traitement
transition
transmission
travail
trou
tuyère
type
utilisateur
utilisation
valeur
variation
version
vie
visibilité
vue
warning

C

PRÉAMBULE DE L'ENQUÊTE

Consentement libre et éclairé :

En remplissant ce questionnaire :

Je certifie avoir donné mon accord pour participer à une étude de sciences humaines. J'accepte volontairement de participer à cette étude et je comprends que ma participation n'est pas obligatoire et que je peux stopper ma participation à tout moment sans avoir à me justifier ni encourir aucune responsabilité. Mon consentement ne décharge pas les organisateurs de la recherche de leurs responsabilités et je conserve tous mes droits garantis par la loi.

Au cours de cette enquête, j'accepte que soient recueillies des données sur mes réponses. Je comprends que les informations recueillies sont strictement confidentielles et que les résultats anonymés sont à usage exclusif de la communauté scientifique dans son ensemble. J'ai été informé(e) que mon identité n'apparaîtra dans aucun rapport ou publication et que toute information me concernant sera traitée de façon confidentielle. J'accepte qu'une fois anonymées, les données enregistrées à l'occasion de cette étude puissent être conservées dans une base de données et faire l'objet d'un traitement informatisé non nominatif en vue de publication scientifique.

J'ai pris connaissance de ces informations et donne mon accord à cette enquête / Je ne souhaite pas continuer

---

Questionnaire sur les exigences :

Bonjour et d'avance merci pour votre participation à ce questionnaire portant sur la rédaction et la compréhension des exigences techniques ! Celui-ci s'inscrit dans le cadre d'une thèse de doctorat poursuivie à l'Université Toulouse - Jean Jaurès (laboratoire CLLE-ERSS, France) et cofinancée par le CNES (Centre National d'Études Spatiales) et le Conseil régional de Midi-Pyrénées. Les résultats nous permettront d'améliorer les spécifications techniques de systèmes spatiaux et, plus généralement, de faire progresser la recherche scientifique. **Votre avis nous intéresse, que vous soyez ou non familier de la rédaction technique et/ou du domaine spatial.**

Contexte :
Une exigence peut se définir comme "l'expression d'une condition ou d'une fonc-

tionnalité à laquelle doit répondre un système ou un logiciel" (specief.org). Concrètement, il s'agit ici de phrases exprimant les besoins du client (= une description du mode de fonctionnement attendu des différentes parties du système concerné), auxquelles le fournisseur sera tenu de se conformer au moment de la réalisation du système. Ces exigences expriment donc une obligation et devront être respectées dans tous les cas, d'où l'importance qu'elles soient correctement interprétées par tous les intervenants. Les exigences sont regroupées dans des documents appelés spécifications.

Tâche :
Le questionnaire qui suit contient 20 questions (une par page). Pour chacune d'entre elles, la consigne est la même : nous vous proposons entre 2 et 6 formulations différentes, mais qui sont censées exprimer le même besoin. Nous voudrions savoir lesquelles vous semblent les plus adaptées (c'est-à-dire les plus claires, les moins ambiguës, etc.) et lesquelles sont, selon vous, à éviter (car pouvant prêter à confusion, par exemple), en imaginant que vous êtes la personne en charge de lire ces exigences et de les respecter. Pour cela, il vous est demandé d'**attribuer la note maximale (5/5) à la meilleure option (ou aux meilleures options, si vous estimez qu'elles se valent)** ***sur la page*****, puis d'attribuer une note relative (entre 1 et 4) à toutes les autres options** (une ligne du tableau = une option).

Conseils et remarques :
- Il n'y a pas de "bonnes" ou de "mauvaises" réponses, nous sommes intéressés par votre ressenti.
- Pour chaque question, vous avez la possibilité (si vous le souhaitez) de justifier votre choix. Nous serions très intéressés par vos remarques.
- Nous vous conseillons de vous installer dans un endroit tranquille, à un moment où vous aurez suffisamment de temps devant vous et où vous ne serez pas dérangé jusqu'à la fin du questionnaire.

Vous êtes prêt ! Vous pouvez commencer le questionnaire.

D

EXIGENCES UTILISÉES POUR L'ENQUÊTE

D.1 MODALISATION

I.	L'opérateur du générateur de TCH pourra rentrer l'adresse (champ adresse) directement ou indirectement en précisant le nom d'un label LVC.
II.	[...] La configuration est stockée à bord. Elle ne peut être modifiée que par deux événements : [...]
III.	[...] L'ensemble des TC sera saisi en BD, avec une contrainte de taille du mnémonique BD de 12 caractères maximum. [...]
IV.	Le DUPC doit combiner en une seule séquence MDP de lecture l'ensemble des ordres de programmation des trois canaux TMI.
V.	[...] Sur cet ordre, le LVC déroute l'exécution du MDP principal vers le MDP DT. [...]
VI.	[...] Le plan TC devra respecter les contraintes décrites dans [DR20].

D.2 NOMBRE DE PHRASES

I.	[...] L'automate de la fonction IOS passe à ON, et l'ensemble des traitements sur les équipements Senseur solaire, magnétomètres, magnéto-coupleurs, roue à réaction est séquencé.
II.	Il est possible d'exporter et d'importer toutes les règles de transfert déclarées. [...]
III.	Les champs SM_ID et FM_ID seront extraits à partir de la BDS, aucun contrôle n'est donc nécessaire (on considère la BDS validée).
IV.	Si le transfert en cours d'un fichier échoue après le nombre de tentatives fixées, le fichier sur le destinataire est effacé et l'information est remontée aux IHM ainsi qu'au journal de bord.

D.3 ANAPHORE

I. [...] Le paquet ne sera généré que s'il est activé par le LVC.

II. [...] La configuration est stockée à bord. Elle ne peut être modifiée que par deux événements : [...]

III. En cas de réception d'un fichier vide, celui-ci n'est pas transmis et un délais d'attente configurable sera respecté avant de tenter une nouvelle connexion. [...]

IV. La liste des TCD est définie en BDS. Elle est donnée ici à titre informatif : [...]

V. L'utilisateur doit pouvoir savoir à tout moment s'il est connecté sur le serveur GIDE nominal ou redondant.

D.4 AGENT

I. [...] Remarque : la surveillance ne sera exécutée que si la surveillance fonctionnelle est activée par le LVC.

II. Le vidage des tables sera contrôlé par le CCC. [...]

III. [...] On recommande donc une durée d'observation de l'ordre de 25 à 100 sec, incluse dans le mode MCO considéré. [...]

IV. On ne listera ici que les TC associées aux familles F_EMETTEUR et F_MANOEUVRES. [...]

V. [...] Cette TC est refusée si la taille du paquet TM programmé dépasse la taille maximum d'un paquet TM telle que définie dans l'IF-1/6-10-CN.

E

RÉSULTATS DÉTAILLÉS DE L'ENQUÊTE

E.1 MODALISATION

I.

a. *L'opérateur a le choix du mode de saisie : il doit pouvoir rentrer l'adresse directement ou indirectement en précisant le nom d'un label.*

Notes maximales :	36
Moyenne générale :	3,69
Moyenne des experts :	4,17
Moyenne des non experts :	3,10
Différence significative ?	oui

b. *L'opérateur a le choix du mode de saisie : il peut rentrer l'adresse directement ou indirectement en précisant le nom d'un label.*

Notes maximales :	25
Moyenne générale :	3,40
Moyenne des experts :	2,98
Moyenne des non experts :	3,92
Différence significative ?	oui

c. *L'opérateur a le choix du mode de saisie : il rentre l'adresse directement ou indirectement en précisant le nom d'un label.*

Notes maximales :	22
Moyenne générale :	3,18
Moyenne des experts :	3,04
Moyenne des non experts :	3,36
Différence significative ?	non

d. *L'opérateur a le choix du mode de saisie : il pourra rentrer l'adresse directement ou indirectement en précisant le nom d'un label.*

Notes maximales :	16
Moyenne générale :	3,09
Moyenne des experts :	2,65
Moyenne des non experts :	3,64
Différence significative ?	oui

e. *L'opérateur a le choix du mode de saisie : il devra pouvoir rentrer l'adresse directement ou indirectement en précisant le nom d'un label.*

Notes maximales :	12
Moyenne générale :	2,87
Moyenne des experts :	3,33
Moyenne des non experts :	2,31
Différence significative ?	oui

f. *L'opérateur a le choix du mode de saisie : il rentrera l'adresse directement ou indirectement en précisant le nom d'un label.*

Notes maximales :	8
Moyenne générale :	2,85
Moyenne des experts :	2,71
Moyenne des non experts :	3,03
Différence significative ?	non

II.

a. *La configuration doit pouvoir être modifiée par deux événements.*

Notes maximales :	32
Moyenne générale :	3,69
Moyenne des experts :	4,17
Moyenne des non experts :	3,10
Différence significative ?	oui

b. *La configuration peut être modifiée par deux événements.*

Notes maximales :	31
Moyenne générale :	3,23
Moyenne des experts :	2,77
Moyenne des non experts :	3,79
Différence significative ?	oui

c. *La configuration pourra être modifiée par deux événements.*

Notes maximales :	18
Moyenne générale :	3,10
Moyenne des experts :	2,71
Moyenne des non experts :	3,59
Différence significative ?	oui

d. *La configuration devra pouvoir être modifiée par deux événements.*

Notes maximales :	14
Moyenne générale :	3,05
Moyenne des experts :	3,17
Moyenne des non experts :	2,90
Différence significative ?	non

e. *La configuration est modifiée par deux événements.*

Notes maximales :	17
Moyenne générale :	2,64
Moyenne des experts :	2,21
Moyenne des non experts :	3,18
Différence significative ?	oui

f. *La configuration sera modifiée par deux événements.*

Notes maximales :	4
Moyenne générale :	2,56
Moyenne des experts :	2,23
Moyenne des non experts :	2,97
Différence significative ?	oui

III.

a. *L'ensemble des TC doit être saisi en base de données.*

Notes maximales :	47
Moyenne générale :	4,28
Moyenne des experts :	4,38
Moyenne des non experts :	4,15
Différence significative ?	non

b. *L'ensemble des TC devra être saisi en base de données.*

Notes maximales :	20
Moyenne générale :	3,57
Moyenne des experts :	3,65
Moyenne des non experts :	3,49
Différence significative ?	non

c. *L'ensemble des TC sera saisi en base de données.*

Notes maximales :	13
Moyenne générale :	3,20
Moyenne des experts :	2,83
Moyenne des non experts :	3,64
Différence significative ?	oui

d. *L'ensemble des TC est saisi en base de données.*

Notes maximales :	23
Moyenne générale :	3,16
Moyenne des experts :	2,94
Moyenne des non experts :	3,44
Différence significative ?	non

IV.

a. *Le DUPC doit combiner en une seule séquence de lecture l'ensemble des ordres de programmation des trois canaux.*

Notes maximales :	40
Moyenne générale :	4,14
Moyenne des experts :	4,40
Moyenne des non experts :	3,82
Différence significative ?	oui

b. *Le DUPC combine en une seule séquence de lecture l'ensemble des ordres de programmation des trois canaux.*

Notes maximales :	36
Moyenne générale :	3,55
Moyenne des experts :	3,13
Moyenne des non experts :	4,08
Différence significative ?	oui

c. *Le DUPC devra combiner en une seule séquence de lecture l'ensemble des ordres de programmation des trois canaux.*

Notes maximales :	19
Moyenne générale :	3,53
Moyenne des experts :	3,67
Moyenne des non experts :	3,36
Différence significative ?	non

d. *Le DUPC combinera en une seule séquence de lecture l'ensemble des ordres de programmation des trois canaux.*

Notes maximales :	13
Moyenne générale :	3,13
Moyenne des experts :	2,98
Moyenne des non experts :	3,31
Différence significative ?	non

V.

a. *Sur ordre du système, le LVC doit dérouter l'exécution du MDP principal vers le MDP secondaire.*

Notes maximales :	36
Moyenne générale :	3,89
Moyenne des experts :	4,33
Moyenne des non experts :	3,33
Différence significative ?	oui

b. *Sur ordre du système, le LVC devra dérouter l'exécution du MDP principal vers le MDP secondaire.*

Notes maximales :	21
Moyenne générale :	3,68
Moyenne des experts :	3,81
Moyenne des non experts :	3,51
Différence significative ?	non

c. *Sur ordre du système, le LVC déroute l'exécution du MDP principal vers le MDP secondaire.*

Notes maximales :	27
Moyenne générale :	3,43
Moyenne des experts :	3,08
Moyenne des non experts :	3,85
Différence significative ?	oui

d. *Sur ordre du système, le LVC déroutera l'exécution du MDP principal vers le MDP secondaire.*

Notes maximales :	14
Moyenne générale :	3,29
Moyenne des experts :	3,02
Moyenne des non experts :	3,62
Différence significative ?	oui

VI.

a. *Le plan de télécommande doit respecter les contraintes décrites dans le document DR20.*

Notes maximales :	51
Moyenne générale :	4,45
Moyenne des experts :	4,63
Moyenne des non experts :	4,23
Différence significative ?	oui

b. *Le plan de télécommande devra respecter les contraintes décrites dans le document DR20.*

Notes maximales :	24
Moyenne générale :	3,71
Moyenne des experts :	3,67
Moyenne des non experts :	3,77
Différence significative ?	non

c. *Le plan de télécommande respectera les contraintes décrites dans le document DR20.*

Notes maximales :	7
Moyenne générale :	3,07
Moyenne des experts :	3,02
Moyenne des non experts :	3,13
Différence significative ?	non

d. *Le plan de télécommande respecte les contraintes décrites dans le document DR20.*

Notes maximales :	20
Moyenne générale :	3,06
Moyenne des experts :	2,75
Moyenne des non experts :	3,44
Différence significative ?	oui

E.2 NOMBRE DE PHRASES

I.

a. *L'automate de la fonction IOS passe à ON et l'ensemble des traitements sur les équipements senseur solaire, magnétomètres, magnéto-coupleurs, roue à réaction est séquencé.*

Notes maximales :	43
Moyenne générale :	4,01
Moyenne des experts :	3,88
Moyenne des non experts :	4,18
Différence significative ?	non

b. *L'automate de la fonction IOS passe à ON.
L'ensemble des traitements sur les équipements senseur solaire, magnétomètres, magnéto-coupleurs, roue à réaction est séquencé.*

Notes maximales :	43
Moyenne générale :	3,95
Moyenne des experts :	3,77
Moyenne des non experts :	4,18
Différence significative ?	non

II.

a. *Il est possible d'importer et d'exporter toutes les règles de transfert déclarées.*

Notes maximales :	67
Moyenne générale :	4,52
Moyenne des experts :	4,12
Moyenne des non experts :	5,00
Différence significative ?	oui

b. *Il est possible d'importer toutes les règles de transfert déclarées.
Il est possible d'exporter toutes les règles de transfert déclarées.*

Notes maximales :	23
Moyenne générale :	3,59
Moyenne des experts :	4,13
Moyenne des non experts :	2,92
Différence significative ?	oui

III.

a. *Les champs SM_ID et FM_ID seront extraits à partir de la BDS.*

Notes maximales :	70
Moyenne générale :	4,62
Moyenne des experts :	4,33
Moyenne des non experts :	4,97
Différence significative ?	oui

b. *Le champ SM_ID sera extrait à partir de la BDS.
Le champ FM_ID sera extrait à partir de la BDS.*

Notes maximales :	23
Moyenne générale :	3,68
Moyenne des experts :	4,13
Moyenne des non experts :	3,13
Différence significative ?	oui

IV.

a. *Si le transfert d'un fichier de télémesure échoue après le nombre de tentatives fixées, le fichier sur le destinataire est effacé et l'information est remontée aux IHM.*

Notes maximales :	59
Moyenne générale :	4,33
Moyenne des experts :	4,10
Moyenne des non experts :	4,62
Différence significative ?	non

b. *Si le transfert d'un fichier de télémesure échoue après le nombre de tentatives fixées, le fichier sur le destinataire est effacé.
L'information est remontée aux IHM.*

Notes maximales :	28
Moyenne générale :	3,68
Moyenne des experts :	3,67
Moyenne des non experts :	3,69
Différence significative ?	non

E.3 ANAPHORE

I.

a. *Le paquet cyclique ne sera généré que s'il est activé par le LVC.*

Notes maximales :	68
Moyenne générale :	4,70
Moyenne des experts :	4,58
Moyenne des non experts :	4,85
Différence significative ?	oui

b. *Le paquet cyclique ne sera généré que si ce paquet est activé par le LVC.*

Notes maximales :	11
Moyenne générale :	3,37
Moyenne des experts :	3,63
Moyenne des non experts :	3,05
Différence significative ?	oui

c. *Le paquet cyclique ne sera généré que si ce paquet cyclique est activé par le LVC.*

Notes maximales :	16
Moyenne générale :	3,02
Moyenne des experts :	3,46
Moyenne des non experts :	2,49
Différence significative ?	oui

d. *Le paquet cyclique ne sera généré que si le paquet est activé par le LVC.*

Notes maximales :	4
Moyenne générale :	2,37
Moyenne des experts :	2,29
Moyenne des non experts :	2,46
Différence significative ?	non

II.

a. *La configuration du système est stockée à bord. Elle ne peut être modifiée que par deux événements : événement 1 et événement 2.*

Notes maximales :	49
Moyenne générale :	4,24
Moyenne des experts :	4,12
Moyenne des non experts :	4,38
Différence significative ?	non

b. *La configuration du système est stockée à bord. Cette configuration ne peut être modifiée que par deux événements : événement 1 et événement 2.*

Notes maximales :	34
Moyenne générale :	4,08
Moyenne des experts :	4,04
Moyenne des non experts :	4,13
Différence significative ?	non

c. *La configuration du système est stockée à bord. Cette configuration du système ne peut être modifiée que par deux événements : événement 1 et événement 2.*

Notes maximales :	17
Moyenne générale :	3,34
Moyenne des experts :	3,81
Moyenne des non experts :	2,77
Différence significative ?	oui

d. *La configuration du système est stockée à bord. La configuration ne peut être modifiée que par deux événements : événement 1 et événement 2.*

Notes maximales :	4
Moyenne générale :	2,84
Moyenne des experts :	2,79
Moyenne des non experts :	2,90
Différence significative ?	non

III.

a. *En cas de réception d'un fichier vide, ce fichier n'est pas transmis.*

Notes maximales :	48
Moyenne générale :	4,22
Moyenne des experts :	4,10
Moyenne des non experts :	4,36
Différence significative ?	non

b. *En cas de réception d'un fichier vide, il n'est pas transmis.*

Notes maximales :	32
Moyenne générale :	3,64
Moyenne des experts :	3,54
Moyenne des non experts :	3,77
Différence significative ?	non

c. *En cas de réception d'un fichier vide, ce fichier vide n'est pas transmis.*

Notes maximales :	20
Moyenne générale :	3,48
Moyenne des experts :	3,83
Moyenne des non experts :	3,05
Différence significative ?	oui

d. *En cas de réception d'un fichier vide, le fichier vide n'est pas transmis.*

Notes maximales :	8
Moyenne générale :	2,89
Moyenne des experts :	3,02
Moyenne des non experts :	2,72
Différence significative ?	non

IV.

a. *La liste des TCD est définie dans le document C12 ; cette liste est exhaustive.*

Notes maximales :	54
Moyenne générale :	4,29
Moyenne des experts :	4,10
Moyenne des non experts :	4,51
Différence significative ?	non

b. *La liste des TCD est définie dans le document C12 ; elle est exhaustive.*

Notes maximales :	24
Moyenne générale :	3,61
Moyenne des experts :	3,67
Moyenne des non experts :	3,54
Différence significative ?	non

c. *La liste des TCD est définie dans le document C12 ; cette liste des TCD est exhaustive.*

Notes maximales :	18
Moyenne générale :	3,52
Moyenne des experts :	3,79
Moyenne des non experts :	3,18
Différence significative ?	oui

d. *La liste des TCD est définie dans le document C12 ; la liste des TCD est exhaustive.*

Notes maximales :	9
Moyenne générale :	3,03
Moyenne des experts :	3,40
Moyenne des non experts :	2,59
Différence significative ?	oui

V.

a. *L'utilisateur doit savoir à tout moment s'il est connecté sur le serveur GIDE nominal ou redondant.*

Notes maximales :	83
Moyenne générale :	4,96
Moyenne des experts :	4,96
Moyenne des non experts :	4,95
Différence significative ?	non

b. *L'utilisateur doit savoir à tout moment si l'utilisateur est connecté sur le serveur GIDE nominal ou redondant.*

Notes maximales :	3
Moyenne générale :	2,01
Moyenne des experts :	2,00
Moyenne des non experts :	2,03
Différence significative ?	non

c. *L'utilisateur doit savoir à tout moment si cet utilisateur est connecté sur le serveur GIDE nominal ou redondant.*

Notes maximales :	4
Moyenne générale :	2,00
Moyenne des experts :	1,98
Moyenne des non experts :	2,03
Différence significative ?	non

E.4 AGENT

I.

a. *L'opération n'est nécessaire que si la surveillance fonctionnelle est autorisée par le LVC.*

Notes maximales :	52
Moyenne générale :	4,46
Moyenne des experts :	4,42
Moyenne des non experts :	4,51
Différence significative ?	non

b. *L'opération n'est nécessaire que si le LVC autorise la surveillance fonctionnelle.*

Notes maximales :	40
Moyenne générale :	4,09
Moyenne des experts :	4,08
Moyenne des non experts :	4,10
Différence significative ?	non

c. *L'opération n'est nécessaire que si la surveillance fonctionnelle est autorisée.*

Notes maximales :	8
Moyenne générale :	2,83
Moyenne des experts :	2,69
Moyenne des non experts :	3,00
Différence significative ?	non

d. *L'opération n'est nécessaire que si on autorise la surveillance fonctionnelle.*

Notes maximales :	1
Moyenne générale :	1,97
Moyenne des experts :	1,69
Moyenne des non experts :	2,31
Différence significative ?	oui

II.

a. *Le vidage des tables sera contrôlé par le CCC.*

Notes maximales :	48
Moyenne générale :	4,44
Moyenne des experts :	4,29
Moyenne des non experts :	4,62
Différence significative ?	non

b. *Le CCC contrôlera le vidage des tables.*

Notes maximales :	52
Moyenne générale :	4,43
Moyenne des experts :	4,56
Moyenne des non experts :	4,26
Différence significative ?	non

c. *Le vidage des tables sera contrôlé.*

Notes maximales :	6
Moyenne générale :	2,37
Moyenne des experts :	2,21
Moyenne des non experts :	2,56
Différence significative ?	non

d. *On contrôlera le vidage des tables.*

Notes maximales :	0
Moyenne générale :	1,76
Moyenne des experts :	1,58
Moyenne des non experts :	1,97
Différence significative ?	oui

III.

a. *La durée d'observation imposée est comprise entre 20 et 100 secondes.*

Notes maximales :	77
Moyenne générale :	4,83
Moyenne des experts :	4,75
Moyenne des non experts :	4,92
Différence significative ?	non

b. *On impose une durée d'observation comprise entre 20 et 100 secondes.*

Notes maximales :	11
Moyenne générale :	3,06
Moyenne des experts :	2,88
Moyenne des non experts :	3,28
Différence significative ?	non

IV.

a. *Les TC associées aux familles F_EMETTEUR et F_MANOEUVRES sont listées ici.*

Notes maximales :	72
Moyenne générale :	4,76
Moyenne des experts :	4,87
Moyenne des non experts :	4,62
Différence significative ?	non

b. *On liste ici les TC associées aux familles F_EMETTEUR et F_MANOEUVRES.*

Notes maximales :	19
Moyenne générale :	3,47
Moyenne des experts :	3,35
Moyenne des non experts :	3,62
Différence significative ?	non

V.

a. *La TC est refusée si la taille du paquet programmé dépasse la taille maximum.*

Notes maximales :	77
Moyenne générale :	4,75
Moyenne des experts :	4,73
Moyenne des non experts :	4,77
Différence significative ?	non

b. *On refuse la TC si la taille du paquet programmé dépasse la taille maximum.*

Notes maximales :	8
Moyenne générale :	2,95
Moyenne des experts :	2,54
Moyenne des non experts :	3,49
Différence significative ?	oui

BIBLIOGRAPHIE

AFIS (2012). *Guide Bonnes Pratiques en Ingénierie des Exigences*. Toulouse, France : Cépaduès. URL : http://www.cepadues.com/livres/guide-bonnes-pratiques-ingenierie-des-exigences-9782364930209.html.

AeroSpace and Defence Industries Association of Europe (2007). *Simplified Technical English. Specification ASD-STE100. International specification for the preparation of maintenance documentation in a controlled language. Issue 4*.

Akman, V. (2000). « Rethinking context as a social construct ». In : *Journal of Pragmatics* 32.6, p. 743–759. ISSN : 0378-2166. DOI : 10.1016/S0378-2166(99)00067-3. URL : http://www.sciencedirect.com/science/article/pii/S0378216699000673.

Albrecht, M. v. (2003). *Cicero's Style : A Synopsis*. Leiden/Boston : Brill. ISBN : 978-90-474-0197-1.

Andersson, L., Lupu, M., Palotti, J. R., Piroi, F., Hanbury, A. et Rauber, A. (2014). « Insight to Hyponymy Lexical Relation Extraction in the Patent Genre Versus Other Text Genres. » In : *Proceedings of the First International Workshop on Patent Mining and Its Applications (IPAMIN)*. Hildesheim, Allemagne.

Anthony, L. (2018). *AntConc*. Version 3.5.7. Tokyo, Japon. URL : http://www.laurenceanthony.net/software.

Arora, C., Sabetzadeh, M., Briand, L. et Zimmer, F. (2015). « Automated Checking of Conformance to Requirements Templates Using Natural Language Processing ». In : *IEEE Transactions on Software Engineering* 41.10, p. 944–968. ISSN : 0098-5589. DOI : 10.1109/TSE.2015.2428709.

Aubin, S. et Hamon, T. (2006). « Improving Term Extraction with Terminological Resources ». In : *Advances in Natural Language Processing*. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, p. 380–387. ISBN : 978-3-540-37334-6 978-3-540-37336-0. DOI : 10.1007/11816508_39. URL : https://link.springer.com/chapter/10.1007/11816508_39.

Aughterson, K. (2000). « Redefining the Plain Style : Francis Bacon, Linguistic Extension, and Semantic Change in "The Advancement of Learning" ». In : *Studies in Philology* 97.1, p. 96–143. ISSN : 0039-3738. URL : http://www.jstor.org/stable/4174661.

Auroux, S. (1998). *La raison, le langage et les normes*. Paris : Presses universitaires de France.

Austin, J. L. (1991). *Quand dire, c'est faire*. Paris, France : Éditions du Seuil.

Barcellini, F., Albert, C., Grosse, C. et Saint-Dizier, P. (2012). « Risk Analysis and Prevention : LELIE, a Tool dedicated to Procedure and Requirement Authoring. » In : *LREC*, p. 698–705.

Bartlett, J. M. (1964). « Frege : On the Scientific Justification of a Concept-Script ». In : *Mind* 73.290, p. 155–160. issn : 0026-4423. url : http://www.jstor.org/stable/2251806.

Béchet, N., Cellier, P., Charnois, T., Crémilleux, B. et Quiniou, S. (2013). « SDMC : un outil en ligne d'extraction de motifs séquentiels pour la fouille de textes ». In : *Conférence Francophone sur l'Extraction et la Gestion des Connaissances (EGC'13)*.

Beauvisage, T. (2001). « Exploiter des données morphosyntaxiques pour l'étude statistique des genres – application au roman policier ». In : *Traitement automatique des langues* 43.

Bendinelli, M. (2017). « Segments phraséologiques et séquences textuelles. Questions méthodologiques et approches topologiques ». In : *Corpus* 17. issn : 1638-9808. url : http://journals.openedition.org/corpus/2844.

Benezech, D. (1996). « La norme : une convention structurant les interrelations technologiques et industrielles ». In : *Revue d'économie industrielle* 75.1, p. 27–43. doi : 10.3406/rei.1996.1603. url : https://www.persee.fr/doc/rei_0154-3229_1996_num_75_1_1603.

Bertels, A. et Speelman, D. (2016). « Analyse sémantique distributionnelle dans un corpus technique : les enjeux sémantiques dans un jeu de statistiques ». In : *Actes de JADT 2016*, p. 273–283.

Bestgen, Y. (2013). « Inadequacy of the chi-squared test to examine vocabulary differences between corpora ». In : *Literary and Linguistic Computing* 29.2, p. 164–170.

Bhatia, V. K. (1993). *Analysing genre*. Londres, Royaume-Uni : Longman.

Biber, D. (1988). *Variation across speech and writing*. Cambridge : Cambridge University Press.

Biber, D. (1993). « Using register-diversified corpora for general language studies ». In : *Computational linguistics* 19.2, p. 219–241.

Biber, D. (2006). *University language : A corpus-based study of spoken and written registers*. T. 23. John Benjamins Publishing.

Biber, D. (2009). « A corpus-driven approach to formulaic language in English ». In : *International journal of corpus linguistics* 14.3, p. 275–311.

Biber, D. (2010). « Corpus-based and Corpus-driven Analyses ». In : Heine, B. et Narrog, H. *The Oxford Handbook of Linguistic Analysis*. Oxford University Press. isbn : 978-0-19-967707-8.

Blancafort, H., Heid, U., Gornostay, T., Méchoulam, C., Daille, B. et Sharoff, S. (2011). « User-centred Views on Terminology Extraction Tools : Usage Scenarios and Integration into MT and CAT Tools ». In : *Translation Careers and Technologies : Convergence Points for the Future (TRALOGY)*. Paris, France : INIST. URL : https://hal.archives-ouvertes.fr/hal-00818657.

Bostian, L. R. (1983). « How Active, Passive and Nominal Styles Affect Readability of Science Writing ». In : *Journalism Quarterly* 60.4, p. 635–670. ISSN : 0022-5533. DOI : 10.1177/107769908306000408. URL : https://doi.org/10.1177/107769908306000408.

Botel, M. et Granowsky, A. (1972). « A formula for measuring syntactic complexity : A directional effort ». In : *Elementary English* 49.4, p. 513–516.

Bouquet, S. (2007). « Contribution à une linguistique néo-saussurienne des genres de la parole (1) : une grammaire du morphème on ». In : *Linx. Revue des linguistes de l'université Paris X Nanterre* 56, p. 143–156. ISSN : 0246-8743. DOI : 10.4000/linx.376. URL : http://journals.openedition.org/linx/376.

Bowker, L. et Pearson, J. (2002). *Working with specialized language : a practical guide to using corpora*. Londres : Routledge.

Bres, J. (2009). « Dialogisme et temps verbaux de l'indicatif ». In : *Langue française* 163, p. 21–39. ISSN : 0023-8368. DOI : 10.3917/lf.163.0021. URL : https://www.cairn.info/revue-langue-francaise-2009-3-page-21.html.

Bulot, T. et Blanchet, P. (2013). *Une introduction à la sociolinguistique : pour l'étude des dynamiques de la langue française dans le monde*. Archives contemporaines. 166 p. URL : https://hal.archives-ouvertes.fr/hal-01436656.

Cabré, M. T. (1998). *La terminologie : théorie, méthode et applications*. Trad. par M. C. Cormier et J. Humbley. Ottawa/Paris : Les Presses de l'Université d'Ottawa/Armand Collin. ISBN : 978-2-200-21891-1.

Calvet, L. J. et Moreau, M.-L. (1998). *Une ou des normes ? Insécurité linguistique et normes endogènes en Afrique francophone*. Paris : Didier Érudition.

Carlier, A. (2002). « Les propriétés aspectuelles du passif ». In : *Cahiers Chronos* 10, p. 41–63.

Carlson, N. et Laplante, P. (2014). « The NASA automated requirements measurement tool : a reconstruction ». In : *Innovations in Systems and Software Engineering* 10.2, p. 77–91.

Cellier, P., Charnois, T., Plantevit, M., Rigotti, C., Crémilleux, B., Gandrillon, O., Kléma, J. et Manguin, J.-L. (2015). « Sequential pattern mining for discovering gene interactions and their contextual information from biomedical texts ». In : *Journal of Biomedical Semantics* 6. ISSN : 2041-1480. DOI : 10.1186/s13326-015-0023-3. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4436157/.

Charnock, R. (1999). « Les langues de spécialité et le langage technique : considérations didactiques ». In : *ASp. la revue du GERAS* 23, p. 281–302. ISSN : 1246-8185. DOI : 10.4000/asp.2566. URL : http://journals.openedition.org/asp/2566.

Charnock, R. (2008). « Raisonnement linguistique chez les juges anglais : le cas de l'interprétation des testaments ». In : *Lidil. Revue de linguistique et de didactique des langues* 38, p. 109–124. ISSN : 1146-6480. URL : http://journals.openedition.org/lidil/2781.

Charnois, T. (1999). « Détection d'anomalies sémantiques dans les textes de spécifications. Application au problème des interactions de services de télécommunications ». Thèse de doctorat. Paris 13.

Charrow, R. et Charrow, V. (1979). « Making legal language understandable : A psycholinguistic study of jury instructions ». In : *Columbia law review* 79.7, p. 1306–1374.

Cheminot, E. (1999). « Formalisation de spécifications de logiciels : traitement d'annotations en langue naturelle contrôlée ». Thèse de doctorat. INP Grenoble.

Chervak, S., Drury, C. G. et Ouellette, J. P. (1996). « Field Evaluation of Simplified English for Aircraft Workcards ». In : *Proceedings of the 10th FAA/AAM Meeting on Human Factors in Aviation Maintenance and Inspection*.

Chevrot, J.-P. (2012). « Que reste-t-il de la distinction compétence-performance dans les théories fondées sur l'usage ? » In : *Mémoires XXI de la Société Linguistique de Paris*. Louvain : Peeters Publisher, p. 49–68. URL : https://hal.archives-ouvertes.fr/hal-00850507.

Choi, K.-S. et Isahara, H. (2014). « Toward ISO Standard for Controlled Natural Language ». In : *Proceedings of the Workshop "Controlled Natural Language Simplifying Language Use" at LREC2014*, p. 1–2.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Massachusetts : MIT Press.

Clark, P., Murray, W. R., Harrison, P. et Thompson, J. (2009). « Naturalness vs. Predictability : A Key Debate in Controlled Languages ». In : *Controlled Natural Language*. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, p. 65–81. ISBN : 978-3-642-14417-2 978-3-642-14418-9. DOI : 10.1007/978-3-642-14418-9_5. URL : https://link.springer.com/chapter/10.1007/978-3-642-14418-9_5.

Cohen, K. B., Baumgartner, W. A. et Temnikova, I. (2016). « SuperCAT : The (New and Improved) Corpus Analysis Toolkit ». In : *International Conference on Language Resources and Evaluation* 2016, p. 2784–2788. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5860820/.

Coleman, E. B. (1964). « The comprehensibility of several grammatical transformations. » In : *Journal of Applied Psychology* 48.3, p. 186.

Coleman, E. B. (1962). « Improving comprehensibility by shortening sentences. » In : *Journal of Applied Psychology* 46.2, p. 131.

Condamines, A. (2003). « Sémantique et corpus spécialisés : Constitution de Bases de Connaissances Terminologiques ». Mémoire d'Habilitation à Diriger les Recherches. Université Toulouse Le Mirail. URL : https://halshs.archives-ouvertes.fr/tel-01321042/document.

Condamines, A., Lopez, S. et Warnier, M. (2017). « The Linguist and Workplace Norms ». In : *Cahiers de lexicologie* 2017, p. 171–190.

Condamines, A. et Narcy-Combes, J.-P. (2015). « La linguistique appliquée comme science située ». In : *Cultures de recherche en linguistique appliquée*. Sous la dir. de F. Carton, J.-P. Narcy-Combes, M.-F. Narcy-Combes et D. Toffoli. Riveneuve éditions. URL : https://hal.archives-ouvertes.fr/hal-01286390.

Condamines, A. et Picton, A. (2014). « Étude du fonctionnement des nominalisations déverbales dans un contexte de déspécialisation ». In : *Actes du 3e Congrès Mondial de Linguistique Française.* Berlin, Allemagne, p. 697–711.

Condamines, A. et Warnier, M. (2014). « Linguistic Analysis of Requirements of a Space Project and Their Conformity with the Recommendations Proposed by a Controlled Natural Language ». In : *Controlled Natural Language*. Lecture Notes in Computer Science. Springer, p. 33–43. ISBN : 978-3-319-10222-1 978-3-319-10223-8. DOI : 10.1007/978-3-319-10223-8_4. URL : https://link.springer.com/chapter/10.1007/978-3-319-10223-8_4.

Condamines, A. et Warnier, M. (2017). « Towards the creation of a CNL adapted to requirements writing by combining writing recommendations and spontaneous regularities : example in a space project ». In : *Language Resources and Evaluation* 51.1, p. 221–247. ISSN : 1574-020X, 1574-0218. DOI : 10.1007/s10579-016-9368-1. URL : https://link.springer.com/article/10.1007/s10579-016-9368-1.

Cornu, G. (1990). *Linguistique juridique*. Paris, France : Éditions Montchrestien.

Coseriu, E. (1952). *Sistema, norma y habla*. Universidad de la República. Facultad de Humanidades y Ciencias. Instituto e Filología. Departamento de Lingüística.

Coutherut, M. (2017). « L'électronicien francophone devant le manuel d´utilisation d´appareil rédigé ou à rédiger en anglais ». In : *ASp. la revue du GERAS* 71, p. 69–92. ISSN : 1246-8185. DOI : 10.4000/asp.4971. URL : http://journals.openedition.org/asp/4971.

Crabbé, B. et Candito, M. (2008). « Expériences d'analyse syntaxique statistique du français ». In : *15ème conférence sur le Traitement Automatique des Langues Naturelles-TALN'08*, pp–44.

Creissels, D. (2000). « L'emploi résultatif de être + participe passé en français ». In : *Cahiers Chronos* 6, p. 133–142.

Cutts, M. (2013). *Oxford Guide to Plain English*. Oxford, Royaume-Uni : OUP Oxford. ISBN : 978-0-19-164911-0.

Da Silva, J. (2014). « Langage contrôlé et traduction automatique pour le jargon médical : étude d'un corpus de textes de Traducteurs Sans Frontières ». Mémoire de Master. Université de Genève. URL : https://archive-ouverte.unige.ch/unige:55591.

Daille, B. (2003). « Conceptual structuring through term variations ». In : *ACL 2003 Workshop on Multiword Expressions : Analysis, Acquisition and Treatment*. Sous la dir. de F. Bond, A. Korhonen, D. MacCarthy et A. Villacicencio. Sapporo, Japon, p. 9–16. URL : https://hal.archives-ouvertes.fr/hal-00456518.

Davies, J. C. (1970). « Some Observations on the Early Development of Cicero's Plain Style ». In : *Latomus* 29.3, p. 729–736. ISSN : 0023-8856. URL : http://www.jstor.org/stable/41527737.

De Jong, M. D. et Lentz, L. R. (1996). « Expert Judgments Versus Reader Feedback : A Comparison of Text Evaluation Techniques ». In : *Journal of Technical Writing and Communication* 26.4, p. 507–519.

De Winter, J. C. et Dodou, D. (2010). « Five-point Likert items : t test versus Mann-Whitney-Wilcoxon ». In : *Practical Assessment, Research & Evaluation* 15.11.

Dendale, P. (1994). « Devoir épistémique, marqueur modal ou évidentiel ? » In : *Langue française* 102.1, p. 24–40. DOI : 10.3406/lfr.1994.5712. URL : https://www.persee.fr/doc/lfr_0023-8368_1994_num_102_1_5712.

Depecker, L. (1996). « Terminologie et standardisation ». In : *Hieronymus Complutensis* 3, p. 85–88.

Dick, J., Hull, E. et Jackson, K. (2005). *Requirements engineering*. Londres/Berlin : Springer.

Douay, C. (2003). « Des modalités de l'interlocution au système des modaux ». In : *Corela. Cognition, représentation, langage* 1. ISSN : 1638-5748. DOI : 10.4000/corela.636. URL : http://journals.openedition.org/corela/636.

Drouin, P. (2007). « Identification automatique du lexique scientifique transdisciplinaire ». In : *Revue française de linguistique appliquée* XII.2, p. 45–64. ISSN : 1386-1204. URL : https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2007-2-page-45.htm.

DuBay, W. H. (2004). *The Principles of Readability*. Impact Information. URL : https://eric.ed.gov/?id=ED490073.

Duranti, A. et Goodwin, C. (1997). *Rethinking context : Language as an interactive phenomenon*. T. 11. Cambridge : Cambridge University Press.

Emeneau, M. B. (1955). « India and Linguistics ». In : *Journal of the American Oriental Society* 75.3, p. 145–153. ISSN : 0003-0279. DOI : 10.2307/595166. URL : http://www.jstor.org/stable/595166.

Eveleens, J. et Verhoef, C. (2010). « The rise and fall of the Chaos report figures ». In : *IEEE Software* 27.1, p. 30–36. issn : 0740-7459. doi : 10.1109/MS.2009.154.

Fabre, C. et Lenci, A. (2015). « Distributional Semantics Today. Introduction to the special issue ». In : *Traitement Automatique des Langues* 56.2, p. 7–20. url : https://hal.archives-ouvertes.fr/hal-01259695.

Fabre, C., Hathout, N., Sajous, F. et Tanguy, L. (2014). « Ajuster l'analyse distributionnelle à un corpus spécialisé de petite taille ». In : 21e Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2014), p. 266–279. url : https://hal.archives-ouvertes.fr/hal-01022171/document.

Farfeleder, S., Moser, T., Krall, A., Stålhane, T., Zojer, H. et Panis, C. (2011a). « DODT : Increasing requirements formalism using domain ontologies for improved embedded systems development ». In : *Design and Diagnostics of Electronic Circuits & Systems (DDECS), 2011 IEEE 14th International Symposium on*. IEEE, p. 271–274.

Farfeleder, S., Moser, T., Krall, A., Stålhane, T., Omoronyia, I. et Zojer, H. (2011b). « Ontology-Driven Guidance for Requirements Elicitation ». In : *The Semanic Web : Research and Applications*. Extended Semantic Web Conference. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, p. 212–226. url : https://link.springer.com/chapter/10.1007/978-3-642-21064-8_15.

Fehr, E. et Fischbacher, U. (2004). « Social norms and human cooperation ». In : *Trends in Cognitive Sciences* 8.4, p. 185–190. issn : 1364-6613. doi : 10.1016/j.tics.2004.02.007.

Felker, D. B., Pickering, F., Charrow, V., Holland, M. et Redish, J. (1981). *Guidelines for Document Designers*.

Femmer, H., Fernández, D. M., Juergens, E., Klose, M., Zimmer, I. et Zimmer, J. (2014). « Rapid Requirements Checks with Requirements Smells : Two Case Studies ». In : *Proceedings of the 1st International Workshop on Rapid Continuous Software Engineering*. RCoSE 2014. New York, NY, USA : ACM, p. 10–19. url : http://doi.acm.org/10.1145/2593812.2593817.

Fish, S. E. (1980). *Is there a text in this class ? : The authority of interpretive communities*. Harvard University Press.

Flesch, R. F. (1944). « How basic is Basic English ? » In : *Harper's Magazine* 188.1126, p. 339–343.

Flesch, R. F. (1949). *Art of readable writing*. New York : Harper & Row.

Fleury, S. et Zimina, M. (2014). « Trameur : A framework for annotated text corpora exploration ». In : *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics : System Demonstrations*, p. 57–61.

Fludernik, M. (1991). « Shifters and deixis : Some reflections on Jakobson, Jespersen, and reference ». In : *Semiotica* 86.3, p. 193–230.

Forsbom, E. (2005). *Feature Extraction for Genre Classification*. URL : http://stp.lingfil.uu.se/~evafo/gslt/statmet/statmet05forsbom.pdf.

François, T. (2011). « Les apports du traitement automatique du langage à la lisibilité du français langue étrangère ». Thèse de doctorat. Université Catholique de Louvain.

Fuchs, C. (1995). « Ambigüité et ambivalence : le discret et le continu ». In : *Ambigüités/ambivalences : actes du colloque international organisé par le Centre de Recherches Iberiques et Ibero-Américaines de l'Université de Rouen*. Rouen : Université de Rouen, p. 7–24.

Fuchs, C. (2008). « L'incertitude interprétative dans l'activité de langage ». In : *Actes de Savoirs* 5, p. 41–57. URL : https://halshs.archives-ouvertes.fr/halshs-00340631.

Fuchs, N., Schwertel, U. et Schwitter, R. (1998). « Attempto Controlled English — Not Just Another Logic Specification Language ». In : *International Workshop on Logic Programming Synthesis and Transformation*. Lecture Notes in Computer Science. Berlin/Heidelberg : Springer, p. 1–20. ISBN : 978-3-540-65765-1 978-3-540-48958-0. DOI : 10.1007/3-540-48958-4_1. URL : https://link.springer.com/chapter/10.1007/3-540-48958-4_1.

Gadet, F. (2007). *La variation sociale en français*. Paris : Ophrys. ISBN : 978-2-7080-1154-0.

Ganier, F. (2002). « Évaluer l'efficacité des documents techniques procéduraux : un panorama des méthodes ». In : *Le travail humain* 65.1. ISSN : 0041-1868, 2104-3663. DOI : 10.3917/th.651.0001. URL : http://www.cairn.info/revue-le-travail-humain-2002-1-page-1.htm.

Ganier, F. (2006). « La révision de textes procéduraux ». In : *Langages* 164, p. 71–85. ISSN : 0458-726X. DOI : 10.3917/lang.164.0071. URL : https://www.cairn.info/revue-langages-2006-4-page-71.html.

García-Flores, J., Ivanova, E., Desclés, J.-P. et Djioua, B. (2006). « Annotation automatique de relations de contrôle dans des spécifications des besoins informatiques ». In : *Actes de TALN 2006*. Louvain, Belgique.

Gaudin, F. (2003). « Terminologie et politique linguistique ». In : *Socioterminologie. Une approche sociolinguistique de la terminologie*. Louvain-la-Neuve : De Boeck Supérieur.

Gaudin, F. (2005). « La socioterminologie ». In : *Langages* 157, p. 81–93. ISSN : 0458-726X. DOI : 10.3917/lang.157.0081. URL : https://www.cairn.info/revue-langages-2005-1-page-81.htm.

Gautier, L. (2002). « Terme, phraséoterme, phrasème : questions de délimitation en langue spécialisée ». In : *Le Continuum en linguistique*. Sousse, Tunisie, p. 153–172. URL : https://hal.archives-ouvertes.fr/hal-00201271.

Gibbons, J. (2003). *Forensic Linguistics : An Introduction to Language in the Justice System*. Hoboken : Wiley. isbn : 978-0-631-21247-8.

Gjesdal, A. M. (2008). « Étude sémantique du pronom ON dans une perspective textuelle et contextuelle ». Thèse de doctorat. The University of Bergen.

Gledhill, C. (2000). *Collocations in Science Writing*. Gunter Narr Verlag. isbn : 978-3-8233-4945-7.

Gledhill, C. et Kübler, N. (2015). « How Trainee Translators Analyse Lexico-Grammatical Patterns ». In : *Journal of Social Sciences* 11.3, p. 162–178.

Gledhill, C. et Kübler, N. (2016). « What can linguistic approaches bring to English for Specific Purposes ? » In : *ASp. la revue du GERAS* 69, p. 65–95. issn : 1246-8185. doi : 10.4000/asp.4804. url : http://journals.openedition.org/asp/4804.

Glinz, M. (2000). « Problems and Deficiencies of UML As a Requirements Specification Language ». In : *Proceedings of the 10th International Workshop on Software Specification and Design*. IWSSD '00. Washington, DC, USA : IEEE Computer Society, p. 11–. isbn : 978-0-7695-0884-9. url : http://dl.acm.org/citation.cfm?id=857171.857222.

Génova, G., Fuentes, J. M., Llorens, J., Hurtado, O. et Moreno, V. (2013). « A framework to measure and improve the quality of textual requirements ». In : *Requirements engineering* 18.1, p. 25–41.

Gougenheim, G. (1962). « Principes nouveaux pour l'enseignement du français ». In : *Esprit* 311, p. 585–594. issn : 0014-0759. url : http://www.jstor.org/stable/24258952.

Gough, P. B. (1965). « Grammatical transformations and speed of understanding ». In : *Journal of Verbal Learning and Verbal Behavior* 4.2, p. 107–111. issn : 0022-5371. doi : 10.1016/S0022-5371(65)80093-7. url : http://www.sciencedirect.com/science/article/pii/S0022537165800937.

Guță, A. (2006). « La focalisation dans le discours législatif ». In : *Analele Universitatea din Craiova* 2, p. 96–102.

Guilbert, T. (2014). « Les genres du discours dans l'articulation des approches qualitatives et quantitatives ». In : *Corela. Cognition, représentation, langage* (HS-15). issn : 1638-5748. doi : 10.4000/corela.3577. url : http://journals.openedition.org/corela/3577.

Gumperz, J. (2009). « The Speech Community ». In : Duranti, A. *Linguistic Anthropology : A Reader*. Hoboken : Wiley-Blackwell, p. 66–73. isbn : 978-1-4051-2633-5.

Guéry, F. (1995). « Leibniz et les langues ». In : *Revue Philosophique de la France et de l'Étranger* 185.2, p. 239–250. issn : 0035-3833. url : http://www.jstor.org/stable/41097446.

Gylling, M. (2013). *The Structure of Discourse : A Corpus-Based Cross-Linguistic Study*. 276 p. isbn : 978-87-92977-88-5.

Halliday, M. A. K. (2004). *The Language of Science*. Londres/New York : Continuum.

Halliday, M. A. K. (1974). « La base fonctionnelle du langage ». In : *Langages* 8.34, p. 54–73. doi : 10.3406/lgge.1974.2259. url : https://www.persee.fr/doc/lgge_0458-726x_1974_num_8_34_2259.

Halliday, M. A. K. (1978). *Language as social semiotic : the social interpretation of language and meaning*. Edward Arnold. isbn : 978-0-7131-5967-7.

Halliday, M. A. K. et Hasan, R. (1976). *Cohesion in English*. Londres, Royaume-Uni : Longman.

Harber, J. R. (1979). « Syntactic Complexity : A Necessary Ingredient in Predicting Readability ». In : *Journal of Learning Disabilities* 12.7. doi : 10.1177/002221947901200702. url : http://journals.sagepub.com/doi/pdf/10.1177/002221947901200702.

Hargis, G. (2000). « Readability and Computer Documentation ». In : *ACM Journal of Computer Documentation* 24.3, p. 122–131. issn : 1527-6805. doi : 10.1145/344599.344634. url : http://doi.acm.org/10.1145/344599.344634.

Harris, W. V. (1988). *Interpretive Acts : In Search of Meaning*. Oxford : Oxford University Press UK.

Hathout, N., Namer, F. et Dal, G. (2002). « An experimental constructional database : the MorTAL project ». In : *Many morphologies*, p. 178–209.

Hatier, S. (2013). « Extraction des mots simples du lexique scientifique transdisciplinaire dans les écrits de sciences humaines : une première expérimentation ». In : *Actes de la conférence TALN-RECITAL*, p. 85–100.

Heid, U. (2001). « Collocations in Sublanguage Texts : Extraction from Corpora ». In : Wright, S. E. et Budin, G. *Handbook of Terminology Management : Application-oriented terminology management*. T. 2. Amsterdam/Philadelphia : John Benjamins Publishing Company, p. 788–808.

Heid, U. (2008). « Computational phraseology. An overview ». In : *Phraseology : An Interdisciplinary Perspective ; Granger, S., Meunier, F., Eds*, p. 337–360.

Heid, U. et Freibott, G. (1991). « Collocations dans une base de données terminologique et lexicale ». In : *Meta : Journal des traducteurs/Meta : Translators' Journal* 36.1, p. 77–91.

Heiden, S., Magué, J.-P. et Pincemin, B. (2010). « TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement ». In : 10th International Conference on the Statistical Analysis of Textual Data - JADT 2010. T. 2. Edizioni Universitarie di Lettere Economia Diritto, p. 1021–1032. url : https://halshs.archives-ouvertes.fr/halshs-00549779/document.

Heilman, M., Collins-Thompson, K., Callan, J. et Eskenazi, M. (2007). « Combining lexical and grammatical features to improve readability measures for first and second language texts ». In : *Human Language Technologies 2007 : The Conference of the North American Chapter of the Association for Computational Linguistics ; Proceedings of the Main Conference*, p. 460–467.

Hirst, G. (1997). « Context as a spurious concept ». In : *Context in Knowledge Representation and Natural Language, AAAI Fall Symposium*. Cambridge : MIT.

Hjelmslev, L. (1942). « Langue et parole ». In : *Cahiers Ferdinand de Saussure* 2, p. 29–44. issn : 0068-516X. url : http://www.jstor.org/stable/27757874.

Huijsen, W.-O. (1998). « Controlled Language - An Introduction ». In : *Proceedings of the Second International Workshop on Controlled Language Applications - CLAW 98*. Sous la dir. de T. Mitamura. Pittsburgh, Pennsylvania : Language Technologies Institute, Carnegie Mellon University, p. 1–15.

Humbley, J. (2001). « Quelques enjeux de la dénomination en terminologie ». In : *Cahiers de praxématique* 36, p. 93–115. issn : 0765-4944. url : http://journals.openedition.org/praxematique/338.

Hupet, M. et Costermans, J. (1976). « Un passif : Pour quoi faire ?, Quinze années de travaux psycholinguistiques ». In : *La Linguistique* 12.2, p. 3–26. issn : 0075-966X. url : http://www.jstor.org/stable/30248325.

Hymes, D. (2005). « Models of the Interaction of Language and Social Life : Toward a Descriptive Theory ». In : Kiesling, S. F. et Paulston, C. *Intercultural Discourse and Communication : The Essential Readings*. Hoboken : John Wiley & Sons, p. 4–16. isbn : 978-0-470-75828-1.

IEEE (1990). « Standard Glossary of Software Engineering Terminology ». In : *IEEE*, p. 1–83.

ISO/IEC/IEEE29148 (2011). « International Standard - Systems and software engineering – Life cycle processes – Requirements engineering ». In : *ISO/IEC/IEEE 29148 :2011(E)*, p. 1–94. doi : 10.1109/IEEESTD.2011.6146379.

ISO/TS24620-1 (2015). « International Standard - Language resource management – Controlled natural language (CNL) – Part 1 : Basic concepts and principles ». In : *ISO/TS 24620-1 :2014*, p. 1–9.

International Council on Systems Engineering (INCOSE) (2011). *Guide for Writing Requirements. Version 1*.

Jacques, M.-P. et Aussenac-Gilles, N. (2006). « Variabilité des performances des outils de TAL et genre textuel. » In : *TAL* 47.1, p. 11–32.

Jahchan, N., Condamines, A. et Cannesson, E. (2016). « To What Extent Does Text Simplification Entail a More Optimized Comprehension in Human-Oriented CNLs ? » In : *International Workshop on Controlled Natural Language*. Lecture Notes in Computer Science. Springer, p. 69–80. isbn : 978-3-319-41497-3 978-

3-319-41498-0. DOI : 10.1007/978-3-319-41498-0_7. URL : https://link.springer.com/chapter/10.1007/978-3-319-41498-0_7.

JAKOBSON, R. (1960). « Linguistics and poetics ». In : *Style in language*. MA : MIT Press, p. 350–377.

JOHANNESSEN, V. (2012). « CESAR-text vs. boilerplates : What is more efficient - requirements written as free text or using boilerplates (templates) ? » Master's Thesis. Institutt for datateknikk og informasjonsvitenskap.

KAIL, M. (1979). « Coréférence et thématisation ». In : *L'Année psychologique* 79.2, p. 411–427. DOI : 10.3406/psy.1979.28277. URL : https://www.persee.fr/doc/psy_0003-5033_1979_num_79_2_28277.

KAMSTIES, E. et PEACH, B. (2000). « Taming ambiguity in natural language requirements ». In : *Proceedings of the Thirteenth International Conference on Software and Systems Engineering and Applications*.

KANDEL, L. et MOLES, A. (1958). « Application de l'indice de Flesch à la langue française ». In : *Cahiers Études de Radio-Télévision* 19, p. 253–274.

KARWOWSKI, W. (2005). « Ergonomics and human factors : the paradigms for science, engineering, design, technology and management of human-compatible systems ». In : *Ergonomics* 48.5, p. 436–463. ISSN : 0014-0139. DOI : 10.1080/00140130400029167. URL : https://doi.org/10.1080/00140130400029167.

KÜBLER, N. et FRÉROT, C. (2003). « Verbs in specialised corpora : From Manual Corpus-Based Description to Automatic Extraction in an English-French Parallel Corpus ». In : *UCREL Technical Paper* 16, p. 429–438.

KÜBLER, N. et PECMAN, M. (2012). « The ARTES bilingual LSP dictionary : from collocation to higher order phraseology ». In : GRANGER, S. et PAQUOT, M. *Electronic Lexicography*. Oxford : Oxford University Press. ISBN : 978-0-19-174596-6. URL : http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199654864.001.0001/acprof-9780199654864-chapter-10.

KENNEDY, G. A. (1957). « Theophrastus and Stylistic Distinctions ». In : *Harvard Studies in Classical Philology* 62, p. 93–104. ISSN : 0073-0688. DOI : 10.2307/310969. URL : http://www.jstor.org/stable/310969.

KERBRAT-ORECCHIONI, C. (2000). « Est-il bon, est-il méchant ? » In : WAUTHION, M. et SIMON, A.-C. *Politesse et idéologie : rencontres de pragmatique et de rhétorique conversationnelles*. Louvain : Peeters Publishers, p. 21–35. ISBN : 978-90-429-0949-6.

KERBRAT-ORECCHIONI, C. (2012). « Le contexte revisité ». In : *Corela. Cognition, représentation, langage* (HS-11). ISSN : 1638-5748. DOI : 10.4000/corela.2627. URL : http://journals.openedition.org/corela/2627.

KILGARRIFF, A. (2005). « Language is never, ever, ever, random ». In : *Corpus linguistics and linguistic theory* 1.2, p. 263–276.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. et Suchomel, V. (2014). « The Sketch Engine : ten years on ». In : *Lexicography* 1.1, p. 7–36.

Kitagawa, C. et Lehrer, A. (1990). « Impersonal uses of personal pronouns ». In : *Journal of Pragmatics* 14.5, p. 739–759.

Kittredge, R. (2005). « Sublanguages and Controlled Languages ». In : Mitkov, R. *The Oxford Handbook of Computational Linguistics*. Oxford, Royaume-Uni : Oxford University Press, p. 430–447. isbn : 978-0-19-174357-3. url : http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199276349.001.0001/oxfordhb-9780199276349.

Kittredge, R. et Lehrberger, J. (1982). *Sublanguage : Studies of Language in Restricted Semantic Domains*. Berlin : Walter de Gruyter. isbn : 978-3-11-008244-9.

Klare, G. (1988). « The formative years ». In : *Readability : Its past, present, and future*. Sous la dir. de B. L. Zakaluk et S. J. Samuels. Newark, Delaware : International Reading Association, p. 27–47.

Kleiber, G. (1999). « Il y a contexte et contexte ». In : *L'emprise du sens. Structures linguistiques et interprétations*. Sous la dir. de M. Plénat, M. Aurnague, A. Condamines, J.-P. Maurel, C. Molinier et C. Muller. Amsterdam / Atlanta : Rodopi, p. 167–181.

Knecht, H. H. (1981). *La Logique chez Leibniz : Essai sur le rationalisme baroque*. Lausanne : L'Age d'Homme.

Koehn, P. (2005). « Europarl : A parallel corpus for statistical machine translation ». In : *MT summit*. T. 5, p. 79–86.

Kraif, O. et Diwersy, S. (2012). « Le Lexicoscope : un outil pour l'étude de profils combinatoires et l'extraction de constructions lexico-syntaxiques ». In : *Proceedings of the Joint Conference JEP-TALN-RECITAL 2012, volume 2 : TALN*, p. 399–406.

Krisch, J. et Houdek, F. (2015). « The myth of bad passive voice and weak words. An empirical investigation in the automotive industry ». In : *Requirements Engineering Conference (RE), 2015 IEEE 23rd International*. IEEE, p. 344–351.

Krisch, J., Dick, M., Jauch, R. et Heid, U. (2016). « A Lexical Resource for the Identification of "Weak Words" in German Specification Documents ». In : *LREC*, p. 2846–2850.

Kuhn, T. (2009). « An evaluation framework for controlled natural languages ». In : *International Workshop on Controlled Natural Language*. Springer, p. 1–20.

Kuhn, T. (2014). « A survey and classification of controlled natural languages ». In : *Computational Linguistics* 40.1, p. 121–170.

Labov, W. (1976). *Sociolinguistique*. Paris : Éditions de Minuit.

Lafon, P. et Salem, A. (1983). « L'inventaire des segments répétés d'un texte ». In : *Mots. Les langages du politique* 6.1, p. 161–177. doi : 10.3406/mots.1983.1101. url : https://www.persee.fr/doc/mots_0243-6450_1983_num_6_1_1101.

Lami, G., Gnesi, S., Fabbrini, F., Fusani, M. et Trentanni, G. (2004). « An automatic tool for the analysis of natural language requirements ». In : *Informe técnico, CNR Information Science and Technology Institute*.

Lamsweerde, A. v. (2001). « Goal-oriented requirements engineering : a guided tour ». In : *Proceedings Fifth IEEE International Symposium on Requirements Engineering*, p. 249–262. doi : 10.1109/ISRE.2001.948567.

Landhausser, M., Korner, S. J., Tichy, W. F., Keim, J. et Krisch, J. (2015). « DeNom : a tool to find problematic nominalizations using NLP ». In : *2015 IEEE Second International Workshop on Artificial Intelligence for Requirements Engineering (AIRE)*, p. 1–8. doi : 10.1109/AIRE.2015.7337623.

Landragin, F., Poibeau, T. et Victorri, B. (2012). « ANALEC : a new tool for the dynamic annotation of textual data ». In : *International Conference on Language Resources and Evaluation (LREC 2012)*, p. 357–362.

Le Pesant, D. (2002). « La détermination dans les anaphores fidèles et infidèles ». In : *Langages* 36.145, p. 39–59. doi : 10.3406/lgge.2002.906. url : https://www.persee.fr/doc/lgge_0458-726x_2002_num_36_145_906.

Le Querler, N. (2004). « Les modalités en français ». In : *Revue belge de Philologie et d'Histoire* 82.3, p. 643–656. doi : 10.3406/rbph.2004.4850. url : https://www.persee.fr/doc/rbph_0035-0818_2004_num_82_3_4850.

Lee, D. (2002). « Genres, registers, text types, domains and styles : clarifying the concepts and navigating a path through the BNC jungle ». In : *Language and Computers* 42, p. 247–292.

Lerat, P. (1995). *Les langues spécialisées*. Paris : Presses universitaires de France.

Lerat, P. (1997). « Approches linguistiques des langues spécialisées ». In : *ASp. la revue du GERAS* 15, p. 1–10. issn : 1246-8185. doi : 10.4000/asp.2926. url : http://journals.openedition.org/asp/2926.

Lopez, S. (2013). « Norme(s) et usage(s) langagiers : le cas des communications pilote-contrôleur en anglais ». Thèse de doctorat. Université Toulouse le Mirail - Toulouse II. url : https://tel.archives-ouvertes.fr/tel-00944009.

Lux, V. (1998). « Elaboration d'un francais rationalisé étendu modulaire (FREM) pour les manuels de maintenance d'aéronefs ». Thèse de doctorat. Paris 7.

Malrieu, D. (2007). « Contribution à une linguistique néo-saussurienne des genres de la parole (2) : analyse des valeurs d'indexicalité interlocutoire de on selon les genres textuels ». In : *Linx. Revue des linguistes de l'université Paris X Nanterre* 56, p. 157–178. issn : 0246-8743. doi : 10.4000/linx.377. url : http://journals.openedition.org/linx/377.

Malrieu, D. et Rastier, F. (2000). « Genres et variations morphosyntaxiques ». In : *Traitement Automatique des Langues* 42.2, p. 547–577. URL : https://halshs.archives-ouvertes.fr/halshs-00161624.

Marlaud, S. (2013). « Les langues artificielles sont-elles des langues ? Étude contrastive de l'espéranto et de la caractéristique universelle ». In : *Syntaxe et sémantique* 14, p. 85–117. ISSN : 1623-6742. DOI : 10.3917/ss.014.0085. URL : https://www.cairn.info/revue-syntaxe-et-semantique-2013-1-page-85.htm.

Marsden, P. V. et Wright, J. D. (2010). *Handbook of Survey Research*. 2e édition. Bingley, Royaume-Uni : Emerald Group Publishing. 905 p. ISBN : 978-1-84855-224-1.

Mathieu, Y. (1993). « Quelques passifs avec agent obligatoire ». In : *Langages* 27.109, p. 35–36.

Mavin, A., Wilkinson, P., Harwood, A. et Novak, M. (2009). « Easy Approach to Requirements Syntax (EARS) ». In : *2009 17th IEEE International Requirements Engineering Conference*. 2009 17th IEEE International Requirements Engineering Conference, p. 317–322. DOI : 10.1109/RE.2009.9.

Mel'čuk, I. (2013). « Tout ce que nous voulions savoir sur les phrasèmes, mais... » In : *Cahiers de lexicologie* 102.1, p. 129–149.

Merdy, E. (2016). « Identifier et catégoriser l'ambiguïté dans les spécifications techniques de conceptions de systèmes ». In : *JEP-TALN-RÉCITAL 2016*. Paris, France. URL : https://hal.archives-ouvertes.fr/hal-01351067.

Merdy, E., Kang, J. et Tanguy, L. (2016). « Identification of fuzzy and underspecified terms in technical documents : an experiment with distributional semantics ». In : *Atelier "Risque et TAL" dans le cadre de la conférence TALN*. Paris, France. URL : https://hal.archives-ouvertes.fr/hal-01365926.

Meyer, B. (1993). « On formalism in specifications ». In : *Program Verification*. Springer, p. 155–189.

Nair, D. K. et Somé, S. S. (2006). « A Formal Approach to Requirement Verification. » In : *SEDE*, p. 148–153.

Naumchev, A. et Meyer, B. (2017). « Seamless requirements ». In : *Computer Languages, Systems & Structures* 49, p. 119–132.

Nef, F. (1986). *Sémantique de la référence temporelle en français moderne*. T. 32. Herbert Lang.

Noailly, M. (1996). « Le vide des choses ». In : *Cahiers de praxématique* 27, p. 73–90. ISSN : 0765-4944. URL : http://journals.openedition.org/praxematique/2999.

O'Brien, S. (2003). « Controlling Controlled English. An analysis of several controlled language rule sets ». In : *Proceedings of EAMT-CLAW*, p. 105–114.

O'Brien, S. (2006). « Machine-translatability and post-editing effort : an empirical study using translog and choice network analysis ». Thèse de doctorat. Dublin City University. School of Applied Language et Intercultural Studies. URL : http://doras.dcu.ie/18118/.

O'Brien, S. (2010). « Controlled language and readability ». In : *Translation and cognition* 15, p. 143–168.

O'Hayre, J. (1966). *Gobbledygook has gotta go*. U.S. Dept. of the Interior, Bureau of Land Management.

O'Leary, N., Braines, D., Preece, A. et Webberley, W. (2017). « Conversational homes ». In : *9th International Conference on Advanced Cognitive Technologies and Applications (COGNITIVE'17)*.

O'Regan, G. (2006). « Z Specification Language ». In : *Mathematical Approaches to Software Quality*. Springer, London, p. 75–91. ISBN : 978-1-84628-242-3 978-1-84628-435-9. DOI : 10.1007/1-84628-435-9_4. URL : https://link.springer.com/chapter/10.1007/1-84628-435-9_4.

Ogden, C. K. (1944). « Will Basic English Become the Second Language ? » In : *Public Opinion Quarterly* 8.1, p. 3–9.

Orwell, G. (1946). *Politics and the English Language*. URL : https://iis.berkeley.edu/sites/default/files/pdf/politics_and_the_english_language.pdf.

Ownby, R. L. (2005). « Influence of Vocabulary and Sentence Complexity and Passive Voice on the Readability of Consumer-Oriented Mental Health Information on the Internet ». In : *AMIA Annual Symposium Proceedings* 2005, p. 585–588. ISSN : 1942-597X. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1560876/.

Paltridge, B. (1996). « Genre, text type, and the language learning classroom ». In : *ELT Journal* 50.3, p. 237–243. ISSN : 0951-0893. DOI : 10.1093/elt/50.3.237. URL : https://academic.oup.com/eltj/article/50/3/237/2924374.

Pecman, M. (2007). « Approche onomasiologique de la langue scientifique générale ». In : *Revue française de linguistique appliquée* XII.2, p. 79–96. ISSN : 1386-1204. URL : https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2007-2-page-79.html.

Perreault, W. D. (1975). « Controlling order-effect bias ». In : *The Public Opinion Quarterly* 39.4, p. 544–551.

Picton, A. (2009). « Diachronie en langue de spécialité. Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus. Un exemple appliqué au domaine spatial. » Thèse de doctorat. Université Toulouse le Mirail-Toulouse II.

Poudat, C. (2006). « Etude contrastive de l'article scientifique de revue linguistique dans une perspective d'analyse des genres ». Thèse de doctorat. Université d'Orléans.

Poudat, C. et Rinck, F. (2006). « Contrastes internes et variations stylistiques du genre de l'article scientifique de linguistique ». In : 8èmes Journées d'analyse des données textuelles (JADT), p. 785–796. URL : https://halshs.archives-ouvertes.fr/halshs-00418166.

Périnet, A. et Hamon, T. (2015). « Analyse distributionnelle appliquée aux textes de spécialité - Réduction de la dispersion des données par abstraction des contextes ». In : *Traitement Automatique des Langues* 56.2, p. 77–102. URL : https://hal.archives-ouvertes.fr/hal-01259695.

Quiniou, S., Cellier, P., Charnois, T. et Legallois, D. (2012a). « Fouille de données pour la stylistique : cas des motifs séquentiels émergents ». In : *Journées Internationales d'Analyse Statistique des Données Textuelles (JADT'12)*, p. 821–833.

Quiniou, S., Cellier, P., Charnois, T. et Legallois, D. (2012b). « What about sequential data mining techniques to identify linguistic patterns for stylistics ? » In : *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, p. 166–177.

Rabatel, A. (2001). « La valeur de « on » pronom indéfini/pronom personnel dans les perceptions représentées ». In : *L'information grammaticale* 88.1, p. 28–32. DOI : 10.3406/igram.2001.2726. URL : https://www.persee.fr/doc/igram_0222-9838_2001_num_88_1_2726.

Rajan, A. et Wahl, T. (2013). *CESAR : Cost-efficient Methods and Processes for Safety-relevant Embedded Systems*. 978-3709113868. Springer.

Randall, J. H. et Graf, L. R. (2014). « Linguistics meets "legalese" : syntax, semantics, and jury instruction reform ». In : *LSA Annual Meeting Extended Abstracts* 5, p. 25–1–5. ISSN : 2377-3367. DOI : 10.3765/exabs.v0i0.2381. URL : https://journals.linguisticsociety.org/proceedings/index.php/ExtendedAbs/article/view/2381.

Rasmussen, K. W. (1999). « Le passif vu à travers le contrat bilatéral : fonction de focalisation ». In : *La langue, les signes et les êtres : actes du colloque de l'Institut d'Etudes Romanes de l'Université de Copenhague, le 3 octobre 1998*. T. 45. Museum Tusculanum Press, p. 321.

Rastier, F. (1995). « Le terme : entre ontologie et linguistique ». In : *La banque des mots* 7, p. 35–65.

Rastier, F. (1998). « Le problème épistémologique du contexte et le statut de l'interprétation dans les sciences du langage ». In : *Langages* 32.129, p. 97–111. DOI : 10.3406/lgge.1998.2149. URL : https://www.persee.fr/doc/lgge_0458-726x_1998_num_32_129_2149.

Rastier, F. et Pincemin, B. (1999). « Des genres à l'intertexte ». In : *Cahiers de praxématique* 33, p. 83–111. ISSN : 0765-4944. URL : http://journals.openedition.org/praxematique/1974.

Rebeyrolle, J. et Tanguy, L. (2000). « Repérage automatique de structures linguistiques en corpus : le cas des énoncés définitoires ». In : *Cahiers de Gram-*

maire 25, p. 153–174. URL : https://halshs.archives-ouvertes.fr/halshs-01322256/document.

REDISH, J. et SELZER, J. (1985). « The Place of Readability Formulas in Technical Communication ». In : *Technical Communication* 32.4, p. 46–52. ISSN : 0049-3155. URL : http://www.jstor.org/stable/43094562.

RENOUF, A. et SINCLAIR, J. (1991). « Collocational frameworks in English ». In : *English corpus linguistics*, p. 128–143.

REY, A. (1972). « Usages, jugements et prescriptions linguistiques ». In : *Langue française* 16.1, p. 4–28. DOI : 10.3406/lfr.1972.5701. URL : https://www.persee.fr/doc/lfr_0023-8368_1972_num_16_1_5701.

RICHARD, I. (2006). « L'évolution de l'emploi de shall, de must et du présent simple dans le discours juridique normatif dans le cadre du Plain Language Movement ». In : *ASp. la revue du GERAS* 49, p. 137–153. ISSN : 1246-8185. DOI : 10.4000/asp.742. URL : http://journals.openedition.org/asp/742.

RICHAUDEAU, F. (1969). « Les structures des phrases ». In : *Communication & Langages* 1.1, p. 19–25. DOI : 10.3406/colan.1969.3706. URL : https://www.persee.fr/doc/colan_0336-1500_1969_num_1_1_3706.

ROGERS, M. (2000). « Genre and terminology ». In : TROSBORG, A. *Analysing Professional Genres*. Amsterdam/Philadelphia : John Benjamins, p. 3–24.

ROLAND, D., DICK, F. et ELMAN, J. L. (2007). « Frequency of basic English grammatical structures : A corpus analysis ». In : *Journal of Memory and Language* 57.3, p. 348–379. ISSN : 0749-596X. DOI : 10.1016/j.jml.2007.03.002. URL : http://www.sciencedirect.com/science/article/pii/S0749596X07000307.

ROUSKI, M. (2015). « Le passif de l'autorité ». In : *Corela. Cognition, représentation, langage* 13. ISSN : 1638-5748. DOI : 10.4000/corela.3968. URL : http://journals.openedition.org/corela/3968.

RUPP, C. et JOPPICH, R. (2010). *Templates – Construction Plans for Requirements and for More*. URL : https://www.sophist.de/en/information-pool/downloads/open-download-area/?tx_mtdownloads_pi1%5Baction%5D=getviewclickeddownload&tx_mtdownloads_pi1%5Buid%5D=275&tx_mtdownloads_pi1%5Bcid%5D=27689&no_cache=1.

RYAN, R. (2011). « Traduction technique : les langues contrôlées au service de l'ergonomie documentaire ». In : *ILCEA. Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie* 14. ISSN : 1639-6073. URL : http://journals.openedition.org/ilcea/1090.

RYCHLÝ, P. (2007). « Manatee/bonito-a modular corpus manager ». In : *1st Workshop on Recent Advances in Slavonic Natural Language Processing*, p. 65–70.

SADOUN, D. (2014). « Des spécifications en langage naturel aux spécifications formelles via une ontologie comme modèle pivot ». Thèse de doctorat. Univer-

sité Paris Sud - Paris XI. URL : https://tel.archives-ouvertes.fr/tel-01060540/document.

Sahraoui, H.-T., Holat, P., Cellier, P., Charnois, T. et Ferré, S. (2017). « Exploration of Textual Sequential Patterns ». In : *14th International Conference on Formal Concept Analysis*, p. 99.

Saint-Dizier, P. (2018). « Mining incoherent requirements in technical specifications : Analysis and implementation ». In : *Data & Knowledge Engineering*. ISSN : 0169-023X. DOI : 10.1016/j.datak.2018.05.006. URL : http://www.sciencedirect.com/science/article/pii/S0169023X17305232.

Salvage, P. (2016). *Droit pénal général*. 8e édition. Grenoble : Presses universitaires de Grenoble. ISBN : 978-2-7061-2605-5.

Sateli, B., Angius, E., Rajivelu, S. S. et Witte, R. (2012). « Can text mining assistants help to improve requirements specifications ? » In : *Mining Unstructured Data (MUD 2012), Canada*.

Saussure, F. (1995). *Cours de linguistique générale*. Sous la dir. de C. Bally, C.-A. Sechehaye, A. Riedlinger et T. De Mauro. Grande bibliothèque Payot. Paris : Payot. ISBN : 978-2-228-88942-1.

Schmid, H. (1995). « Treetagger : a language independent part-of-speech tagger ». In : *Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart* 43, p. 28.

Searle, J. R. (1972). *Les Actes de Langage. Essai de Philosophie du Langage*. Paris : Hermann.

Shubert, S. K., Spyridakis, J. H., Holmback, H. K. et Coney, M. B. (1995). « The Comprehensibility of Simplified English in Procedures ». In : *Journal of Technical Writing and Communication* 25.4, p. 347–369. ISSN : 0047-2816. DOI : 10.2190/WG69-D74B-4DLL-2WBK. URL : https://doi.org/10.2190/WG69-D74B-4DLL-2WBK.

Siouffi, G. et Steuckardt, A. (2007). *Les linguistes et la norme : aspects normatifs du discours linguistique*. Peter Lang. ISBN : 978-3-03911-216-6.

Somers, H. (1998). « An Attempt to Use Weighted Cusums to Identify Sublanguages ». In : *Proceedings of the Joint Conferences on New Methods in Language Processing and Computational Natural Language Learning*. Stroudsburg, PA, USA : Association for Computational Linguistics, p. 131–139. ISBN : 978-0-7258-0634-7. URL : http://dl.acm.org/citation.cfm?id=1603899.1603922.

Sperber, D. et Wilson, D. (1989). *La pertinence : communication et cognition*. Les Editions de Minuit. 396 p. ISBN : 978-2-7073-1305-8.

Stage, L. (2002). « Les modalités épistémique et déontique dans les énoncés au futur (simple et composé) ». In : *Revue Romane* 1. ISSN : 1600-0811. URL : https://tidsskrift.dk/revue_romane/article/view/31059.

Stamatatos, E., Fakotakis, N. et Kokkinakis, G. (2000). « Automatic text categorization in terms of genre and author ». In : *Computational linguistics* 26.4, p. 471–495.

Stewart, K. M. (1998). « Effect of AECMA simplified English on the comprehension of aircraft maintenance procedures by non-native English speakers ». Master Thesis. University of British Columbia.

Suchowolec, K. (2014). « Are Style Guides Controlled Languages ? » In : *International Workshop on Controlled Natural Language*. Lecture Notes in Computer Science. Springer, p. 112–122. ISBN : 978-3-319-10222-1 978-3-319-10223-8. DOI : 10.1007/978-3-319-10223-8_11. URL : https://link.springer.com/chapter/10.1007/978-3-319-10223-8_11.

Sullivan, G. M. et Artino, A. R. (2013). « Analyzing and Interpreting Data From Likert-Type Scales ». In : *Journal of Graduate Medical Education* 5.4, p. 541–542. ISSN : 1949-8349. DOI : 10.4300/JGME-5-4-18. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3886444/.

Swales, J. (2016). « Reflections on the concept of discourse community ». In : *ASp. la revue du GERAS* 69, p. 7–19. ISSN : 1246-8185. DOI : 10.4000/asp.4774. URL : http://journals.openedition.org/asp/4774.

Tanguy, L. et Hathout, N. (2001). « Webaffix : un outil d'acquisition morphologique dérivationnelle à partir du Web ». In : *9e Conférence Annuelle sur le Traitement Automatique des Langues Naturelles (TALN-2002)*, p. 254. URL : https://halshs.archives-ouvertes.fr/halshs-00287679/document.

Tanguy, L., Sajous, F. et Hathout, N. (2015). « Évaluation sur mesure de modèles distributionnels sur un corpus spécialisé : comparaison des approches par contextes syntaxiques et par fenêtres graphiques ». In : *Traitement Automatique des Langues* 56.2. URL : https://halshs.archives-ouvertes.fr/halshs-01322234/document.

Tanguy, L. et Tulechki, N. (2009). « Sentence complexity in french : A corpus-based approach ». In : *Intelligent information systems (IIS)*, p. 131–145.

Temnikova, I., Baumgartner, W. A., Hailu, N. D., Nikolova, I., McEnery, T., Kilgarriff, A., Angelova, G. et Cohen, K. B. (2014). « Sublanguage Corpus Analysis Toolkit : A tool for assessing the representativeness and sublanguage characteristics of corpora. » In : *LREC*, p. 1714–1718.

Todiraşcu, A., Heid, U., Ştefănescu, D., Tufiş, D., Gledhill, C., Weller, M. et Rousselot, F. (2008). « Vers un dictionnaire de collocations multilingue ». In : *Cahiers de linguistique* 33.1, p. 161–186.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. T. 6. Amsterdam/Philadelphia : John Benjamins Publishing.

Tommila, T. et Pakonen, A. (2014). *Controlled natural language requirements in the design and analysis of safety critical I&C systems*.

Tutin, A. (2007a). « Modélisation linguistique et annotation des collocations : une application au lexique transdisciplinaire des écrits scientifiques ». In : *Formaliser les langues avec l'ordinateur : actes des sixièmes, Sofia 2003, et septièmes, Tours 2004, journées Intex-Nooj* 3, p. 189.

Tutin, A. (2007b). « Traitement sémantique par analyse distributionnelle des noms transdisciplinaires des écrits scientifiques ». In : *Actes de Traitement Automatique des Langues Naturelles (TALN)*, p. 283–292.

Urieli, A. (2013). « Robust French syntax analysis : reconciling statistical methods and linguistic knowledge in the Talismane toolkit ». Thèse de doctorat. Université Toulouse le Mirail - Toulouse II. URL : https://tel.archives-ouvertes.fr/tel-01058143/document.

Urieli, A. (2015). « Stratégies pour l'étiquetage et l'analyse syntaxique statistique de phénomènes difficiles en français : études de cas avec Talismane. » In : *Traitement Automatique des Langues* 56.1.

Vetters, C. (2012). « Modalité et évidentialité dans pouvoir et devoir : typologie et discussions ». In : *Langue française* 173, p. 31–47. ISSN : 0023-8368. DOI : 10.3917/lf.173.0031. URL : https://www.cairn.info/revue-langue-francaise-2012-1-page-31.html.

Victorri, B. et Fuchs, C. (1996). *La polysémie - construction dynamique du sens*. Hermès. URL : https://halshs.archives-ouvertes.fr/halshs-00713735/document.

Vinet, M.-T. (1978). « Le passif et les nominalisations ». In : *Cahier de linguistique* 8, p. 457–477.

Volanschi, A. et Kübler, N. (2010). « Building an electronic combinatory dictionary as a writing aid tool for researchers in biology ». In : *e-Lexicography in the 21st Century : New Applications, New Challenges*, p. 343–355.

Wandel, A. (2015). « How many people speak esperanto ? Esperanto on the Web ». In : *Interdisciplinary Description of Complex Systems : INDECS* 13.2, p. 318–321.

Warnier, M. (2015). « How can corpus linguistics help improve requirements writing ? Specifications of a space project as a case study ». In : *2015 IEEE 23rd International Requirements Engineering Conference (RE)*, p. 388–392. DOI : 10.1109/RE.2015.7320456.

Warnier, M. et Condamines, A. (2015). « A Methodology for Identifying Terms and Patterns Specific to Requirements as a Textual Genre Using Automated Tools ». In : Terminology and Artificial Intelligence (TIA'2015). URL : https://halshs.archives-ouvertes.fr/halshs-01379533/document.

Warnier, M. et Condamines, A. (2016). « Analyse d'un corpus d'exigences pour améliorer la rédaction des spécifications de systèmes spatiaux au CNES ». In : *Journées d'Analyse des Données Textuelles*.

Warnier, M. et Condamines, A. (2017a). « A Case Study on Evaluating the Relevance of Some Rules for Writing Requirements Through an Online Survey ». In : *2017 IEEE 25th International Requirements Engineering Conference (RE)*, p. 243–252. doi : 10.1109/RE.2017.11.

Warnier, M. et Condamines, A. (2017b). « Improving Requirement Boilerplates Using Sequential Pattern Mining ». In : *Europhras 2017*. Londres, Royaume-Uni. url : https://hal.archives-ouvertes.fr/hal-01672313.

Williams, C. (2005). « Vagueness in Legal Texts : Is There a Future for Shall ? » In : Bhatia, V., Engberg, J., Gotti, M. et Heller, D. *Vagueness in Normative Texts*. Berne, Suisse : Peter Lang. isbn : 978-3-03910-653-0. url : https://fair.unifg.it/handle/11369/10554.

Wilson, D. et Sperber, D. (1979). « L'interprétation des énoncés ». In : *Communications* 30.1, p. 80–94. doi : 10.3406/comm.1979.1448. url : https://www.persee.fr/doc/comm_0588-8018_1979_num_30_1_1448.

Wooten, C. W. (1987). *Hermogenes' "On types of style"*. UNC Press Books.

Wyner, A. et al. (2009). « On Controlled Natural Languages : Properties and Prospects ». In : *Controlled Natural Language*. International Workshop on Controlled Natural Language. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, p. 281–289. isbn : 978-3-642-14417-2 978-3-642-14418-9. doi : 10.1007/978-3-642-14418-9_17. url : https://link.springer.com/chapter/10.1007/978-3-642-14418-9_17.

Zamanian, M. et Heydari, P. (2012). « Readability of texts : State of the art ». In : *Theory and Practice in Language Studies* 2.1, p. 43.

Zhang, Q. (1998). « Fuzziness - vagueness - generality - ambiguity ». In : *Journal of Pragmatics* 29.1, p. 13–31. issn : 0378-2166. doi : 10.1016/S0378-2166(97)00014-3. url : http://www.sciencedirect.com/science/article/pii/S0378216697000143.

Zowghi, D. et Gervasi, V. (2002). « The Three Cs of requirements : consistency, completeness, and correctness ». In : *International Workshop on Requirements Engineering : Foundations for Software Quality, Essen, Germany : Essener Informatik Beitiage*, p. 155–164.

INDEX DES AUTEURS

AeroSpace and Defence Industries Association of Europe 73
AFIS 46, 91, 172
Akman, V. 11
Albert, C. 51
Albrecht, M. v. 31
Andersson, L. 23
Angelov, K. 30
Angelova, G. 107
Angius, E. 51
Anthony, L. 106
Arora, C. 171
Artino, A. R. 198
Aubin, S. 107
Aughterson, K. 31
Auroux, S. 17, 18
Aussenac-Gilles, N. 177
Austin, J. L. 88

Baisa, V. 107
Barcellini, F. 51
Bartlett, J. M. 32
Barzdins, G. 30
Baumgartner, W. A. 107
Béchet, N. 162
Beauvisage, T. 23
Bendinelli, M. 60
Benezech, D. 16
Bertels, A. 185
Bestgen, Y. 110
Bhatia, V. K. 24
Biber, D. 23, 24, 58, 59, 65, 71, 163, 173
Blancafort, H. 177
Blanchet, P. 17
Bostian, L. R. 92, 99
Botel, M. 83
Bouquet, S. 97
Bowker, L. 105
Braines, D. 39
Bres, J. 87
Briand, L. 171
Bulot, T. 17
Bušta, J. 107

Cabré, M. T. 176
Callan, J. 92
Calvet, L. J. 17
Candito, M. 108
Cannesson, E. 42
Carlier, A. 94
Carlson, N. 51
Cellier, P. 162, 166
Charnock, R. 29, 101
Charnois, T. 50, 162, 166
Charrow, R. 92
Charrow, V. 92, 99
Cheminot, E. 30
Chervak, S. 43
Chevrot, J.-P. 18
Choi, K.-S. 40
Chomsky, N. 18
Clark, P. 40, 73
Cohen, K. B. 107
Coleman, E. B. 83, 92, 99
Collins-Thompson, K. 92
Condamines, A. 7, 20, 42, 52, 71, 83, 98, 99, 161, 181, 187
Coney, M. B. 43
Cornu, G. 88
Coseriu, E. 18
Costermans, J. 92, 210
Coutherut, M. 47
Crabbé, B. 108
Creissels, D. 95
Crémilleux, B. 162
Cutts, M. 35

Da Silva, J. 30
Daille, B. 107, 177
Dal, G. 107
Damljanovic, D. 30

Davies, J. C. 31
Davis, B. 30
De Jong, M. D. 198
De Winter, J. C. 198
Dendale, P. 89
Depecker, L. 21
Desclés, J.-P. 93
Dick, F. 23
Dick, J. 47
Dick, M. 103
Diwersy, S. 106
Djioua, B. 93
Dodou, D. 198
Douay, C. 88
Drouin, P. 105, 176
Drury, C. G. 43
DuBay, W. H. 36
Duranti, A. 8

Elman, J. L. 23
Emeneau, M. B. 16
Eskenazi, M. 92
Eveleens, J. 47

Fabbrini, F. 51
Fabre, C. 184, 185
Fakotakis, N. 177
Farfeleder, S. 51, 171
Fehr, E. 16
Felker, D. B. 99
Femmer, H. 51
Fernández, D. M. 51
Ferré, S. 162
Fischbacher, U. 16
Fish, S. E. 10
Flesch, R. F. 33, 37, 81
Fleury, S. 106
Fludernik, M. 153
Forsbom, E. 177
François, T. 85
Freibott, G. 60
Frérot, C. 25, 105
Fuchs, C. 11–13
Fuchs, N. 30, 38
Fuentes, J. M. 51
Fusani, M. 51

Gadet, F. 17, 18
Gandrillon, O. 162
Ganier, F. 47, 187
García-Flores, J. 93
Gaudin, F. 21, 22, 175
Gautier, L. 175
Gervasi, V. 50
Gibbons, J. 101
Gjesdal, A. M. 98
Gledhill, C. 23, 60, 61, 94, 105
Glinz, M. 50
Gnesi, S. 51
Génova, G. 51
Goodwin, C. 8
Gornostay, T. 177
Gougenheim, G. 33
Gough, P. B. 92
Graf, L. R. 92
Granowsky, A. 83
Grosse, C. 51
Guță, A. 93
Guilbert, T. 26
Gumperz, J. 16
Guéry, F. 32
Gylling, M. 85

Hailu, N. D. 107
Halliday, M. A. K. 14, 15, 35, 109
Hamon, T. 107, 185
Hanbury, A. 23
Harber, J. R. 83
Hargis, G. 132
Harris, W. V. 11
Harrison, P. 40, 73
Harwood, A. 172
Hasan, R. 109
Hathout, N. 107, 184
Hatier, S. 176
Heid, U. 60, 61, 103, 105, 177
Heiden, S. 108
Heilman, M. 92
Heydari, P. 37
Hirst, G. 10
Hjelmslev, L. 18
Hoefler, S. 30
Holat, P. 162

Holland, M. 99
Holmback, H. K. 43
Houdek, F. 150
Huijsen, W.-O. 75
Hull, E. 47
Humbley, J. 21, 175
Hupet, M. 92, 210
Hurtado, O. 51
Hymes, D. 29

IEEE 45
International Council on Systems Engineering (INCOSE) 73
Isahara, H. 40
ISO/IEC/IEEE29148 1, 45, 46, 48, 82, 90
ISO/TS24620-1 40
Ivanova, E. 93

Jackson, K. 47
Jacques, M.-P. 177
Jahchan, N. 42
Jakobson, R. 13–15
Jakubíček, M. 107
Jauch, R. 103
Johannessen, V. 171
Jones, K. 30
Joppich, R. 172
Juergens, E. 51

Kail, M. 93
Kaljurand, K. 30
Kamsties, E. 49, 52
Kandel, L. 85
Kang, J. 103, 185
Karwowski, W. 20
Keim, J. 51, 159
Kennedy, G. A. 31
Kerbrat-Orecchioni, C. 9, 15
Kilgarriff, A. 107, 110
Kitagawa, C. 97
Kittredge, R. 27, 75
Klare, G. 36
Kleiber, G. 9
Kléma, J. 162
Klose, M. 51
Knecht, H. H. 32
Koehn, P. 85
Kokkinakis, G. 177
Korner, S. J. 51, 159
Kovář, V. 107
Kraif, O. 106
Krall, A. 51, 171
Krisch, J. 51, 103, 150, 159
Kübler, N. 25, 60, 61, 105
Kuhn, T. 30, 38–40, 43, 75

Labov, W. 17
Lafon, P. 173
Lami, G. 51
Lamsweerde, A. v. 50
Landhausser, M. 51, 159
Landragin, F. 106
Laplante, P. 51
Le Pesant, D. 194
Le Querler, N. 87, 88
Lee, D. 24, 25
Legallois, D. 162, 166
Lehrberger, J. 27
Lehrer, A. 97
Lenci, A. 184
Lentz, L. R. 198
Lerat, P. 28
Llorens, J. 51
Lopez, S. 15, 20
Lupu, M. 23
Luts, M. 30
Lux, V. 39, 100, 117

Magué, J.-P. 108
Malrieu, D. 23, 97, 98
Manguin, J.-L. 162
Marlaud, S. 32
Marsden, P. V. 190
Mathieu, Y. 92
Mavin, A. 172
McEnery, T. 107
Méchoulam, C. 177
Mel'čuk, I. 175
Merdy, E. 103, 185
Meyer, B. 50
Michelfeit, J. 107
Moles, A. 85
Moreau, M.-L. 17
Moreno, V. 51

Moser, T. 51, 171
Murray, W. R. 40, 73

Nair, D. K. 50
Namer, F. 107
Narcy-Combes, J.-P. 7
Naumchev, A. 50
Nef, F. 87
Nikolova, I. 107
Noailly, M. 93
Novak, M. 172

O'Brien, S. 30, 37, 40, 43, 83, 92
Ogden, C. K. 33
O'Hayre, J. 34
O'Leary, N. 39
Omoronyia, I. 51
O'Regan, G. 50
Orwell, G. 34
Ouellette, J. P. 43
Ownby, R. L. 92

Pakonen, A. 45, 51
Palotti, J. R. 23
Paltridge, B. 24
Panis, C. 171
Peach, B. 49, 52
Pearson, J. 105
Pecman, M. 28, 60, 176
Perreault, W. D. 190
Pickering, F. 99
Picton, A. 98, 99, 105, 181
Pincemin, B. 10, 108
Piroi, F. 23
Plantevit, M. 162
Poibeau, T. 106
Pool, J. 30
Poudat, C. 23, 26, 98
Preece, A. 39
Périnet, A. 185

Quiniou, S. 162, 166

Rabatel, A. 97
Rajan, A. 172
Rajivelu, S. S. 51
Randall, J. H. 92
Rasmussen, K. W. 93
Rastier, F. 9, 10, 23, 175
Rauber, A. 23
Rebeyrolle, J. 153
Redish, J. 37, 99
Renouf, A. 173
Rey, A. 17
Richard, I. 91
Richaudeau, F. 83
Rigotti, C. 162
Rinck, F. 23, 26
Rogers, M. 177
Roland, D. 23
Rosner, M. 30
Rouski, M. 93, 94
Rousselot, F. 60, 105
Rupp, C. 172
Ryan, R. 30
Rychlý, P. 107

Sabetzadeh, M. 171
Sadoun, D. 30
Sahraoui, H.-T. 162
Saint-Dizier, P. 51
Sajous, F. 184
Salem, A. 173
Salvage, P. 101
Sateli, B. 51
Saussure, F. 18
Schmid, H. 108
Schwertel, U. 39
Schwitter, R. 30, 39
Searle, J. R. 88
Selzer, J. 37
Sharoff, S. 177
Shubert, S. K. 43
Sinclair, J. 173
Siouffi, G. 16, 17
Somé, S. S. 50
Somers, H. 27
Sowa, J. 30
Speelman, D. 185
Sperber, D. 11
Spyridakis, J. H. 43
Stage, L. 86–88
Stamatatos, E. 177
Steuckardt, A. 16, 17

Stewart, K. M. 43
Stålhane, T. 51, 171
Suchomel, V. 107
Suchowolec, K. 41
Sullivan, G. M. 198
Swales, J. 24, 29

Tanguy, L. 103, 107, 153, 184, 185
Ştefănescu, D. 60, 105
Temnikova, I. 107
Thompson, J. 40, 73
Tichy, W. F. 51, 159
Todiraşcu, A. 60, 105
Tognini-Bonelli, E. 58, 65
Tommila, T. 45, 51
Trentanni, G. 51
Tufiş, D. 60, 105
Tulechki, N. 103
Tutin, A. 176, 184

Urieli, A. 107

Verhoef, C. 47
Vetters, C. 89
Victorri, B. 11, 106
Vinet, M.-T. 99
Volanschi, A. 61

Wahl, T. 172
Wandel, A. 32
Warnier, M. 20, 52, 71, 161, 187
Webberley, W. 39
Weller, M. 60, 105
Wilkinson, P. 172
Williams, C. 91
Wilson, D. 11
Witte, R. 51
Wooten, C. W. 31
Wright, J. D. 190
Wyner, A. 30

Zamanian, M. 37
Zhang, Q. 43
Zimina, M. 106
Zimmer, F. 171
Zimmer, I. 51
Zimmer, J. 51
Zojer, H. 51, 171
Zowghi, D. 50

INDEX THÉMATIQUE

ambiguïté, 10
analyse distributionnelle, 184

boilerplate, 51, 170

communauté de discours, 29
conjonction, 78, 119, 193, 203
contexte, 8
corpus, 58
corpus-based, 59, 71
corpus-driven, 59, 161

exigence, 45, 86

fonctions du langage, 13

genre textuel, 23

langue contrôlée, 30, 73
langue naturelle, 8
langue spécialisée, 28
lisibilité, 35
longueur de phrase, 82, 129

modalisation, 86, 141, 192, 199
motif, 162

nominalisation déverbale, 97, 156
normaison, 21
normalisation, 21
norme, 16

on (pronom), 97, 151, 195, 208

phraséologie, 60
plain language, 34
pronom, 76, 109, 194, 205

représentativité (corpus), 65

sous-langage, 27
spécification, 46

temps verbal, 86, 134
terme, 175

test (statistique), 110, 197

voix passive, 92, 145, 195, 208