HAL open science

Une étude linguistique de l'emploi des phraséologismes pragmatiques à fonction expressive de la conversation quotidienne chez des locuteurs japonophones du français de niveau avancé

Alexis Ladreyt

To cite this version:

Alexis Ladreyt. Une étude linguistique de l'emploi des phraséologismes pragmatiques à fonction expressive de la conversation quotidienne chez des locuteurs japonophones du français de niveau avancé. Linguistique. Université Grenoble Alpes [2020-..], 2022. Français. NNT : 2022GRALL017 . tel-03925334

HAL Id: tel-03925334

https://theses.hal.science/tel-03925334

Submitted on 5 Jan 2023

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.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : LLSH- Langues, Littératures et Sciences Humaines
Spécialité : Sciences du langage Spécialité Didactique et Linguistique
Unité de recherche : Laboratoire de Linguistique et Didactique des Langues Etrangères et Maternelles

Une étude linguistique de l'emploi des phraséologismes pragmatiques à fonction expressive de la conversation quotidienne chez des locuteurs japonophones du français de niveau avancé

A linguistic study of pragmatic phraseological units with expressive function used in daily conversation : the case of japanese advanced speakers of french

Présentée par :

Alexis LADREYT

Direction de thèse :

Agnès TUTIN PROFESSEURE DES UNIVERSITES, Université Grenoble Alpes	Directrice de thèse
Tomoko HIGASHI Université Grenoble Alpes	Co-directeur de thèse

Rapporteurs :

Gaétane DOSTIE
PROFESSEUR, Université de Sherbrooke
Hisae AKIHIRO
MAITRE DE CONFERENCE, Tokyo University of Foreign Studies

Thèse soutenue publiquement le **20 septembre 2022**, devant le jury composé de :

Agnès TUTIN PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes	Directrice de thèse
Gaétane DOSTIE PROFESSEUR, Université de Sherbrooke	Rapporteure
Hisae AKIHIRO MAITRE DE CONFERENCE, Tokyo University of Foreign Studies	Rapporteure
Cristelle CAVALLA PROFESSEUR DES UNIVERSITES, Université Paris 3Sorbonne Nouvelle	Présidente
Chevrot JEAN-PIERRE PROFESSEUR DES UNIVERSITES EMERITE, Université Grenoble Alpes	Examinateur

Invités :

Tomoko Higashi
MAITRE DE CONFERENCE, Université Grenoble Alpes

Remerciements

Me voici à la fin de ce périple qui a commencé il y a un peu moins de 4 ans. Périple difficile, exigeant, périple éprouvant, mais périple duquel je ressors indéniablement grandi. Bien que ce travail d'écriture fût très solitaire, le travail de thèse lui ne l'était pas. Il est le fruit de rencontres, parfois inopinées, de discussions stimulantes entre collègues, de réflexions et de prises de conscience nourries par le contact à autrui. Il est aussi le résultat d'un soutien inconditionnel, qu'il soit moral ou scientifique. Ce sont ces personnes qui ont directement ou indirectement contribué à l'achèvement de ce travail de thèse que je souhaite remercier ici.

En tout premier lieu, je remercie très chaleureusement mes deux directrices de recherche, Agnès Tutin et Tomoko Higashi. Elles m'ont accompagné et encadré dans mon parcours scientifique d'abord en master, puis en thèse. Ce chemin parcouru ensemble a profondément contribué à façonner le chercheur que je suis aujourd'hui. Dès le début du master, elles n'ont eu de cesse de m'accompagner et de soutenir mes choix scientifiques avec un contact humain et une disponibilité extraordinaire. Durant la thèse, elles m'ont encadré, guidé, parfois recadré avec un sens du détail et une rigueur scientifique qui m'ont profondément marqué et que je souhaite transmettre à mon tour. Je ne mentirais pas en disant que je les considère comme deux mentors, deux références sur lesquelles s'appuie grandement mon parcours scientifique. Le sujet de recherche abordé dans cette thèse n'était pas facile, mais elles ont été très présentes pour me pousser à me dépasser, me laissant la liberté de creuser, d'essayer, de tenter, d'échouer, en somme, de faire ma recherche. Je me rends compte grâce à elles qu'il n'y a pas meilleure école que la prise de risque, l'innovation, la recherche de nouvelles approches, l'envie d'aller là où personne n'a encore été. Elles me l'ont montré à plusieurs reprises, elles-mêmes s'aventurant parfois dans domaines qui ne sont pas les leurs pour me guider au mieux. La route est encore longue, mais grâce à ces années passées à me former à leur côté, je suis maintenant sûr que cette direction est la bonne.

Je tiens également à remercier Hisae Akihiro, Gaétane Dostie, Cristelle Cavalla et Jean-Pierre Chevrot qui me font le grand honneur de faire partie du jury de thèse et qui ont accepté d'évaluer mon travail malgré les nombreuses contraintes temporelles et géographiques. Vous avez toute ma gratitude.

Durant toutes ces années où je me suis consacré pleinement à ma formation scientifique, j'ai pu compter sur le soutien sans faille et le réconfort de ma famille. Tout d'abord, celui de mon frère jumeau, Florian, « mon pote » qui durant toutes ces années a toujours cru en moi et a travaillé pour deux afin de m'offrir la stabilité financière et le contexte idéal à ma formation en master et au début de ma thèse. Tu as ma gratitude éternelle. Ensuite celui de mes parents, Éric et Isabelle qui m'ont guidé et épaulé, m'ont remonté quand j'étais au plus bas, ont cru également en mon potentiel. Je pense également à mon petit frère Arthur, qui a toujours été d'un grand soutien et qui ne dit jamais non à une bonne discussion « entre frangins ». Je remercie également mes grands-parents : Fabienne et Jacky, Nicole et Manuel. Ils ont été formidables, m'ont épaulé, m'ont aidé à me ressourcer et à me retrouver. L'âge faisant expérience, ils ont été de très bons conseils dans les moments de doute. Il me vient une pensée émue envers ma mère biologique, Marianne, qui nous a quittés il y a longtemps, trop longtemps. Je lui dédie ce travail. Je remercie également tous les autres membres de ma famille, ils se reconnaitront.

Que serait ce travail de thèse sans l'entourage amical et professionnel qui ont contribué à stimuler ma curiosité et à renforcer ma motivation ? Ma première pensée va vers Jonathan, mon plus vieil ami qui au bout de 25 ans à se côtoyer ne s'est toujours pas lassé de ma présence. Je remercie également mes amis de longue date dont leur joie et leur bonne humeur ont grandement soutenu ma motivation et m'ont aidé dans les moments les plus difficiles : Robin, Fabien, Alexandre, Jérémy, Julien, Frédéric, Marc, Éric et Yoan. Je remercie aussi Kôsuké, avec lequel j'ai eu la joie de nouer des liens scientifiques et amicaux au cours des dernières années. Les discussions du dimanche lors de nos réunions de travail m'ont permis d'endurer la solitude de l'écriture.

Depuis le début de ce doctorat, j'ai eu le grand honneur et le bonheur de côtoyer et de représenter la communauté doctorante LIDILAIME. Mes chèr.e.s collègues doctorant.e.s ont été une réelle source d'inspiration et de motivation dans ce périple. Leur compagnie, les discussions très stimulantes et les bons moments passés ensemble ont véritablement contribué au succès de cette thèse et à sa qualité. Merci Roxanne, Iris, Laurence, Luca, Émile, Claire W., Claire H., Lan, Yujing, Wenjie, William, Hélène, Grace, Mahault, Chaeyoung, Kim, Marie, Manon, Aurélie, Rim, Azadeh, Lizandro, Nicolas, Doriane, Sylvain ainsi que tous les autres doctorants de la communauté.

Les collègues que j'ai côtoyés au sein du LIDILEM ont été également une grande source d'inspiration et de motivation dans ce travail. Merci à Béatrice Villa, Alice Henderson, Françoise Boch, Julie Sorba, Cyril Trimaille, Marinette Matthey, Laura Abou-Haidar, Thierry Soubrié, Jérôme Barona, Stéphanie Galligani, Catherine Muller, Olivier Kraif, Catherine Frier, Elke Niessen, Claudine Moïse, Laure Gardelle, Samia Ounoughi, Lucia Gomez, Tatiana Aleksandrova, Lidia Miladi, Francis Grossmann, Isabelle Rousset, Jean-Marc Colletta, Iva Novakova, Jean-Pascal Simon et à tout le laboratoire. J'adresse également un remerciement tout particulier à Sanae Harada et Simon Tuchais de l'université de Sophia (Tokyo) et à Mathieu Mangeot. Leurs qualités scientifique et humaine ont indéniablement contribué à cette thèse. Enfin, j'adresse une pensée à tous les collègues des autres universités que j'ai pu côtoyer, ils se reconnaitront.

Mes derniers remerciements vont à ma chère épouse, Tomoko. Elle a été l'épaule bienveillante et aimante qui m'a donné la stabilité et le soutien dans les moments difficiles. Elle a pendant ces dernières années fait preuve d'une patience extraordinaire, supportant un époux tête en l'air et peut-être un peu trop absorbé par son travail. Elle a très tôt vu le potentiel en moi, et m'a insufflé cette passion pour la langue japonaise. Sa rencontre a été en quelque sorte l'heureux déclencheur de cette envie d'explorer la langue et ses mécanismes. Je lui dois tout et plus encore…

Une étude linguistique de l'emploi des phraséologismes pragmatiques à fonction expressive de la conversation quotidienne chez des locuteurs japonophones du français de niveau avancé

MOTS-CLÉS : phraséologie, acquisition L2, compétence lexicopragmatique, pragmatique

RÉSUMÉ

Notre étude s'inscrit à l'interface de la phraséologie et de l'acquisition de la langue seconde. Elle vise à étudier la compétence lexicopragmatique de locuteurs japonophones du français de niveau avancé mobilisée dans l'emploi de phraséologismes pragmatique à fonction expressive de la conversation quotidienne. Nous nous focaliserons notamment sur les difficultés rencontrées et sur les stratégies mobilisées en compréhension et en production. Nous tenterons également de caractériser les spécificités linguistiques de la variété non-native impliquée dans l'emploi des phraséologismes qui nous intéressent. Afin d'observer ces éléments, nous mobilisons un protocole expérimental en trois phases. La première phase constitue un diagnostic des spécificités de la compétence pragmatique mobilisée dans la production des PhPex chez des locuteurs japonophones du français de niveau intermédiaire. Les premières observations issues de ce diagnostic serviront à la deuxième phase de notre expérimentation qui constitue un test linguistique mesurant la compétence d'usage du lexique pragmatique chez les locuteurs japonophones du français de niveau avancé. La troisième phase constitue une série d'entretien de niveau métapragmatique menée avec les locuteurs ayant passé le test de la phase 2. Nous questionnerons notamment les stratégies d'emploi et de compréhension employées par les locuteurs non natifs lors de la passation du test. Nous proposerons ensuite une modélisation des spécificités d'emploi des phraséologismes pragmatique à fonction expressive chez les locuteurs observés. Cette modélisation sera fondée sur une analyse outillée et qualitative des résultats du protocole.

A linguistic study of the use of pragmatic phraseologisms with an expressive function in everyday conversation by advanced Japanese speakers of French

KEYWORDS : phraseology, L2 acquisition, lexicopragmatic competence, pragmatics

ABSTRACT

This study is at the intersection of phraseology and second language acquisition. It aims to study the lexicopragmatic competence of advanced Japanese speakers of French in the use of everyday conversation's pragmatic phraseologisms with an expressive function. In particular, we will focus on the difficulties encountered and the strategies used in comprehension and production. We will also try to characterise the linguistic specificities of the non-native variety involved in the use of the phraseologisms we are interested in. In order to observe these elements, we mobilise an experimental protocol in three phases. The first phase involves a diagnosis of the specificities of the pragmatic competence mobilised in the production of PhPex in intermediate-level Japanese speakers of French. The first observations resulting from this diagnosis will be used in the second phase of the experiment, which is a linguistic test measuring the competence of use of the pragmatic lexicon of advanced-level Japanese speakers of French. The third phase is a series of metapragmatic interviews conducted with speakers who have passed the test in phase 2. In particular, we will question the use and comprehension strategies employed by non-native speakers during the test. We will then present a model of the specificities of the use of pragmatic phraseologisms with an expressive function by the speakers observed. This modelling will be grounded on a tool-based and qualitative analysis of the results of the protocol.

言語学の観点からフランス語の日本人上級学習者の日常会話における感情を表す語用論的定型表現の研究

キーワード: 定型表現分析、第二言語習得、語用論的語彙能力、語用論

要旨

本研究は、定型表現分析 (phraséologie)と第二言語習得の分野をめぐるものである。本研究は、フランス語の日本人上級学習者の日常会話における感情を表す語用論的定型表現に関する語用論的語彙能力の研究を目的とする。特に、フランス語の日本人上級学習者の定型表現に関する理解や発言の困難さ、ストラテジーについて分析する。また、フランス語の日本人上級学習者の語用論的定型表現の使用において、どのような言語的特徴があるのかを明らかにする。 これらの特徴を観察するために、本研究では三段階の調査を行った。第一段階の調査では、中級のフランス語の日本人学習者の PhPex の使用と解釈に関する語用論能力の特徴の考察を行った。ここから得られた予備結果を踏まえ、第二段階の調査、すなわちフランス語の日本人上級学習者の語用論的語彙能力を測定する言語テストを実施した。第三段階の調査では、第二段階のテストを受けたフランス語の日本人上級学習者に対して様々なメタ語用論的観点からの面接を行った。特に、この三段階目の調査で、日本人上級学習者が第二段階のテスト中に用いたフランス語の運用とその理解に関するストラテジーについて検討する。その考察を元に、日本人上級学習者による感情を表す語用論的定型表現の使用に関する特徴のモデル化を提示する。このモデル化は、調査結果のコーパス言語的、統計的および定性的分析に基づいているものである。

Sommaire

REMERCIEMENTS 2

SOMMAIRE 8

NORME DE PRESENTATION DES CITATIONS EN JAPONAIS 12

INTRODUCTION 14

PARTIE 1 - CADRAGE THÉORIQUE 20

PREMIER VOLET DU CADRE THEORIQUE - DE LA PHRASEOLOGIE GENERALE AUX PHRASEOLOGISMES PRAGMATIQUES A FONCTION EXPRESSIVE 22

CHAPITRE 1 - ÉTAT DES LIEUX DU DOMAINE DE LA PHRASEOLOGIE 26
1. *Le champ de la phraséologie* *26*
2. *Le domaine de la phraséologie au Japon* *32*
3. *Extension récente du domaine de la phraséologie* *39*

CHAPITRE 2 - LA PHRASEOLOGIE PRAGMATIQUE : VERS UN NOUVEAU PARADIGME 42
1. *Le champ de la pragmatique* *42*
2. *La phraséologie pragmatique au sens large* *44*
3. *La phraséologie pragmatique des interactions orales* *45*
4. *Revue des différents paradigmes de phraséologismes pragmatiques apparentés aux PhPex* *46*

CHAPITRE 3 - LES PHRASEOLOGISMES PRAGMATIQUES A FONCTION EXPRESSIVE : FORME, SENS, USAGE 70
1. *Choix de dénomination* *70*
2. *Définitions des spécificités des PhPex* *74*
3. *Les PhPex observés dans cette étude* *82*

DEUXIEME VOLET DU CADRE THEORIQUE - DES PHRASEOLOGISMES PRAGMATIQUES A FONCTION EXPRESSIVE A L'ACQUISITION DE LA COMPETENCE LEXICOPRAGMATIQUE 85

CHAPITRE 4 - CONTEXTUALISATION DU CHAMP DE LA PRAGMATIQUE DE LA LANGUE SECONDE 87
1. *Tour d'horizon historique et théorique* *87*
2. *Le champ de la pragmatique de la langue seconde* *92*

CHAPITRE 5 - UNE APPROCHE INTERCULTURELLE, INTERLINGUISTIQUE ET CONTRASTIVE 95
1. *La pragmatique transculturelle (Cross-Cultural pragmatics)* *95*
2. *La pragmatique interlangagière (Interlanguage Pragmatics)* *97*
3. *Analyse Contrastive de l'interlangue (Contrastive Interlanguage Analysis)* *98*
4. *À propos des travaux sur l'acquisition des phraséologismes pragmatique de la conversation* *100*

CHAPITRE 6 - DE LA COMPETENCE COMMUNICATIVE A LA COMPETENCE LEXICOPRAGMATIQUE 103
1. *Évolution de la définition de la compétence pragmatique* *103*
2. *Vers une nouvelle approche de la compétence pragmatique ?* *112*
3. *La compétence lexicopragmatique : définition et modélisation* *117*

CHAPITRE 7 - METHODE D'ELICITATION ET D'EVALUATION DE LA PRODUCTION PRAGMATIQUE EN L2 127
1. *Les protocoles non-interactifs* *127*
2. *Les protocoles interactifs* *136*

SYNTHESE DU CADRE THEORIQUE 145

PARTIE 2 - PROTOCOLE EXPÉRIMENTAL ET ANALYSE 149

AVANT-PROPOS 151

PREMIERE PHASE DU PROTOCOLE EXPERIMENTAL : 153

CHAPITRE 8 - CADRAGE EXPERIMENTAL DE LA PHASE DIAGNOSTIQUE 155
1. *Questionnement de départ et description de l'étude* *155*
2. *Le diagnostic de la compétence lexicopragmatique et la notion d'erreur.* *157*
3. *Méthodologie de l'étude diagnostique* *159*

CHAPITRE 9 - ANALYSE DES RESULTATS DU DIAGNOSTIC 185
1. Synthèse des résultats quantitatifs par catégorie d'annotation 185
2. Analyse qualitative des anomalies pragmatiques et interactionnelles 194
3. Diagnostic et conclusion 218

DEUXIEME PHASE DU PROTOCOLE EXPERIMENTAL : 225

POINT ETAPE 1 DU PROTOCOLE EXPERIMENTAL 227
CHAPITRE 10 – CADRAGE EXPERIMENTAL DU TEST LEXICOPRAGMATIQUE 231
1. Les informateurs 231
2. Le test lexicopragmatique 233
3. Traitement des données 241
CHAPITRE 11 - ANALYSE QUANTITATIVE ET STATISTIQUE DES DONNEES 269
1. Étude quantitative des résultats obtenus au test lexicopragmatique 270
2. Étude statistique des variables liées à l'apprentissage et les variables liées à la pratique du français 278
3. Synthèse de l'analyse quantitative et statistique des données 292
4. Discussion des résultats quantitatifs et statistiques 295
5. Remarques finales 300
CHAPITRE 12 - ANALYSE QUALITATIVE ET COMPARATIVE DES DONNEES 305
1. Analyse linguistique de l'activité 1 (complétion) 306
2. Analyse linguistique de l'activité 2 (reformulation) 314
3. Analyse linguistique de l'activité 3 (équivalence fonctionnelle) 327
4. Analyse linguistique de l'activité 4 (QCM lexical) 345
5. Analyse linguistique de l'activité 5 (QCM sémantique) 355
6. Synthèse des résultats du test lexicopragmatique 362

TROISIEME PHASE DU PROTOCOLE EXPERIMENTAL : 375

POINT ETAPE 2 DU PROTOCOLE EXPERIMENTAL 377
CHAPITRE 13 - METHODOLOGIE DE RECUEIL ET D'ANALYSE DES DONNEES DES ENTRETIENS SEMI-DIRECTIFS 379
1. Questionnement de départ et objectif 379
2. Principe et structure de l'entretien d'élicitation métapragmatique semi-directif 381
3. Méthodologie d'analyse 389
CHAPITRE 14 - SYNTHESE DES DISCUSSIONS METAPRAGMATIQUES 393
1. Les difficultés de discernement de sens phraséologique et littéral 393
2. Difficulté liée à des analogies erronées avec la langue maternelle 394
3. Difficultés liées au repérage de la fonction pragmatique 396
4. Difficulté d'accommodation du registre au contexte de communication 396
5. Difficultés liées aux objectifs de communication 398
6. Difficulté en production liée à l'asymétrie entre compétence de compréhension et diversité du répertoire lexicopragmatique 399
7. L'influence de l'éthos communicatif sur la production en LC 400

CONCLUSION 403

1. Synthèse des résultats 403
2. Limites et améliorations possibles de notre approche 407
3. Mot de la fin et ouverture 408

BIBLIOGRAPHIE 411

SIGLES ET ABREVIATIONS UTILISES 442

TABLE DES ILLUSTRATIONS 444

LISTES DES TABLEAUX 448

ANNEXES 450

TABLE DES ANNEXES 451

TABLE DES MATIERES 482

Norme de présentation des citations en japonais

Pour les besoins de cette thèse, les termes et énoncés japonais seront transcrits en se basant sur le système *Hepburn modifié* (ANSI Z39.11-1972). Il existe d'autres normes de transcriptions, notamment le *kunreishiki* 訓令式 (ISO 3602), mais le système Hepburn semble à notre sens plus intuitif, car proche des représentations phonétiques françaises, ce qui facilitera la lecture pour le public non averti. Par ailleurs, il est utile de préciser que le japonais ne comporte pas d'espaces à l'origine, nous segmenterons donc les énoncés avec une barre oblique « / » si besoin.

En ce qui concerne les voyelles, les graphies « a » et « o » seront toujours des voyelles fermées. La graphie « e » sera systématiquement prononcée à la manière d'un [e] fermé. Le graphème « u » renverra à un [u] neutre très fermé et non arrondi. Pour ce qui est de l'allongement vocalique, il sera noté comme suit :

- [u] allongé sera noté « û »
- [e] allongé sera noté « ei »
- [o] allongé sera noté « ô »

Pour ce qui est des consonnes, les seules différences à noter concernent les graphies qui vont suivre. La graphie « ch » renverra à l'affriquée [t͡ʃ]. La graphie « h » sera toujours aspirée. Les graphies « y » et « w » renverront aux semi-voyelles [j] et [w]. Enfin, la graphie du « r » sera prononcée comme une consonne liquide battue dont la prononciation se situe entre le [l] et le [r]. En outre, le redoublement d'une consonne renverra à l'accentuation de la longueur et de la plosivité des consonnes à la manière des géminées en italien. Si nous prenons le toponyme « ro**pp**ongi », le [p] redoublé sera prononcé de manière très plosive avec un léger allongement du temps de plosion. Le « n » syllabique « ん » sera transcrit « n », devant les labiales « p », « b » ou « m » (ex. とんぼ *to**n**bo*). Afin d'écarter toute ambiguïté, le « n » syllabique sera retranscrit par un « n' » lorsque celui-ci précèdera une voyelle (Ex. たんい *ta**n'i*** (unité) *vs* たに *ta**ni*** (vallée)). Enfin, les particules は、を、へ seront transcrites *wa*, *o* et *e*.

Introduction

Notre pratique de la conversation quotidienne est jalonnée par l'emploi très fréquent de séquences lexicopragmatiques stéréotypées qui réalisent des fonctions communicatives complexes indispensables au bon déroulement de l'acte de communication. En voici un exemple :

[PAT] hop
[JUD] non mais laisse euh je vais
[PAT] oh je fais une-
[JUD] bah je vais ranger
[JUD] ne t'en fais pas
[JUD] tu veux plus rien non pour la route

Corpus CLAPI (Baldauf-Quilliatre *et al.*, 2016), sous-corpus « Apéritif_glasgow »

Dans cet extrait d'interaction orale, nous observons l'occurrence de la séquence *ne t'en fais pas* qui est employée ici comme expression d'atténuation permettant de rassurer l'interlocuteur. Indirectement, le locuteur employant cette expression demande au locuteur de ne pas se soucier du rangement, car il souhaite s'en occuper. Cet exemple montre que le locuteur va mobiliser lors de la conversation tout un inventaire de routines conversationnelles (Coulmas, 1981) plus ou moins préfabriquées sur les plans sémantiques, syntaxiques et pragmatiques, afin d'exprimer une grande diversité de fonctions illocutoires, interpersonnelles ou afin de structurer son propos. Nous désignerons dans cette thèse ces structures par *phraséologismes pragmatiques*. Au sein de cette classe des phraséologismes pragmatiques, nombreuses sont les structures qui ont pour fonction principale l'expression d'un état émotionnel et psychologique conditionné par l'évaluation subjective que projette le locuteur sur les différents événements qui constituent son expérience de sujet parlant. Nous désignerons ces expressions par *phraséologismes pragmatiques à fonction expressive* (désormais *PhPex).* Afin d'illustrer cette sous-classe, nous pouvons citer par exemple des expressions comme *ça va pas la tête !* ou *tu m'en diras tant !* qui employées en contextes spécifiques actualisent des fonctions pragmatiques qui ne peuvent être prédites à l'aide des constituants lexicaux. Dans le premier cas, le PhPex permet au locuteur d'exprimer son indignation face à un événement déclencheur qui l'affecte négativement ou qu'il ne cautionne pas. Dans le second cas, l'expression permet au locuteur d'exprimer sa désillusion face à un événement ayant de faibles chances de se réaliser ou d'exprimer un refus.

La maitrise du lexique préfabriqué de l'oral quotidien apparaît bien souvent comme l'un des indicateurs (sans être le seul) d'une bonne maitrise de la langue cible (ou *LC*) chez l'apprenant en langue étrangère (Gonzalez-Rey 2015). Il n'apparait donc pas étonnant que les questionnements récents en didactique des langues s'intéressent de plus en plus au lexique préfabriqué chez l'apprenant non natif du français. On voit essaimer de nombreuses initiatives, ainsi que de nouveaux cadres théoriques. Ceux-ci ont à cœur d'explorer les mécanismes de compréhension et de production de ces phraséologismes pragmatiques chez les natifs et les non-natifs, et de produire du matériel pédagogique adapté à l'apprentissage de ces objets linguistiques particuliers. Cavalla (2009) mentionne que « L'enseignement de la phraséologie […] est actuellement au cœur de nos interrogations […] [Les] auteurs du Cadre européen Commun de Référence pour les Langues, le CECRL (Conseil de l'Europe, 2001, p. 87), recommandent désormais l'enseignement de la phraséologie quasiment au même titre que le reste du lexique » (Cavalla 2009 : 1). En écho à ces initiatives, les auteurs du CECRL mentionnent également l'importance de la maitrise des préfabriqués linguistiques, en plaçant cette compétence comme condition nécessaire à l'accession au dernier stade du cursus d'apprentissage, à savoir le niveau C2[1]. En outre, l'arrivée très récente du volume complémentaire du CECRL en 2018 a montré qu'il existe un besoin réel en termes de formation et d'évaluation de la compétence liée à l'usage du lexique préfabriqué de l'oral quotidien. Il y a donc là tout un enjeu crucial pour le domaine de la *phaséodidactique* (González-Rey 2019), la nécessité d'un meilleur traitement des phraséologismes pragmatiques dans le matériel pédagogique destiné aux apprenants se faisant de plus en plus grande (Catena et Corral 2017).

Le constat « paradoxal » servant de point de départ à cette étude est en lien avec ces préoccupations didactiques grandissantes à l'égard de la phraséologie des interactions orales, et plus globalement, de la compétence de communication. Il n'est en effet pas rare de constater qu'en dépit d'une maitrise approfondie de la langue cible sanctionnée par l'obtention d'une certification de niveau avancé (B2 à C2 CECRL) des difficultés sont

---

[1] « Je peux participer sans effort à toute conversation ou discussion et je suis aussi très à l'aise avec les expressions idiomatiques et les tournures courantes. Je peux m'exprimer couramment et exprimer avec précision de fines nuances de sens. En cas de difficulté, je peux faire marche arrière pour y remédier avec assez d'habileté pour que cela passe inaperçu » (Extrait du *Tableau 2 - Niveaux communs de compétences – Grille pour l'auto-évaluation*, CECRL).

observables. Cette complexité se manifeste souvent sous la forme d'énoncés acceptables et bien structurés au niveau grammatical, mais qui donnent une sensation d'étrangeté lors de l'emploi. Cette sensation d'étrangeté résulte notamment de plusieurs facteurs tels que le choix lexical, l'ajustement de l'illocutoire ou la variation du registre. Il peut s'agir par exemple de l'emploi du tutoiement dans un contexte de vouvoiement, ou du suremploi des formes de remerciement ou d'excuse lorsque cela n'est pas nécessaire.

Ce paradoxe a déjà été observé dans une série de travaux menés sur l'anglais. Les travaux de Rose & Ng (2001) et de Tateyama (2001) ont démontré qu'une simple exposition à la langue cible de l'apprentissage ne suffit bien souvent pas et qu'une formation explicite menant le locuteur non natif (ou *LNN*) à s'interroger sur les contraintes pragmatiques inhérentes à la langue est nécessaire. En outre, les travaux d'Olshtain & Blum-Kulka (1984) et de Bardovi-Harlig & Hartford (1993) montrent qu'un locuteur non natif avec des compétences très développées dans la grammaire de la langue cible peut présenter de grandes lacunes dans la compétence pragmatique de cette même langue, ce qui a pour résultat l'usage pragmatiquement déficient de formes grammaticalement correctes. La simple maitrise de la grammaire ne semble pas suffire à permettre la pleine réalisation des compétences pragmatiques et interactionnelles. Un travail linguistique de modélisation du fonctionnement pragmatique des unités lexicopragmatiques et des difficultés rencontrées par les locuteurs non natifs semble indispensable. Partant de ce constat paradoxal, nous avons entrepris de mieux comprendre le fonctionnement de l'acquisition de la compétence pragmatique et les difficultés liées à la production et la compréhension des PhPex.

Notre étude s'inscrit à l'interface de la phraséologie et de l'acquisition de la langue seconde. Il s'agit d'un thème encore peu exploré, tout du moins en ce qui concerne la classe des phraséologismes pragmatiques. Il existe quelques études rigoureuses sur les routines conventionnelles (Bardovi-Harlig 2018) ou le lexique préfabriqué (Forsberg 2010), mais aucune d'entre elles ne s'est encore intéressée à la conversation quotidienne et plus particulièrement au lexique préfabriqué expressif. En outre, les études actuelles sur l'acquisition de la compétence pragmatique ou du lexique préfabriqué à fonction pragmatique se sont intéressées principalement à des locuteurs non natifs évoluant en milieu académique (échange universitaire notamment). Cependant, peu d'entre elles proposent une étude menée sur des locuteurs non natifs immergés dans un milieu professionnel ou civil, hors apprentissage guidé. C'est partant de ces différents manques en ressources sur

l'acquisition de ce type de lexique que nous avons entrepris notre étude. Cette étude se fonde sur trois principes :

1) L'étude de la compétence mobilisée dans l'usage des PhPex fondée sur une analyse linguistique fine des productions de LNN recueillies à l'issue de notre protocole expérimental.

2) L'observation des difficultés d'emploi rencontrées par les LNN dans une perspective lexicale et contrastive.

3) La mise en place d'un protocole expérimental outillé et validé par des procédures d'accord interjuge et interannotateur.

Notre perspective met donc au centre de ses préoccupations l'observation du matériel linguistique mobilisé par les locuteurs non natifs. L'analyse des données se focalisera principalement sur cette entrée. Les locuteurs observés dans le cadre de cette étude sont des locuteurs japonophones du français. Nous observerons deux tranches de niveau : le niveau intermédiaire (A2/B1, phase 1 du protocole) et avancé (B2-C2 phase 2 protocoles). Ce choix s'est fait pour deux raisons : 1) nous voulions observer un cas où la langue maternelle (ou *LM*) du locuteur est typologiquement éloignée de la langue cible où les PhPex sont employés, 2) le couple de langue français/japonais est très peu étudié, que ce soit en phraséologie ou en didactique. Ce travail de thèse vise à répondre aux questions de recherche suivantes :

- Quelles sont la nature, la distribution et la fréquence des emplois de PhPex spécifiques aux locuteurs japonophones et comment diffèrent-elles d'un groupe de locuteur à l'autre ?
- Quelles sont les différentes difficultés rencontrées par les locuteurs japonophones ?
- Observe-t-on des récurrences dans les difficultés observées chez ces locuteurs ?
- Quels sont les différents niveaux linguistiques concernés par ces difficultés ?
- Est-il possible de caractériser linguistiquement ces spécificités d'usage ?

Afin de répondre à ces questions, nous mobilisons un protocole expérimental en trois phases. La première phase constitue un diagnostic des spécificités de la compétence pragmatique mobilisée dans la production des PhPex chez des locuteurs japonophones du français de niveau intermédiaire. Les premières observations issues de ce diagnostic serviront à la

deuxième phase de notre expérimentation qui constitue un test linguistique mesurant la compétence d'usage du lexique pragmatique chez les locuteurs japonophones du français de niveau avancé. La troisième phase constitue une série d'entretiens de niveau métapragmatique menée avec les locuteurs ayant passé le test de la phase 2. Nous questionnerons notamment les stratégies d'emploi et de compréhension employées par les locuteurs non natifs lors de la passation du test. Ce protocole en trois phases a été pensé afin de permettre d'observer les spécificités d'usage et de compréhension des PhPex sous différents angles que nous détaillerons lors de chaque phase.

La présente thèse se divise en deux grandes parties : le cadre théorique et l'analyse des données du protocole expérimental. La première partie dédiée à la revue de littérature se divise en deux volets : 1) un premier volet consacré à un état de l'art sur le champ de la phraséologie pragmatique et la définition des PhPex, 2) un deuxième volet consacré à l'acquisition de la compétence pragmatique en langue seconde et à la compétence mobilisée dans l'usage du lexique préfabriqué à fonction pragmatique. Nous proposerons ensuite une synthèse rapide de ces deux volets qui permettra de faire le lien avec la suite de la thèse.

La deuxième grande partie de cette thèse est dédiée à la discussion et à l'analyse des données issues de notre protocole expérimental. Cette partie se divise en trois étapes dédiées à chacune des phases de notre protocole expérimental. Afin d'introduire ces trois phases, nous proposerons un chapitre introductif qui permettra d'expliquer plus en détail l'architecture du protocole expérimental et les hypothèses qui seront explorées. Entre chaque phase de cette deuxième partie, nous proposons un point étape qui permettra de faire le point sur la phase précédente et de faire le lien avec ce qui suit.

Enfin, nous achèverons cette thèse sur une conclusion en deux parties. La première partie synthétisera brièvement les différentes étapes et les résultats obtenus lors de notre protocole expérimental. La seconde étape conclura notre thèse et proposera une ouverture sur les suites qui seront données à ce travail.

Nous proposons ici un schéma qui permettra au lecteur de comprendre l'articulation des différentes parties entre elles :

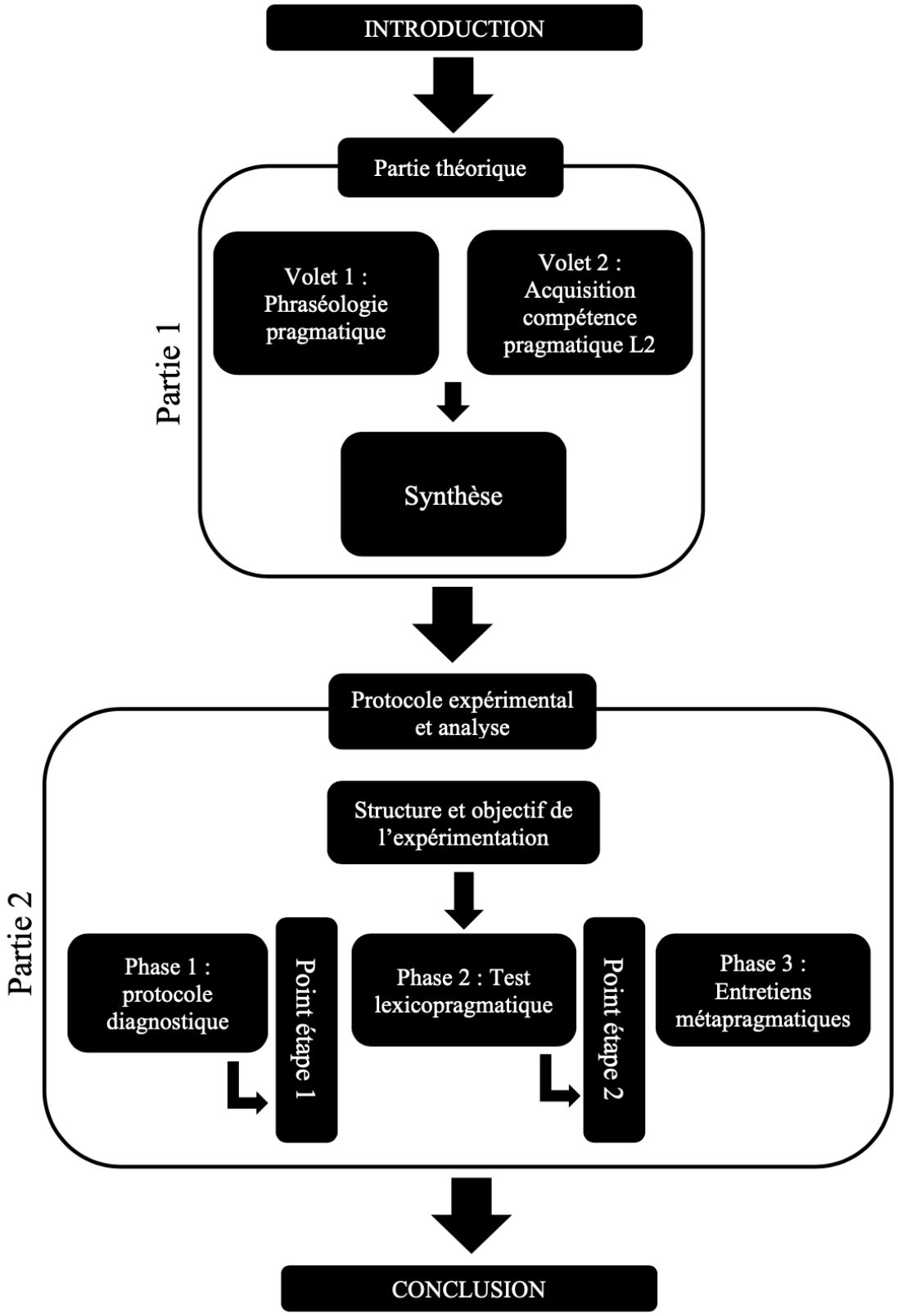


Figure 1 : Structure et articulation de la thèse

PARTIE 1
-
CADRAGE THÉORIQUE

Premier volet du cadre théorique
-
De la phraséologie générale aux phraséologismes pragmatiques à fonction expressive

Ce travail de synthèse se décline en 2 volets. Le premier volet vise à délimiter le cadre théorique de la phraséologie pragmatique et de notre objet d'étude. Nous procèderons en trois étapes : la première étape constitue un rapide historique du champ de la phraséologie en occident et au Japon qui nous permettra de poser le cadre dans lequel s'est développée la *phraséologie pragmatique*. La deuxième étape vise la description du champ de la phraséologie pragmatique et du sous-champ qui nous intéresse plus particulièrement, celui de la *phraséologie pragmatique des interactions orales*. Nous procèderons dans ce cadre à une revue des principaux objets phraséologiques apparentés à cette catégorie, ce qui permettra de donner un aperçu du maquis théorique dans lequel s'est développée notre réflexion. Enfin la troisième étape constituera une proposition de définition de notre objet d'étude, les *phraséologismes pragmatiques à fonction expressive de la conversation quotidienne* (PhPex).

Le deuxième volet de ce travail de synthèse vise à proposer une vue d'ensemble des différentes notions concernant l'acquisition de la compétence pragmatique en L2 qui seront mobilisées dans l'analyse de nos données. Cette deuxième phase se décompose en 3 étapes : une première étape de contextualisation du champ de l'acquisition de la compétence pragmatique et des outils théoriques qui seront utilisés. Une deuxième étape qui exposera les spécificités de notre angle d'analyse : la *compétence lexicopragmatique*. Enfin une troisième étape qui discutera les différentes propositions de dispositifs de mesure de la compétence pragmatique que nous emploierons dans notre protocole expérimental.

Nos travaux soulèvent certains aspects contrastifs entre le français et le japonais. Nous nous efforcerons ainsi d'apporter des éléments théoriques[2] de deux ordres. D'une part, nous évoquerons le champ de la phraséologie japonaise. D'autre part, nous centrerons autant que possible notre état de l'art en acquisition sur le public japonophone qui est la cible de ce travail de recherche. En outre, ce travail de recherche n'est pas spécifiquement destiné à traiter de la phraséologie ou de l'acquisition de manière exclusive, mais à articuler ces deux

[2] Nous attirons toutefois l'attention du lecteur sur le fait qu'en dépit de nombreux efforts mobilisés dans la recherche documentaire sur les ressources en langue japonaise, en tant que chercheur ne vivant pas sur le sol japonais et non-affilié à son domaine universitaire, nous n'avons pas accès à toutes les ressources documentaires disponibles dans l'archipel. Il se peut donc que certaines ressources bibliographiques aient pu échapper à notre état de l'art sur le sujet ou que certaines de nos observations soient erronées. Nous invitons ainsi le lecteur à nous faire part de toute recommandation ou correction à l'adresse suivante : **ladreyt.alexis@gmail.com.**

champs disciplinaires ensemble. Ce faisant, nous ne procèderons pas à un exposé exhaustif d'un de ces domaines. Nous détaillerons uniquement les notions utiles à notre démonstration, de manière à ne pas complexifier inutilement le propos.

Chapitre 1 - État des lieux du domaine de la phraséologie

Ce chapitre a pour objectif de situer le champ de notre étude et le contexte théorique sur lequel repose toute la réflexion sur les phraséologismes de l'oral qui sera développé dans le chapitre 2. Dans un premier temps, nous aborderons le contexte historique du développement de la discipline de la phraséologie. Dans un second temps, nous définirons les contours notionnels de la notion de *phraséologisme*. Nous ferons un rapide état de l'art du domaine de la phraséologie au Japon afin de mieux comprendre les enjeux de notre travail de recherche sur le japonais. Enfin, nous traiterons de l'extension récente du domaine de la phraséologique, ce qui permettra d'introduire le deuxième chapitre.

1. Le champ de la phraséologie

1.1. Survol historique

L'intérêt porté aux constructions préfabriquées de la langue ne date pas d'aujourd'hui. En effet, on remarque qu'il existait déjà un intérêt ancien, et ce bien avant le 17e siècle, notamment chez les premiers éditeurs de recueils parémiologiques concernant les proverbes populaires, maximes et tout autre recueil de vérité morale (Bàrdosi 1990). Par la suite, ces recueils se sont mis à intégrer toute sorte de séquences préfabriquées, parmi lesquelles celles dont le sens faisait appel à des images ou des métaphores stéréotypées. Les premières attestations du terme *phraséologie* en France sont signalées par le TLFi[3] en 1678 dans le cadre d'un ouvrage didactique puis en 1778 dans des ouvrages de littérature et de stylistique (González-Rey 2021 : 20). Avec l'émergence de la lexicographie comme domaine à part entière au 19e et la séparation entre parémiologie et phraséologie, on va assister aux premières tentatives d'inventaire structuré de suites de mots lexicalisées et présentant un sens complexe. Ces suites seront de plus en plus différenciées des parémies (González-Rey 2021 : 19). Les linguistes européens ont commencé à vraiment s'intéresser à la phraséologie dès la fin 19e avec notamment quelques travaux exploratoires parmi lesquels ceux de l'allemand Paul (1880) sur des expressions figées non compositionnelles, les travaux de l'anglais Sweet (1891) sur les *idioms* et côté français, les observations préliminaires de Bréal

[3] http://atilf.atilf.fr/

(1897) sur plusieurs types de lexiques préfabriqués, figés sur le plan syntaxique et sémantique, qu'il désignera par *formule*, *locution* et *groupe articulé*.

Ce sera Bally (1909), disciple de Saussure et considéré comme l'un des pères fondateurs de la phraséologie en Europe, qui le premier va formaliser tout un ensemble de constructions préfabriquées propre à la conversation. Il va ainsi tenter de proposer une typologie approfondie qu'il développera dans son ouvrage précurseur, le *Traité de stylistique française*. Au sein de sa typologie, deux catégories apparaissent particulièrement intéressantes : les *séries phraséologiques* et les *unités phraséologiques* qu'il va caractériser par différents degrés de *cohésion sémantique* (idiomaticité), une régularité syntaxique et des fonctions pragmatiques particulières et sur lesquelles nous reviendrons dans une autre section. Il sera également l'un des premiers à parler de phraséologie exclamative (Legallois et Tutin 2013 : 5). Fait intéressant, Ferdinand de Saussure évoquait quelques années auparavant la notion de « locution toute faite » sans pour autant l'approfondir, notion que l'on retrouvera plus tard dans son œuvre posthume le *Cours de Linguistique Générale* (1916 : 172-173) et qui sera redéveloppée par un autre de ses disciples, Sechehaye (1921 dans González-Rey 2021 :26), sous la forme d'une étude des locutions et des mots-composés dans une perspective sémantique.

Vinogradov (1946, dans Bàrdosi 1990) qui s'inscrit dans la lignée de Bally va populariser le terme de *phraséologisme* et œuvrer pour l'émergence de la phraséologie comme domaine autonome (González-Rey 2021 : 23). Le champ de la phraséologie sera dès lors fortement impulsé par les linguistes russes et allemands. Dès 1970, les linguistes s'emparent de plus en plus de la question, notamment en ce qui concerne la langue écrite. Le 20e siècle connait dès lors un fort engouement pour la phraséologie, notamment chez les fonctionnalistes anglais (Halliday 1966 ; Firth 1968 ; Cowie 1981 ; Sinclair 1991), chez les linguistes américains (Fillmore, Kay et O'Connor, 1988), chez les linguistes travaillant sur le français (Haussmann 1989 ; M. Gross 1993 ; Mel'čuk 1995 ; G. Gross 1996 ; Mejri 1998), chez les linguistes allemands (Hausermann 1977 ; Gréciano 1983 ; Gläser 1986 ; Burger 1998) et chez bien d'autres encore. Cette forte expansion du domaine a aussi vu l'apparition de deux traditions correspondant à deux traitements différents (mais complémentaires) des

phraséologismes : d'un côté, l'approche anglaise[4] qui s'est développée davantage dans un ancrage fort à la linguistique de corpus, notamment au travers des mesures de fréquence, d'association et de spécificité. De l'autre, la tradition franco-germano-russe qui s'incarne dans une approche de linguistique descriptive et typologique des phénomènes de figement morphosyntaxique et sémantique.

Après avoir développé le contexte historique d'émergence et de constitution du champ de la phraséologie, nous allons définir les caractéristiques de son objet d'étude, les *phraséologismes*.

1.2. Statut et objet de la phraséologie

La phraséologie constitue un domaine de recherche encore instable et peu unifié, pour lequel la définition même du statut reste problématique du fait de sa pluridisciplinarité (lexique, syntaxe, sémantique, stylistique, pragmatique, psycholinguistique, TAL, etc.). Elle a pour objectif d'inventorier, de définir et de typologiser le fonctionnement des séquences polylexicales dont les constituants présentent une forte spécificité de leur cooccurrence et un certain nombre de restrictions au niveau syntaxique, sémantique, pragmatique, voire interactionnel. L'objet phraséologique s'est toutefois solidarisé autour de 3 principaux critères définitoires stables, mais fortement discutés dans la communauté.

1.2.1. La fixité combinatoire

Ce critère est repéré très tôt chez les linguistes de la fin du 19e siècle et constitue un des critères les plus observés dans le domaine des études phraséologiques, considéré par certains comme le « socle fondateur de la phraséologie » (Kauffer 2014 citant Gülich et Krafft, 1997). La notion même est sujette à plusieurs débats très stimulants quant à la nature de la préfabrication, et les différents degrés de contrainte qui peuvent peser sur la distribution des constituants, contrainte souvent désignée par *figement morphosyntaxique*. Selon G. Gross (1996 : 54), « une séquence est figée du point de vue syntaxique quand elle refuse toutes les possibilités combinatoires ou transformationnelles qui caractérisent habituellement une suite

[4] Il est utile de noter que ces deux traditions, autrefois isolées, se fondent maintenant toutes les deux sur la linguistique de corpus et tendent à puiser des éléments théoriques de manière réciproque, l'une s'inspirant des avancées de l'autre, et vice-versa.

de ce type ». Le figement semble donc renvoyer aux contraintes distributionnelles observables sur l'ordre des constituants de la séquence préfabriquée et sur le sens actualisé qui en découle. Pour d'autres, le figement se caractérise également par une *saturation lexicale* de la séquence (Mejri 2004 : 23) ou l'association préférentielle des constituants (Ruiz-Gurillo 1997 : 86 dans Gonzàlez-Martìn 2019 : 36). Ce figement est souvent décrit comme scalaire (G. Gross 1996 ; Martin 1997 ; Mejri 2002, Lamiroy et Klein 2005 ; Kauffer 2013) et multifactoriel (Lamiroy 2008 : 89 ; Bolly 2010 : 13). Ce faisant, il rend compte de plusieurs niveaux de contraintes (niveaux morphologique, phonétique, lexical, syntaxique, schéma communicatif, etc.) et de plusieurs degrés d'intensité (du figement faible au figement total) de cette contrainte sur les constituants. Anscombre (2011) montrera par ailleurs que même à un degré apparemment faible de figement, il est possible de dégager des récurrences structurelles autorisant une certaine liberté combinatoire au niveau paradigmatique qu'il désignera par *matrices lexicales* (par exemple, les matrices des comparaisons à parangon : ê*tre fort comme un bœuf*, *être rapide comme l'éclair*, *être gai comme pinçon*, etc.). G. Gross (1996 : 22) conclut une quinzaine d'années auparavant et à l'issue d'un travail d'ampleur sur les *phrases figées* du français que les séquences complètement figées sont peu représentées en comparaison des séquences présentant des blocages transformationnels partiels. Et même lorsqu'il s'agit de segments à combinatoire « libre », c'est-à-dire produite par le locuteur en appliquant les règles combinatoires prototypiques inhérentes à la langue qu'il emploie, Mejri (2016 : 245) et G. Gross (1996 : 46) démontrent qu'un figement demeure observable sous la forme de contraintes syntaxiques, lexicales, grammaticales et sémantico-logiques qui vont conditionner l'acceptabilité des segments en question. Enfin, comme le rappelle Bolly (2010a), trois approches du figement (qui ne s'excluent pas) semblent se dégager des travaux récents :

- Une approche *fonctionnelle* (Bolly 2010a : 12) fondée sur des critères linguistiques et regroupant tout phénomène imposant une contrainte distributionnelle et transformationnelle sur les constituants, bloquant ainsi toute combinatoire morphosyntaxique ou variation sémantico-logique ;

- Une autre approche plus stricte et de nature *statistique* percevant le figement comme la cooccurrence préférentielle ou la spécificité d'association de plusieurs éléments lexicaux dans l'usage (Bolly 2010a : 12, Kauffer 2014 : 2).

- Une approche *cognitive* (Bolly 2010, Wray et Perkins 2000 :1, Svensson 2004) qui perçoit d'une part le figement comme un processus de « *mémorisation cognitive*, i.e. l'intégration mémorielle des unités phraséologiques comme des unités sémantiquement autonomes (enregistrée en mémoire discursive *as a whole*) » (Bolly 2010 : 13), d'autre part, qui perçoit le figement comme un processus d'institutionnalisation ou de conventionnalisation d'un emploi au sein d'une communauté linguistique donnée (Coulmas 1981 ; Bolly 2010 : 13).

Une tendance récente semble aller vers le remplacement de la notion de *figement* par la notion de *contrainte*, laquelle semble plus appropriée pour rendre compte des dynamiques internes et externes qui sous-tendent la préfabrication.

1.2.2. Le statut d'unité polylexicale

Les travaux menés jusqu'à présent ont dans l'ensemble montré qu'un phraséologisme est de nature intrinsèquement polylexicale, constituant ainsi un syntagme à la fois supérieur au mot et inférieur ou égal à la phrase complète. Ce syntagme est souvent sélectionné ou interprété en bloc par le locuteur lors de la communication. Toutefois, ce critère est contesté et le consensus quant au statut lexical des phraséologismes semble encore loin, ce qui complexifie leur traitement lexicographique. Du fait de cette mémorisation en bloc caractéristique des phraséologismes, il ne parait pas trivial de réexaminer cette notion de polylexicalité, car l'emploi se fait de manière unifiée, comme un tout. Cette unité dans la pluralité propre au phraséologisme donne à réflexion quant au statut lexical de cette construction, si bien que sa dénomination même suscite de vifs débats : doit-on intuitivement parler d'unité, c'est-à-dire d'un polymère de plusieurs éléments ? Doit-on parler de *séquence* et mettre en avant le caractère polylexical et successif ? Ou bien doit-on simplement considérer un tout lexicalisé et solidaire comme un *lexème* ? La question reste entière notamment du fait qu'il existe différents degrés de contrainte et de solidarisation entre les constituants[5]. Dans le cadre de cette thèse, nous emploierons principalement le terme de *phraséologisme* pour rendre compte de la catégorie des expressions figées, dénomination qui a le mérite selon nous de

[5] Sur ce dernier point, Agnès Tutin (LIDILEM) a évoqué lors de la relecture de ce manuscrit le fait que les éléments dont il est question conditionne le degré de cohésion. Ainsi dans le cas de la collocation *vibrant hommage* la question se pose quant au statut d'unité, ce qui ne sera pas le cas de *pomme de terre* par exemple.

prendre en compte le statut variable de préformation de ce type de phénomène, mais aussi le caractère composite de sa structure formelle.

1.2.3. L'idiomaticité sémantique

Le troisième et dernier critère est l'idiomaticité (Gréciano 1983). Ce critère désigne l'opacité[6] sémantique (ou *sens phraséologique*) résultant du processus de fixation de la séquence lexicale dans un schéma communicationnel stéréotypé et donnant lieu à l'apparition d'un sens plus complexe étroitement lié aux spécificités socioculturelles et aux propriétés sémantiques actualisables des constituants qui composent l'expression. L'idiomaticité soulève également la question de la compositionnalité sémantique, c'est-à-dire de la possibilité (ou non) de déduire le sens du tout à partir de l'addition du sens de ses parties. Une majorité des travaux semble s'accorder à dire que les phraséologismes sont généralement non-analysables et dotés d'un sens ou d'une fonction non-prédictible. En ce sens, l'actualisation du sens par les unités qui composent la séquence n'obéit pas au cadre d'association sémantico-logique habituel, rendant toute tentative d'interprétation complexe sans posséder au préalable le savoir conventionnel nécessaire à la compréhension. En outre, G. Gross (1996, p. 69–88) remarque que l'absence de paradigme synonymique (impossibilité de remplacer un constituant de la séquence par un synonyme sans en altérer le sens) est très répandue dans les phraséologismes.

Longtemps élevée en tant que critère définitoire, l'idiomaticité absolue a depuis laissé place à une perception graduelle du niveau d'opacité sémantique. Ainsi, Vinogradov (1946) puis Howarth (1996 : 32) partent de l'idée d'un continuum idiomatique allant d'une faible opacité sémantique et une faible restriction combinatoire à une idiomaticité complète, non analysable et non-prédictible. À sa suite, Schmale (2013) observe que « L'idiomaticité, loin d'être réservée aux seules expressions fortement métaphoriques et/ou imagées, est par conséquent une caractéristique graduelle des UCP[7] et peut aller d'un degré très faible à un degré très développé » (Schmale, 2013 : 36). Enfin le critère de compositionnalité suscite de

---

[6] Nous renvoyons le lecteur au travail de thèse de Svensson (2004) qui propose une typologie très fine des différents critères de définition de l'idiomaticité sémantique, ainsi qu'à une réflexion plus récente sur cette notion dans Svensson (2017).
[7] *Unité de Construction Préformée*

plus en plus de précautions dans son intégration à l'analyse des phraséologismes (Burger 2007 : 91) et peut mener à devoir prendre en compte une certaine part de subjectivité. En effet, un certain nombre de travaux ont montré une forte variabilité interindividuelle lors de l'évaluation de la compositionnalité, une faible concordance lors de l'accord interjuge ou interannotateur, et de manière relativement fréquente, un effet de *remotivation* du phraséologisme observé (Moon 1998 ; Geeraerts 2003 dans Legallois et Tutin 2013).

Suite au cadrage du champ de la phraséologie générale et des spécificités de son objet d'étude, nous allons à présent proposer un rapide tour d'horizon du domaine de la phraséologie au Japon

2. Le domaine de la phraséologie au Japon[8]

2.1. Shiraïshi le précurseur (1942)

La phraséologie au Japon apparait comme une discipline relativement jeune et encore peu développée qui n'a réellement débuté qu'à partir des années 1940 avec les premiers travaux lexicographiques de Shiraïshi (1942*). Le principal intérêt de ces travaux réside dans une première tentative de classification des différents types de structures idiomatiques (1942 : 147*) puis une typologie des procédés de leur construction sémantique (1950 : 43-44*). Se basant sur une revue des ouvrages de référence, Shiraïshi (1942 : 147*) propose une première définition de la notion de 慣用 (*kan'yô* = usage commun habituel) qui va servir de cadre de référence pour le domaine jusque dans les années 1970, en montrant que la grammaire standard ne peut suffire à rendre compte des spécificités de 慣用語 (*kan'yô go* = usage commun + mot) ou *termes idiomatiques*. Il ajoutera ensuite que cette notion renvoie à des structures linguistiques employées en tant que telles de manière conventionnelle, dont la distribution des constituants est contrainte et qui dispose d'un sens et de nuances d'emploi spécifiques. Shiraïshi (1950 : 43-44*) remarque également que le sens de ces expressions idiomatiques ne peut s'analyser uniquement à partir de l'analyse de leurs constituants lexicaux, et que chaque unité prise à part ne peut s'actualiser qu'en présence du terme qu'elle

[8] Pour des précisions plus détaillées et exhaustives sur l'histoire de la phraséologie au Japon, voir le travail de thèse de Wu Lin (2017) sur lequel s'appuient certains éléments de ce bref aperçu proposé dans notre thèse. Nous marquerons les références de seconde main repérées dans le travail de Wu (2016) par un astérisque (*).

précède. Il observera les procédés de construction du sens et remarquera la forte prépondérance de la métaphore, de l'analogie, de la métonymie ou de la synecdoque. Il remarquera que quand bien même les constituants permettraient une inférence sémantique de l'expression, le sens du tout demeure plus complexe que l'addition des sens de ses parties, c'est-à-dire que la concaténation des sens actualise un sens partagé et conventionnellement établi dans la communauté linguistique. Shiraïshi va ensuite jouer un rôle de premier plan dans la construction du domaine en abordant diverses facettes de ce qu'il va appeler ensuite 慣用句 (*kan'yô ku* = usage commun + phrase => *locution idiomatique*), jusqu'à finalement produire un premier dictionnaire des locutions idiomatiques japonaises (国語慣用句辞典 = *kokugo kan'yô ku jiten*) (Shiraïshi 1969*). On peut donc remarquer qu'à l'instar de Bally, Shiraïshi pose un certain nombre d'hypothèses qui vont permettre de cimenter le domaine. Toutefois, son travail reste à ce stade très descriptif et focalisé sur la forme, peu sur l'emploi.

2.2. Stagnation puis émergence du domaine de la phraséologie au Japon

La publication du travail lexicographique de Shiraïshi (1969*) a certes permis de proposer un premier inventaire des phraséologismes, mais ne détaillait pas assez les sous-catégories d'objet ni les critères de différenciation, laissant planer une certaine confusion sur la classification des différents phénomènes observés. Cette confusion se retrouvait par exemple dans le manque de distinction entre les parémies, les expressions rituelles ou les locutions verbales. D'autre part, les travaux de Shiraïshi sont longtemps restés marginaux et peu approfondis sur le plan linguistique et sur la description de l'emploi, ce qui a provoqué une perte de vitesse de la recherche en phraséologie (Wu 2017 :13). Ce n'est qu'à partir des années 1970 que le domaine va connaitre un retour à l'exploration des phraséologismes avec les études de Takaki (1974*), de Miyaji (1974*) et d'Okuda (1978*) qui vont initier de nouvelles approches exploratoires sur corpus littéraires en proposant des critères définitoires tels que la fixité morphosyntaxique, la compositionnalité sémantique, la séquentialité / polylexicalité et l'idiomaticité. Ces premiers travaux ont donné lieu à la concurrence de plusieurs dénominations parmi lesquelles 慣用句 (*kan'yô ku*), 慣用語 (*kan'yô go)* que nous avons vu précédemment, mais aussi :

- 慣用語句 (*kan'yô goku => terme composé d'usage commun*)
- 熟語 (*jukugo => terme composé*)
- イディオム (*idiomu => idiome*)

- 成句 (*seiku* => *locution figée*)
- 連語成句 (*rengo seiku* => *séquence locutionnelle*)
- 定型表現 (*teikei hyôgen* => *expression de forme fixe*)
- 凝結表現 (*gyôketsu hyôgen* => *expression cohésive* (*gyôketsu* = *solidification*))
- 凍結表現 (*tôketsu hyôgen* => *expression figée* (*tôketsu* = *gel*))

Cet inventaire est encore long et souligne à la fois cette difficulté de dénomination comparable à celle rencontrée en occident et fait principalement référence à des critères de forme. Miyaji (1974*) différencie également les parémies des locutions idiomatiques (ce que ne faisait pas Shiraïshi), car bien que fonctionnant sur des mécanismes de métaphorisation ou d'analogie, les locutions idiomatiques ne possèdent pas de valeur gnomique comme cela serait le cas avec les parémies. Il constate aussi que les locutions idiomatiques ne sont pas nécessairement figées de manière absolue et remarque que le figement peut s'analyser sous forme de continuum avec des expressions acceptant certains changements au niveau de leur combinatoire sans pour autant que le sens ou la fonction n'en soient altérés. Ce sera alors l'émergence de la notion de *figement graduel.* Takagi (1974*) de son côté va s'inscrire dans la lignée de Miyaji, tout en apportant la notion de *prêt-à-l'emploi* (既成品性 *kiseihinsei*) et la prise en compte des caractéristiques stylistiques.

Dans la lignée de ces nouveaux travaux, d'autres voient le jour parmi lesquels de nouvelles propositions de typologie de Shiraïshi (1977a* et 1977b*) avec notamment une extension aux onomatopées (擬音語 *giongo*), les termes d'adresse et de salutation (挨拶用語 *aisatsu yôgo*) et les termes mimétiques (擬態語 *gitaïgo*) ; et un approfondissement des précédents travaux chez Miyaji (1977*) au niveau des spécificités sémantiques et de la caractérisation du figement. Okuda (1978 : 40-41*) décrit toute une série de structures collocationnelles à noyau verbal fortement idiomatiques qu'il dénomme *groupement idiomatique* (慣用的な組み合わせ *kan'yôtekina kumiawase*). Il va notamment mettre en évidence, tout comme ses prédécesseurs, la non-compositionnalité de la structure binaire tout en ayant l'intuition que cette structure se compose d'un élément dont le sens reste fixe et d'un autre qui joue le rôle d'actualisateur sémantique du premier. Le travail d'Okuda marquera ainsi l'émergence des recherches sur les collocations en japonais.

2.3. Les années 1980 : Renouveau, structuration, consolidation

S'appuyant sur ses travaux antérieurs et ceux de Shiraïshi, Miyaji (1982 : 238*) propose par la suite une nouvelle typologie des phraséologismes, typologie majeure qui va cimenter le fond théorique du domaine et qui constituera le terreau des études subséquentes sur les phraséologismes. Sa définition des phraséologismes évoluera elle aussi et fera date dans le champ des recherches en phraséologie, mettant en avant un degré de fixité et une idiomaticité sémantique variable, ainsi qu'un usage en tant qu'unité solidaire et sélectionnée en bloc par le locuteur. Cette recherche aura un impact majeur qui va renouveler l'intérêt des chercheurs japonais qui jusque-là s'étaient quelque peu désintéressés du domaine et fixer les assises théoriques du champ. Ce renouveau théorique va inspirer toute une série de travaux, dont ceux de :

- Kunihiro (1985*) qui va beaucoup théoriser les critères définitoires des phraséologismes ;
- Nakamura (1977, 1985*) qui s'est intéressé à la notion de compositionnalité et d'abstraction du sens du tout par rapport à ses constituants dans la perspective de la linguistique cognitive ;
- Muraki (1985*) et Satô (1989) qui ont étudié les phraséologismes du point de vue de leur structure syntaxique ;
- Otsubo (1985*), Momiyama (1997) et Sakamoto (1982) qui ont observé les processus de formation par métaphore du sens des phraséologismes et proposeront là encore une typologie basée sur des critères sémantiques.

On observe également pour la première fois chez Morita (1985*) et Nishio (1985*) des études quantitatives sur de grands ensembles de données. On peut ainsi voir que c'est à cette époque que se stabilise le domaine autour de critères de définition proches de ceux observés en Europe à cette époque.

2.4. À partir des années 1990 : essor du domaine et enjeux

Dans la lignée des travaux de Miyaji (1982), on va assister à un essor de la linguistique phraséologique, notamment concernant les questions du figement, de l'idiomaticité et des phénomènes d'actualisation du sens. Parmi les travaux notables, on pourra citer ceux d'Oku (1990) qui va étudier la reconnaissance entre sens référentiel et sens idiomatique ; de Miura

(1996) qui va s'intéresser dans une perspective cognitive au lien entre audition, vision et compréhension des structures idiomatiques en japonais ; ou d'Ishida (1998) qui travaille sur la variabilité morphosyntaxique des mots composés (複合語 *fukugô go*) et des expressions idiomatiques. Elle proposera différents patrons variationnels en fonction du degré de fixité, puis proposera d'explorer la notion de degré d'idiomaticité en japonais. Ishida, qui a effectué une partie de ses recherches au Canada, va introduire une perspective très proche des courants théoriques européen et états-unien concernant le traitement des phraséologismes. Elle va dès lors constituer l'une des figures majeures de la phraséologie japonaise et anglaise au Japon avec une production scientifique très riche sur les phraséologismes dans plusieurs domaines, dont la linguistique descriptive, la linguistique contrastive et l'acquisition de la L2.

À la suite de ces travaux qui ont fondé le renouveau de la discipline au Japon, on va assister à un essaimage des recherches en syntaxe et en sémantique : Tsuchiya (2009) va proposer de comparer les propriétés de structure et les propriétés sémantiques des phraséologismes et des parémies japonais en se basant sur les dictionnaires. Se fondant sur les travaux de Mejri (2003), Takada (2011) étudie la notion de *stéréotypie sémantique* qu'il applique à la description de quelques expressions idiomatiques du japonais et du français. Fang (2014) va observer les mécanismes de genèse des phraséologismes sur le plan sémantique. Les premières études fonctionnelles observant la relation entre le contexte, le sens et la dimension pragmatique de l'emploi des phraséologismes émergent relativement tardivement dans les années 2000 avec les travaux d'Ishida (2003a, 2003b, 2004), Tsuchiya (2011), Momiyama (2009), Yamaoka (2004), Arizono (2013) et Wu (2016).

En parallèle, on observera une extension du domaine vers les études en didactique de la langue étrangère[9] ou de la LM (Ono et al. 1999) et en acquisition de la langue étrangère (Miyoshi 2007, Hori 2010, Kobayashi 2014, Fang 2016, Hiraï 2016). La plupart des travaux dans ces domaines vont dès lors être centrés sur les besoins d'apprentissage d'une population d'apprenants de plus en plus croissante avec notamment une explosion des travaux sur :

- La description des marqueurs discursifs (Togashi 2002, Iio 2006, Yang 2017, Wakamatsu 2021)

[9] *Langue étrangère* sera parfois abrégé *LE*

- L'emploi des onomatopées ou des termes mimétiques
- L'usage des particules finales, les expressions associées à des actes de langage (requête, excuse, etc.)
- Les expressions liées à des rituels linguistiques (salutations, présentation de soi, flatteries, etc.)

Toutefois, ces descriptions tendent à s'écarter de l'objectif initial de la linguistique descriptive et ne s'appuient que très peu sur les éléments de typologie et de définitions proposées dans les études antérieures. Elle vise plus à s'interroger sur la manière d'enseigner les phraséologismes plutôt que sur leurs spécificités linguistiques. En outre, même si les champs de la didactique et de l'acquisition se sont approprié ces questions, il n'en sera pas de même pour la linguistique descriptive pour laquelle les travaux vont rester assez centralisés au sein d'une petite communauté de chercheurs experts du domaine.

2.5. Unification et stabilisation du domaine autour de la linguistique de corpus

Plus récemment, un certain nombre d'études sur corpus ou d'études sur les techniques d'extraction des phraséologismes ont vu le jour. On pourra citer notamment Murata et Yamazaki (2011) qui ont pu démontrer dans un but didactique qu'il est possible d'observer une relation de spécificité entre structure idiomatique et genre littéraire et les études de Sasaki (2013) et Okada (2014) sur le changement sémantique, notamment le processus de passage du sens référentiel au sens phraséologique. Ishida (2014, 2015, 2019) s'intéresse au traitement lexicographique des phraséologismes dans le dictionnaire et va discuter leur identification dans le corpus littéraire japonais BCCWJ[10], notamment lorsque l'on observe pour une même expression plusieurs variantes concurrentes. Toujours en se fondant sur des données extraites à partir de corpus littéraires, Ishida (2000, 2015) se basant sur les travaux de Fraser (1970) va proposer une suite de 5 tests afférents permettant d'évaluer le degré de figement sémantique des phraséologismes observés qu'elle va appliquer sur des phraséologismes utilisant les parties du corps, catégorie extrêmement productive en japonais (cf. Inventaire des expressions basé sur les parties du corps, Oki 2003). Suzuki (2018) de son côté observera le phénomène de transformation sémantique au sein de phraséologismes

[10] https://clrd.ninjal.ac.jp/bccwj/en/

adjectivaux en se basant sur des données extraites des réseaux sociaux. Wu (2016a, 2016b, 2016c) va discuter dans 3 études fondées sur le corpus romanesque BCCWJ :

- Le phénomène de remotivation sémantique de l'expression 足を洗う (*ashi o arau* = *se laver les pieds* => 's'absoudre de son passé délinquant') (2016a)
- La mesure lexicométrique des phraséologismes à noyau verbal les plus fréquemment employés (2016b)
- La catégorisation sémantique de 900 idiomes (2016c) en fonction de leurs propriétés syntaxiques

La thèse de Wu (2017) qui se situe dans la droite lignée de celle d'Ishida (2002) constitue à ce jour l'un des travaux de recherche les plus aboutis sur la description des phraséologismes du japonais au moyen d'une approche sur corpus.

Du côté du TAL, un nombre croissant d'études sur la caractérisation, l'annotation et l'extraction des phraséologismes ont vu le jour depuis les années 2000, notamment :

- Ojima, Satô et Utsuro (2006) qui traitent de la construction d'une base de données des séquences idiomatiques

- Hashimoto, Satô et Utsuro (2007) puis Hashimoto et Kawahara (2008) qui ont travaillé sur la détection, l'identification et la modélisation des phraséologismes dans des corpus écrits de volume moyen

- Shudo, Takahashi et Tanabe (2012) concernant le traitement et l'annotation automatique des séquences polylexicales en japonais

- Iwatsuki (2021) portant sur l'extraction des séquences formulaires spécifique au japonais scientifique.

On remarque une grande diversité des travaux et un intérêt croissant pour les méthodes outillées depuis les 15 dernières années. Notons toutefois qu'en dépit de la stabilisation et de l'unification récente du champ de la phraséologie au Japon, ce domaine de recherche semble marginalisé au sein de la linguistique japonaise. Par ailleurs, une grande partie des travaux accessibles se fondent majoritairement sur des corpus littéraires, parfois peu représentatifs de l'usage actuel et se focalisent essentiellement sur trois types de

phraséologismes : les parémies, les locutions idiomatiques verbales et les collocations. Grezka et Kijima (2019) mentionnent à ce sujet que « La phraséologie ou le figement ne constituent pas encore des sujets majeurs de la linguistique japonaise » (Grezka et Kijima 2019 : 4), et ajouteront que « […] les études sur les séquences figées au Japon ne dépassent pas celles du classement. » (Grezka et Kijima 2019 : 4). Par ailleurs, force est de constater qu'il n'existe pas à ce jour, tout du moins à notre connaissance, de travaux typologiques et descriptifs fins concernant les phraséologismes pragmatiques de l'interaction orale[11] s'appuyant sur des données authentiques non élicitées.

Ce cadrage théorique concernant le domaine de la phraséologie au Japon nous a permis de faire un parallèle intéressant entre l'école occidentale et l'école japonaise, notamment concernant les approches mobilisées et les critères observés jusqu'à présent. Ce cadrage est tout à fait nécessaire pour comprendre le contexte épistémologique dans lequel s'est développé le domaine de la phraséologie japonaise et les enjeux auxquels elle fait face. Après avoir posé le cadre général de la phraséologie, nous allons à présent affiner notre tour d'horizon, en précisant notamment les raisons du basculement de la phraséologie générale vers d'autres approches.

3. Extension récente du domaine de la phraséologie

Le champ de la phraséologie en Europe constitue aujourd'hui un domaine en pleine effervescence et donne lieu à la production riche et transdisciplinaire que l'on connait actuellement, ainsi qu'à un foisonnement terminologique tout aussi riche reflétant les différents domaines de la linguistique qui se sont intéressés à la question[12] : *phrases préfabriquées*, *séquences figées*, *formules*, *routines*, *phrasèmes*, *idiomes*, *séquences préconstruites*, *séquence phraséologique*, *tournures*, *locutions*, *expressions idiomatiques* et *idiotismes* pour ne citer que les plus connus. Toutefois, la relative jeunesse de ce domaine de recherche par rapport à d'autres domaines plus établis et l'inflation terminologique dont

[11] Hormis les phraséologismes rituels et les marqueurs discursifs qui sont très largement discutés en didactique du japonais langue étrangère.
[12] Cf. Martins-Baltar (1997) et Gonzàlez-Rey (2021) pour une liste non-exhaustive des termes associés, mais rendant bien compte de cette inflation terminologique propre au domaine de la phraséologie.

il fait l'objet font qu'une classification opérationnelle et unifiée de la phraséologie semble encore hors d'atteinte (Pecman 2004 :130, Mel'čuk 2013 :129). Cette forte transdisciplinarité et l'apogée de la linguistique de corpus comme un incontournable des sciences du langage a ainsi donné lieu à une *extension* du domaine de la phraséologie (Legallois et Tutin 2013) vers plusieurs champs de la linguistique appliquée, allant de la terminologie à l'analyse du discours en passant par la didactique, la sociolinguistique, la stylistique, la psycholinguistique ou la linguistique computationnelle. Nous pouvons citer deux exemples concrets de cette extension du domaine de la phraséologie : le premier exemple concerne les travaux menés au sein du LIDILEM[13] sur la caractérisation du genre scientifique à l'aide de l'étude linguistique et la typologisation des phraséologismes spécifiquement associés à ce genre discursif (Jacques et Tutin 2018), avec notamment des applications en FOU[14] et dans la formation des doctorants allophones (Cavalla 2018). Le second exemple concerne les travaux menés dans le cadre de l'ANR Phraséorom (Novakova et Siepmann 2020) visant l'identification des genres littéraires fondés sur les emplois phraséologiques préférentiellement associés à chaque genre, association qui permet de dresser un profilage phraséologique desdits genres.

Une partie de cette extension a été impulsée par la linguistique de corpus et le TAL pour lesquels la caractérisation, le traitement et l'annotation des unités phraséologiques constituent un nouveau défi de taille très stimulant sur le plan scientifique. Les efforts scientifiques déployés dans ces deux domaines ont des retombées principalement dans le domaine de la traduction automatique (notamment avec le perfectionnement des corpus bilingues alignés ou comparables) et dans l'ingénierie pédagogique avec le développement d'outils d'apprentissage des phraséologismes guidés sur corpus. Enfin, une des extensions du domaine les plus intéressantes est sans nul doute celle proposée par Wray (2002, 2008) qui, employant une approche linguistique fondée sur les sciences cognitives, va tenter de caractériser les processus cognitifs de mémorisation, de reconnaissance et d'usage des phraséologismes. Cette approche innovante va notamment inspirer toute une série de travaux en acquisition qui se focaliseront sur les processus mentaux impliqués dans la mémorisation et la production des phraséologismes (Forsberg 2006).

---

[13] Laboratoire de Linguistique et Didactique des Langues Étrangères et Maternelles
[14] Français sur Objectif Universitaire

Un peu plus récemment, et avec l'apparition des premiers grands corpus oraux de la conversation quotidienne au Japon, dont le *nihongo nichijô kaïwa kôpasu* (*corpus du japonais de la conversation quotidienne* ou CEJC, Tanaka *et al.* 2018) constitué par l'Institut National de Linguistique Japonaise (*kokuritsu kokugo kenkyûjo* ou NINJAL), on assiste à un engouement croissant des chercheurs japonais pour la recherche sur les phraséologismes de l'oral. Ces travaux ouvrent de nouvelles perspectives méthodologiques et théoriques quant à la description des PhP, notamment avec la prise en compte multimodale des spécificités d'emploi. On pourra mentionner par exemple, Kumagaï (2014) qui étudie de l'emploi du terme expressif *yabaï* (expressif positif ou négatif) dans une perspective à l'interface de la lexicographie, de l'étude de l'analyse conversationnelle et de la sociolinguistique. On mentionnera également l'étude plus récente de Katô (2020) qui propose une exploration des fonctions pragmatiques du terme évaluatif et réactif *sugoï* (*très, trop*) en se basant sur une étude fine du CEJC. Ces divers travaux fondés sur l'étude de corpus oraux ouvrent la voie à de nombreuses applications en didactique, en acquisition et en analyse des interactions.

Que ce soit dans le monde de la recherche occidentale ou dans la communauté scientifique japonaise dans une moindre mesure, ces différents exemples nous montrent que les 15 dernières années ont assurément constitué un tournant crucial pour le domaine de la phraséologie. Cette extension transdisciplinaire du domaine d'application de la phraséologie a constitué un facteur d'émergence pour une sous-catégorie : la *phraséologie pragmatique*.

Dans le chapitre qui suit, nous allons définir le champ de la phraséologie pragmatique, puis nous procèderons à une synthèse des théories apparentées au domaine de la phraséologie pragmatique. Enfin, nous détaillerons point par point les spécificités de notre objet d'étude.

Chapitre 2 - La phraséologie pragmatique : vers un nouvel élargissement

Ce chapitre s'articule en trois temps. Dans un premier temps, nous allons détailler le sous-domaine d'étude des phraséologismes pragmatique en précisant un peu plus le lien entre pragmatique et phraséologie. Dans un second temps nous allons procéder à une revue des différents objets linguistique présentant des caractéristiques proches de notre objet d'étude. Enfin, nous procèderons à une définition de notre objet d'étude, puis nous conclurons par une synthèse intermédiaire qui ouvrira la voie à la deuxième partie de ce cadre théorique sur l'acquisition de la compétence pragmatique.

1. Le champ de la pragmatique

La pragmatique est une sous-discipline de la linguistique qui s'interroge sur les fonctions particulières que peut avoir un acte d'énonciation lors de l'interaction, tout en prenant en compte ses conditions de production. Elle constitue ce que Vanderveken (1989 : 74) définit par la « théorie générale de la signification du locuteur », ce qui est résumé par le titre de l'ouvrage fondateur d'Austin (1970 [1962]) *How to do things with words*[15].

La pragmatique est apparue aux États-Unis au cours du 19e siècle. Elle est née au sein de la philosophie du langage. Peirce (1931), considéré comme l'un des pères fondateurs de la discipline va poser les bases de la théorie de la pragmatique dans une perspective sémiotique et psycholinguistique (lien entre signe linguistique et pensée). Son disciple Morris (1938), va le premier penser la pragmatique comme système intégré au langage, proposant le trio syntaxe, sémantique et pragmatique pour définir les différentes dimensions interagissant dans la construction du signe linguistique et de ses fonctions. Morris va ainsi redéfinir le rapport entre la production linguistique et le locuteur jusque lors focalisé sur la production.

La pragmatique d'essence linguistique a émergé avec les travaux d'Austin (1970 [1962]) puis de Searle (1975, 1982) qui vont fonder la *Speech Act Theory* (Théorie des Actes du

[15] « Comment faire des choses avec les mots»

Langage). Une autre définition qui nous semble pertinente et appropriée à notre étude est celle de Crystal (1997) qui met en avant les contraintes du contexte de communication :

> *« La pragmatique est l'étude du langage du point de vue de ses locuteurs, notamment des choix qu'ils font, des contraintes qu'ils rencontrent dans l'utilisation du langage lors de l'interaction sociale et des effets de leur utilisation du langage sur les autres participants à l'acte de communication. »*[16] *(Crystal 1997, p.301)*

La pragmatique constitue donc une discipline visant à analyser le lien entre l'énoncé, le contexte et le profil de l'individu et de son interlocuteur. Le but de la communication et toutes les dynamiques qui président à l'usage de la langue en contexte d'interaction sont centraux pour la pragmatique. Ces dynamiques se retrouvent à la fois en production (besoin communicatif, stratégies conversationnelles, éthos communicatif), mais aussi en compréhension (inférence, connaissance partagée, règles conventionnelles, connaissance des rôles sociaux). L'énoncé est ainsi le produit de son énonciation (Ducrot 1980, Reboul et Moeschler 1994 : 22). En ce qui nous concerne, nous opterons pour une définition qui se nourrit à la fois du courant anglo-saxon et de la pragmatique intégrée française. Ce faisant, nous rejoignons Anscombre et Ducrot (1979), Ducrot et Schaeffer (1995) ainsi que Kauffer (2017) qui ont discuté l'importance d'un aller-retour entre réalisations linguistiques concrètes et les différents paramètres contextuels qui président à leur usage. Ce retour permet ainsi une compréhension empirique des dynamiques de la communication et des emplois des structures linguistiques qui lui sont propres. Cette définition appliquée et intégrée à la sémantique de la linguistique pragmatique est d'autant plus pertinente pour l'étude des phraséologismes pragmatique (cf. Kauffer, 2016, 2017 et 2018).

Après avoir évoqué quelques éléments sur la phraséologie et la pragmatique, nous allons à présent resserrer un peu plus notre cadre théorique vers la notion de *phraséologisme pragmatique*[17] (désormais *PhP*).

---

[16] "*Pragmatics is the study of language from the point of view of users, especially of the choices they make, the constraints they encounter in using language in social interaction and the effects their use of language has on other participants in the act of communication*." (Crystal 1997, p. 301)

[17] On doit cette dénomination à Burger (1982) et Hyvärinen (2011) (Kauffer 2017).

2. La phraséologie pragmatique au sens large

La phraséologie pragmatique a émergé dans les années 1970 avec l'apparition des premiers travaux lexicographiques sur les routines conversationnelles et (méta)discursives, ainsi qu'avec l'émergence des travaux sur l'acquisition de la compétence pragmatique en langue étrangère. Les linguistes espagnols[18] (Casares 1969 ; Zuluaga 1980 ; Corpas Pastor 1996 dans López Simó 2016) et allemands (Coulmas 1979 ; Burger 1982 ; Lüger 1983[19] ; Schneider 1989) sont parmi les premiers à s'être intéressé de manière récurrente aux PhP et notamment aux structures préfabriquées de la conversation en proposant des classifications étayées de la catégorie. À leur suite, les linguistes travaillant sur le français (Fónagy 1982 ; Martins Baltar 1995 ; Mel'čuk 1995) et les linguistes anglais (Moon 1998 ; Cowie 2001) vont à partir des années 1990 commencer à proposer quelques travaux parcellaires, mais très centrés sur la typologisation et finalement peu sur l'observation de l'usage. Ce n'est qu'au milieu des années 2000[20] que les chercheurs français vont développer toute une série de travaux centrés sur l'analyse des phraséologismes pragmatiques en adoptant plusieurs angles d'observation (Cf. section 4 de ce chapitre). Nous définirons les phraséologismes pragmatiques comme : 1) des séquences polylexicales très productives (à l'oral notamment), 2) mémorisées, sélectionnées ou interprétées en bloc par le locuteur, 3) dotées d'un figement morphosyntaxique variable, 4) d'une prosodie caractéristique avec des variations intonatives et d'intensité, 5) d'un sens plus ou moins opaque ne correspondant pas forcément au sens actualisé par ses unités, 6) de fonctions pragmatiques, (méta)discursives et interactionnelles et enfin, 7) qui sont sujettes à un usage conventionnalisé. Il peut s'agir par exemple de :

- Marqueurs intersubjectifs : *tu sais*, *tu vois*
- Expressions de confirmation : *c'est pas faux*, *ça roule*
- Rituel d'interaction : Salutation (*bonne soirée*) / Remerciement (*merci beaucoup*)
- Slogan publicitaire : *Des pâtes, oui, mais des Panzani !* (Panzani)
- Insultes : *Va te faire foutre !*

---

[18] Pour plus de détails sur la phraséologie générale en Espagne, consulter la thése de Mireia López Simó (2016) et Gonzalez Rey (2021) qui proposent un historique très détaillé des travaux effectué et de leur articulation au sein de la phraséologie.

[19] Il faut également mentionner Burger (2007) qui propose une typologie relativement fine. L'ouvrage est toutefois en allemand.

[20] A noter que nous ne prenons pas en compte dans cette datation les travaux sur les marqueurs discursifs polylexicaux dont le statut est encore sujet à discussion. Le cas échéant, il conviendrait de prendre en compte toute une série de travaux menés depuis les années 1960 sur les marqueurs discursifs.

- Appellatifs : *Ma chérie, Mes chers compatriotes, Très chers collègues*
- Pragmatèmes (Mel'čuk 2013) : *Défense de fumer ! Garde à vous !*
- Phrases situationnelles (Klein et Lamiroy 2011) : *Les carottes sont cuites !*

Cette liste n'est pas exhaustive et il existe un ensemble de constructions diverses et variées dont l'inventaire reste encore à faire. Ce champ affilié à la phraséologie se nourrit de nombreuses approches parmi lesquelles les théories de la pragmatique, de l'analyse conversationnelle, de la psycholinguistique, de l'analyse du discours et de la sociolinguistique. Cet attrait croissant pour ces constructions préfabriquées fréquemment observées à l'oral a donné lieu récemment à l'apparition d'un sous-domaine que nous allons détailler dans la section suivante.

3. La phraséologie pragmatique des interactions orales

Dans le cadre de ce travail de recherche, nous allons nous focaliser sur un sous-domaine de la phraséologie pragmatique impliqué spécifiquement dans l'étude du lexique préfabriqué de la conversation orale : la *phraséologie pragmatique des interactions orales* (ou *phraséologies des interactions* selon Tutin 2019). Cette sous-catégorie s'intéresse à la fois aux interactions orales de types variés (conversation quotidienne, interaction transactionnelle, discussion de travail, etc.) et aux interactions médiées sur les réseaux sociaux. Cette sous-classe des PhP regroupe essentiellement les expressions qui :

- Sont activées par certains schémas de communication[21]
- Permettent d'agir sur la situation, les interactants et la relation qui les unit
- Expriment une fonction expressive ou conative
- Sont principalement réactives et/ou évaluative
- Varient au niveau de leur forme ou du lexique employé en fonction du type d'interlocuteur et de relation qui les unit

Ce sous-champ de la phraséologie vise donc l'étude des séquences phraséologiques occupant des fonctions cruciales pour l'interaction. L'approche employée est à l'interface de l'étude

[21] Sans pour autant être contraint par le contexte comme le seraient les *pragmatèmes* (Mel'čuk 2013).

du lexique, de la pragmatique et de la description des fonctions communicationnelle de l'interaction orale. Dans notre étude, nous focalisons plus particulièrement notre attention sur un sous-genre de l'interaction orale, celui de la *conversation quotidienne informelle*. Après avoir contextualisé le champ théorique de la phraséologie et défini le niveau de granularité de notre étude, nous allons à présent procéder à une discussion de différents travaux menés sur des objets linguistiques apparentés aux PhPex. Cette étape préparera la définition de notre objet d'étude que nous effectuerons ensuite dans le chapitre suivant.

4. Revue des différents paradigmes de phraséologismes pragmatiques apparentés aux PhPex[22]

Nous allons dans les lignes qui suivent procéder à une revue critique de différents paradigmes de phraséologismes à fonction pragmatique utilisés dans l'interaction quotidienne. Nous procèderons comme suit : 1) aperçu théorique, 2) description des spécificités, 3) discussion. Nous tenterons dès que possible de faire ressortir les différences entre les objets linguistiques étudiés. Mentionnons que les paradigmes qui seront étudiés ici ont été sélectionnés d'une part pour leur proximité avec notre objet d'étude, d'autre part pour les recherches soutenues dont ils ont fait l'objet. Nous proposons un tableau récapitulatif dans l'annexe 1. Cette discussion sera ordonnée de manière thématique.

4.1. Les approches généralistes

Cette section traite des approches généralistes visant à proposer une typologie générale des phraséologismes pragmatiques spécifiques à l'interaction orale. Nous ne détaillerons pas ici l'ensemble des typologies introduites, mais nous focaliserons notre attention sur les sous-classes proches des PhPex.

---

[22] Nous attirons l'attention du lecteur sur le fait que cette partie concerne principalement les recherches menées sur le français. Signalons toutefois qu'il existe un certain nombre de travaux sur les phraséologismes pragmatiques de l'espagnol (cf. López Simó 2016 et Gonzalez Rey 2021), de l'allemand (cf. Kauffer sur les ALS en allemand, 2016) ou de l'anglais (*idioms with pragmatic point* Fillmore *et al.* (1988), *speech formulae* Cowie 2001, *pragmatic routines* Alcón Soler et Sánchez Hernández 2017, entre autres) dont nous ne rendrons pas compte ici, faute de place et d'expertise dans les langues dont il est question. Pour ce qui est du japonais, les ressources documentaires à notre disposition ne nous ont pas permis de trouver des travaux proches de nos PhPex.

4.1.1. Locutions exclamatives, Bally (1909)

Comme évoqué précédemment, Bally est le premier à avoir proposé une typologie des phraséologismes pragmatiques avec une orientation vers la communication orale. Au sein des unités phraséologiques, Bally va distinguer une sous-catégorie regroupant divers procédés linguistiques dont la principale fonction est d'exprimer la subjectivité et le positionnement émotionnel du locuteur qu'il va appeler la *phraséologie exclamative*. Au sein de cette catégorie, il va discuter l'existence des *locutions expressives* dont le rôle principal est d'exprimer l'affect du locuteur (Bally 1909 : 295). Cette catégorie se caractérise notamment par une polylexicalité variable[23], une autonomie syntaxique, un figement syntaxique élevé, un sens non compositionnel, ainsi qu'une dimension expressive forte[24].

Bally aura l'intuition précoce de la nécessité de la prise en compte de la prosodie dans la description de cette catégorie (Bally 1909 : 164). Il montrera en effet l'importance de l'intonation pour désambiguïser le sens d'une locution exclamative lorsqu'elle admet plusieurs acceptions (cf. ex. : ***Allons !*** *du courage* / ***Allons !*** *je me résigne!* /***Allons!*** *ne vous fâchez pas!*, Bally 1909 : 299[25]). Nous verrons dans notre analyse et dans les entretiens post-test menés que ce critère joue un rôle très important dans les stratégies d'interprétation du locuteur non-natif. Sur le plan pragmatique, ces expressions reflétant la subjectivité du locuteur permettent d'exprimer un sentiment et une visée discursive particulière (ex. : « fi donc ! voilà qui est trivial » (idem : 299) => exprime le dégout et la réprobation). En outre, Bally montre l'importance de l'usage de ce type d'expression pour la fluidité relationnelle de l'interaction. Enfin, il montre d'une part la nécessité du contexte pour l'interprétation du sens et de la fonction, d'autre part, que le figement de l'expression résulte d'un processus de routinisation : « […], mais, à force d'être répétés, ils arrivent à recevoir un caractère usuel et à former même des unités indissolubles » (Bally 1909 : 117). Les PhPex observés dans cette thèse se situent dans la droite lignée du travail de Bally, notamment en ce qui concerne les

---

[23] Bally (1909) intègre dans ce paradigme à la fois des expressions polylexicales sémantiquement figées telles que « la belle affaire ! » ou « c'est bien fait ! » et des expressions monolexicales transparentes telles que « allons ! », « voyons ! », « jamais ! ».
[24] Il discute notamment des objets linguistiques qui sont communément intégrés à la catégorie des interjections.
[25] La mise en gras est de notre fait.

spécificités formelles et prosodiques. Nous nous écartons cependant de sa définition par le fait que les PhPex se limitent uniquement aux structures polylexicales à noyau verbal conjugué qui constitue principalement des phrases complètes et autonomes sur le plan syntaxique.

4.1.2. Formules de la conversation, López Simó (2016)

López Simó (2016) définit les *formules de la conversation* comme des énoncés réactifs complets et autonomes, utilisées dans la conversation quotidienne, fixes sur le plan syntaxique et dont la structure peut varier du syntagme (*bon appétit !*) à la phrase (*Tout l'honneur est pour moi !*). En outre, ces expressions routinières se caractérisent à la fois par une dimension rituelle, conventionnelle et comportementale (López Simó 2016 : 147). Cette classe de phraséologismes est large et inclut tout un ensemble d'objets variés tels que les *formules métaénonciatives*, les *phrases rituelles* ou les *marqueurs discursifs*.

Sur le plan pragmatique, les formules de la conversation se définissent par un lien étroit avec le contexte d'énonciation et la visée illocutoire du locuteur (López Simó 2016 : 167). Elles peuvent être réactives face à un élément déclencheur qui peut constituer une énonciation (*de rien !* employé après *merci !*), un état des choses (*Bon appétit !* lorsque le diner est prêt et que tout le monde est à table) ou une action particulière (*tu te fiches de moi !* après un comportement inadéquat). Le stimulus déclencheur de la réaction est désigné par López Simó en tant que *motif d'énonciation.* López Simó (2016 : 168-169) mentionne que le cadre spatio-temporel (lieu, temporalité, type d'interaction) et la relation unissant les interactants (relation hiérarchique, distance émotionnelle, connivence, etc.) constituent des facteurs décisifs dans la caractérisation des spécificités des formules de la conversation. En outre, elle prend également en compte dans sa description le comportement de l'expression au sein de la séquence interactionnelle (López Simó 2016 : 169-170). Au sein de sa typologie, c'est la catégorie des *formules personnelles* qui semble la plus proche des PhPex. La classe des *formules personnelles* renvoie aux formules exprimant la subjectivité du locuteur en contexte. Cette catégorie n'implique pas nécessairement une configuration dyadique. La fonction première des formules personnelles est d'exprimer l'état mental, affectif ou l'attitude du locuteur face à un état des choses de la situation d'énonciation. Elles peuvent aussi être de nature cognitive en expriment une modalité épistémique ou déontique propre au locuteur (López Simó 2016 : 180). López Simó proposera la dénomination *d'expressions*

symptomatiques pour désigner les expressions qui permettent de verbaliser les processus cognitifs et l'état d'esprit du locuteur.

Une des originalités du modèle de López Simó (2016) est la prise en compte des paramètres interactionnels et son ancrage fort aux données orales authentiques. Le travail de description effectué en aval est aussi très fin et permet de donner des pistes intéressantes dans la typologie des phraséologismes pragmatiques de la conversation. Nous nous différencions cependant notre définition de celle de López Simó par le fait que nous incluons dans notre classe uniquement des séquences ayant le statut de phrase, excluant de fait les interjections et les monolexicaux.

4.1.3. Phrases préfabriquées des interactions, Tutin (2019)

Les phrases préfabriquées des interactions (ou *PPI*) constituent des séquences :

- Préconstruites et généralement non compositionnelles
- Fréquemment employées dans l'interaction quotidienne
- Essentiellement polylexicales
- Principalement complètes sur le plan syntaxique et dotées d'un verbe à temps fini[26]

L'emploi des PPI et leur interprétation dépendent du contexte d'emploi, mais elles ne sont pas nécessairement déclenchées par ce contexte, ni même contraintes. En ce sens, les PhPex rejoignent les PPI. Ces phrases peuvent tantôt être une réaction expressive ou conventionnelle face à un stimulus se manifestant dans la situation, tantôt constituer une structure linguistique employée de manière à réguler et fluidifier l'interaction. Le travail de Tutin (2019) constitue avant tout un travail de typologisation fondé sur des critères pragmatiques, syntaxiques et sémantiques et voué à être approfondi dans les prochaines années à l'aide d'une étude empirique basée sur des données orales authentiques[27]. Ce travail

---

[26] A noter que ce critère constitue en réalité un choix méthodologique formulé par l'auteure pour les besoins de son étude et étant donné la grande variabilité des structures syntaxiques observées dans la catégories des expressifs

[27] Cf. Projet PREFABS (Extraction et analyse des phrases préfabriquées de l'oral), sous la coordination d'Agnès Tutin (LIDILEM) : https://lidilem.univ-grenoble-alpes.fr/node/16/axes-recherche/axe-1-description-et-modelisation-linguistiques-corpus-tal/projets-laxe-1/projet-prefabs-extraction-et-analyse-phrases-prefabriquees-loral

de typologie a donné lieu à l'établissement d'une typologie en 4 catégories (Tutin 2019 : 68-72) au sein de laquelle les *phrases réactives situationnelles* présentent de nombreuses similitudes avec les PhPex. Il s'agit de phrases employées en réaction à un événement ou un état des choses apparaissant dans la situation d'énonciation et marquant la subjectivité du locuteur en contexte. Elles sont liées à un schéma conversationnel particulier et peuvent constituer des phrases évaluatives et/ou expressives centrées sur le locuteur (*que veux-tu que je te dise ?*) ou des phrases à fonction interactionnelle pouvant exprimer par exemple l'approbation (*tu l'as dit ! c'est clair !*) ou le désaccord (*tu parles !*).

Ce travail original offre une vue d'ensemble comparable à celui de López Simó (2016). On constate en effet que ces deux travaux ont abouti à un certain nombre de classes similaires. Le travail de Tutin propose une typologie fonctionnelle à l'interface de la structure formelle et de ses réalisations concrètes à l'oral. Il met également l'accent sur la prise en compte des données extralinguistiques, des comportements sociaux et des dynamiques interactionnelles, dimensions bien souvent écartées dans les typologies observées jusqu'à présent. Les PhPex présentent de nombreux points communs avec les phrases réactives situationnelles, notamment leur statut phrastique et leur caractère réactif et expressif.

4.1.4. Phrases préfabriquées, Dostie (2019, 2021)

Dostie (2019) définit les phrases préfabriquées comme :

- Des séquences autonomes et polylexicales
- Mémorisées ou employées telles quelles
- Sujettes à plusieurs degrés de contraintes sur l'axe paradigmatique / syntagmatique
- Souvent associées à un affaiblissement de la compositionnalité sémantique (Dostie 2019 : 35)
- Dotées d'une dimension actionnelle forte permettant la réalisation d'un acte illocutoire.

Les phrases préfabriquées ont une dimension résolument interactionnelle, avec plusieurs niveaux de fonction allant de la régulation phatique (*tu vois*, *tu sais*) à la régulation interpersonnelle (*salutation*, *excuses*, etc.). Elles peuvent ainsi constituer un tour de parole de par leur autonomie syntaxique. Dostie (2019) propose 4 critères définitoires partant du

principe que les phrases préfabriquées « servent fondamentalement des besoins communicationnels » (Dostie 2019 : 36-37) : i) la fonction générale du message, ii) l'intégration et le rôle dans la relation émetteur-récepteur, iii) les connaissances socioculturelles, conventionnelles et comportementales impliquées dans l'interprétation du message et iv) l'*univers de discours* siège de l'emploi du message. À partir de ces 4 critères, Dostie a pu établir une typologie en trois temps (Dostie 2019 : 40-50) au sein de laquelle la classe des *phrases expressives* nous intéresse tout particulièrement. Les phrases expressives constituent des phrases préconstruites et réactives dont la principale fonction est d'exprimer un état émotionnel, psychologique ou une attitude du locuteur face à un état des choses ou un événement se déroulant dans la situation d'énonciation. Elles constituent des traces de la subjectivité du locuteur, mais n'ayant aucune fonction structurante ou ordinatrice comme observé dans une autre catégorie de sa typologie, les *métaphrases*. Elles sont spontanément déclenchées par un stimulus extérieur et ne participent pas spécifiquement à réguler le canal de communication, de ce fait, elles ne sont pas associées aux *phrases interpersonnelles*. Il s'agit d'expressions telles que *c'est le comble ! mon œil !* ou *c'est une honte !*.

Le travail de Dostie (2019) fait écho aux typologies de Tutin (2019) et de López Simó (2016) avec lesquelles on observe le recoupage de certaines classes. Un autre intérêt de ce travail est qu'il s'inscrit dans la dynamique actuelle des efforts entrepris très récemment pour le cadrage théorique, la classification et l'uniformisation terminologique des phraséologismes pragmatiques. Ce travail s'appuie également sur des données authentiques et sur l'observation de formes, mais aussi des fonctions interactionnelles et interpersonnelles dans plusieurs genres communicatifs. La sous-classe des PhPex semble également très proche des phrases expressives définies par Dostie (2019). Les phrases expressives acceptent cependant un plus grand paradigme de constructions syntaxiques, ce qui n'est pas le cas des PhPex qui se limitent aux structures phrastiques verbales à temps conjugué.

4.2. Les approches lexicographiques

Cette section traite des approches lexicographiques orientées autour de la description des phraséologismes expressifs de l'interaction orale. Ces différentes approches ont grandement nourri notre perception des phénomènes linguistiques expressifs et ont contribué à l'identification de certaines des spécificités des PhPex.

4.2.1. Structures figées de la conversation, Bidaud (2002)

Dans le cadre d'un travail lexicographique contrastif (français-italien) à visée didactique fondé essentiellement sur des ressources écrites ou fabriquées orales[28], Bidaud (2002) décrit des expressions figées de registre principalement familier, bivalentes (peut avoir plusieurs contextes d'usage), peu imagées ou figuratives et constituant des réponses expressives à des stimulus. Ces expressions sont caractérisées par ce que Bidaud désigne par *figitude*, c'est-à-dire d'une part un blocage des propriétés combinatoires de l'expression sur le plan syntagmatique (pas de commutabilité) et sur le plan paradigmatique (pas de substituabilité), d'autre part, un blocage sémantique caractérisé par un sens global, indécomposable (non compositionnel), non-prédictible et idiomatique. Cette figitude est le résultat d'un phénomène de répétition amenant la fixation et la conventionnalisation d'un usage qui lui-même favorise l'apparition d'une fonction de communication. Sur le plan sémantique, Bidaud oppose le *sens syntaxique* [29] qui constitue l'acception compositionnelle ou référentielle de l'expression, à son sens situationnel qui constitue son *sens idiomatique* et non compositionnel. Ce sens idiomatique est activé par le contexte habituel d'emploi, contexte dans lequel ces expressions s'imposent comme choix prioritaire. En ce sens, Bidaud rejoint Kauffer (2011) qui relève la même ambiguïté dans ses ALS.

Les structures figées de la conversation sont en principe compactes et économiques sur le plan de la forme. En effet, la conversation courante est de nature rapide, impromptue et imprévisible, laissant peu de temps à la réflexion et à l'élaboration du propos. Pour pallier ce caractère spontané et dynamique, le locuteur aura de fortes propensions à l'emploi de structures préfabriquées, plus faciles à mémoriser et moins couteuses sur le plan cognitif car concentrant un sens complexe dans une forme compacte. En outre, ces expressions se

---

[28] Notamment pour les expressions au sujet desquelles il n'existe à l'heure actuelle aucune ressource lexicographique ou de données orales observables.

[29] Nous prenons ici une certaine distance quant à cette dénomination. En effet, bien que nous comprenions l'idée de l'auteur au travers de cette dénomination, il n'en demeure pas moins curieux d'associer la notion de sens à la dimension syntaxique.

retrouvent principalement dans le registre familier, mais Bidaud rappelle que le concept de registre ou de niveau de langue est très fluctuant et que ces notions ne constituent pas des catégories fixes, certains emplois pouvant très bien naviguer entre plusieurs registres si les paramètres de l'interaction le permettent. Il est également mentionné une dimension expressive, voire émotive, accompagnée de gestuelle, de mimique et de saillance intonative. De par leur caractère expressif et réactif, les structures figées de la conversation ont de faibles chances d'être employées en ouverture de la communication, car elles sont essentiellement insérées dans un échange de tour de parole et nécessitent un contenu propositionnel antécédent ou une action de la part de l'interlocuteur pour être provoquées.

Bien que cet inventaire constitue un travail lexicographique de grande qualité, Bidaud (2002) insiste sur le fait que ce travail n'est que préliminaire et mériterait de nombreux approfondissements du travail de surface déjà effectué sur le plan linguistique, interactionnel et pragmatique. Elle suggère ainsi très succinctement à la fin des prolégomènes de son ouvrage quelques pistes d'approfondissements possibles. Il ne s'agit donc pas d'un travail d'analyse approfondie du fonctionnement des phraséologismes pragmatiques, mais d'un travail précurseur ayant permis l'élaboration d'un inventaire qui sert de repère très utile pour de nombreux travaux sur les phraséologismes pragmatiques.

4.2.2. Actes de langage stéréotypés, Kauffer (2011)

Kauffer (2018) dans une production riche et régulièrement actualisée définit les *Actes de Langage Stéréotypés* (ALS) comme :

- Une expression à fonction pragmatique autonome sur le plan syntaxique
- Figée morphosyntaxiquement (elle résiste aux tests afférents et transformationnels) et pragmatiquement (elle a un sens et une fonction actualisables uniquement en contexte)
- Généralement polylexicale et non- compositionnelle
- Fortement idiomatique
- Dotée d'un substrat émotionnel (indignation, joie, doute, etc.)

Kauffer (2011) ajoute que l'ALS constitue un acte de parole. À ce titre, il constitue une énonciation qui s'inscrit dans le cadre d'une interaction et implique une fonction

actionnelle[30] agissant sur le contexte et l'interlocuteur. Pour illustrer les ALS, nous pouvons citer *à la bonne heure !* (satisfaction) ou *tu vas voir ce que tu vas voir !* (menace). Les travaux de description des ALS de Kauffer se sont principalement orientés vers deux perspectives : la description lexicographique et l'analyse pragmatique.

Kauffer mentionne le fait que les unités qui composent certains ALS peuvent tantôt actualiser un sens littéral tantôt compositionnel lorsqu'ils sont associés à un contexte n'impliquant pas de fonction pragmatique particulière. Par exemple, l'ALS *mon œil !* peut marquer l'incrédulité dans un contexte, mais cette séquence peut tout simplement désigner l'œil du locuteur dans un autre. Nous verrons dans le cadre de notre analyse que cette cohabitation d'un sens phraséologique et d'un sens littéral peut poser des problèmes d'interprétation chez le locuteur non-natif. Concernant le statut d'énoncé, Kauffer (2018) précise que ce statut s'acquière du fait que l'ALS constitue « une unité de communication employée dans un certain contexte » (Kauffer, 2018 : 72), de manière autonome et se suffit à lui-même sur le plan informationnel. Il montrera malgré tout que d'une part, certains ALS peuvent être expansés (ex. : ***La belle affaire*** *que de s'être embrouillé dans la filiation d'une servante !*), d'autre part, certains ALS peuvent se combiner (ex. : *Ben voyons, dans tes rêves !*) (Kauffer 2018).

La fonction de l'ALS est intrinsèquement pragmatique. Il constitue une unité ayant à vocation de produire un acte de langage à la fois illocutoire (par ex. : convaincre son interlocuteur) et perlocutoire (par ex. : faire changer d'avis son interlocuteur en essayant de le convaincre). L'énonciation d'un ALS vise avant tout une dimension actionnelle et fondamentalement interactionnelle. En outre, Kauffer (2018 : 74) remarque que l'ALS est en général réactif en réponse à un stimulus issu du contexte d'énonciation.

Le dernier point soulevé par Kauffer concerne les limites de l'application du cadre des *actes de langage* tels que définis par Austin (1970) et Searle (1975). En effet, le modèle en l'état semble trop restrictif et peu adapté à la réalité d'une interaction réelle et authentique (Kerbrat-Orecchioni, 2009 [2001] : 53). Face à ce problème théorique, Kauffer (2018 : 77-78) proposera des extensions au modèle d'Austin et Searle parmi lesquelles la dimension

---

[30] Par *fonction actionnelle*, nous entendons le fait que les phraséologismes pragmatiques réalisent des fonctions pragmatiques que se manifestent sous la forme d'actes de langage ou de réactions de la part de l'interlocuteur. Le langage n'est alors plus représentatif (ou dénotatif), mais permet d'agir sur différents niveaux de la situation d'énonciation.

affective (les sentiments et l'état psychologique des interlocuteurs engagés dans l'interaction) ou la dimension ***cognitive*** (état des connaissances sur l'interlocuteur, sur la situation de communication ou sur l'état de certitude ou de vérité de ce qui est évoqué). En ce sens, il précède les travaux très récents de Krzyżanowska, Grossmann & Kwapisz-Osadnik (2021) qui mentionnent les mêmes limites de la théorie des actes de langage telle que développée par Austin (1970) et Searle (1975).

Bien que proches des PhPex, l'approche des ALS semble s'écarter de notre définition pour plusieurs raisons. D'une part, le choix des expressions traitées pose la question de l'actualité. En effet, nous avons pu remarquer qu'un certain nombre d'expressions décrites par Kauffer semblaient en perte de vitesse actuellement (par exemple : *macache !, la belle affaire !, des clous !, viens-y !, c'est fort de café !*, etc.). D'autre part, les ALS incluent les monolexicaux ainsi que les phraséologismes averbaux, ce qui n'est pas le cas des PhPex qui incluent principalement des structures verbales à temps conjugué. En outre, nous accordons peu d'importance à la notion d'acte de langage que nous jugeons trop floue pour définir les différents types d'illocutoires des PhPex. Enfin, une bonne partie des analyses des ALS s'appuient sur des romans[31] et l'oral représenté, mais peu sur les données authentiques orales. Dans le cadre de cette thèse, nous avons principalement observé le fonctionnement des PhPex au sein de corpus oraux ou sur les réseaux sociaux afin d'observer leur spécificité d'usage en contexte écologique.

4.2.3. Formules expressives de la conversation, Gharbi (2020) et Krzyżanowska, Grossmann & Kwapisz-Osadnik (2021[32])

Les *formules expressives de la conversation* sont le fruit de plus de deux ans de travail dans le cadre du Polonium PRAGMALEX[33] initié entre les équipes l'Université Marie Curie-Skłodowska, l'Université Grenoble Alpes et l'Université de Silésie, projet ayant pour finalité la description des phrases expressives fréquemment employées dans la conversation dans une perspective contrastive (français-polonais-italien). Krzyżanowska, Grossmann &

---

[31] Notamment en raison du travail contrastif avec l'allemand nécessitant des données comparables.
[32] Les premiers travaux ont en réalité débuté en 2018, l'année mentionnée réfère à la publication de l'ouvrage synthèse du travail effectué par les différentes équipes de ce projet interuniversitaire. Nous mentionnons le fait que nous avons participé à la réalisation de ce travail lexicographique au sien de l'équipe française du projet.
[33] https://www.umcs.pl/fr/descriptif-du-projet,15298.htm

Kwapisz-Osadnik (2021) définissent les formules expressives de la conversation comme des séquences autonomes plus ou moins figées, de compositionnalité variable, dont l'emploi est contraint sur le plan pragmatique et dont la principale fonction est d'exprimer l'état mental et affectif du locuteur face à son acte d'énonciation. Ce sont des énoncés évaluatifs sélectionnés en bloc et produisant des réactions émotionnelles conventionnelles face à un dire, un événement ou un comportement. Ces énoncés sont associés (mais pas contraints) à un cadre de communication stéréotypé, leur emploi est favorisé par rapport à d'autres expressions en raison de leur caractère idoine sur le plan pragmatique. Ils permettent d'exprimer une réaction émotionnelle complexe à l'aide d'expressions compactes et adaptées à la situation de communication.

La dimension interactionnelle est une caractéristique fondamentale des formules expressives de la conversation (Krzyżanowska 2020). Les formules expressives constituent des actes de langage qui s'échangent lors de l'interaction quotidienne. Leur expressivité ne se résume pas seulement à l'affect, mais ces expressions permettent également d'exprimer l'attitude du sujet parlant, son appréciation subjective de l'acte énonciatif. Krzyżanowska, Grossmann & Kwapisz-Osadnik (2021) remarquent que certaines formules expressives peuvent actualiser plusieurs sens en fonction du contexte (polyfonctionnalité) alors que d'autres ne s'emploient que dans un schéma interactionnel restreint. En outre, les expressions sélectionnées s'emploient généralement dans des situations de communication courantes ou familières, c'est-à-dire impliquant un abaissement progressif de la distance interactionnelle.

Gharbi (2020), qui a participé étroitement à la réalisation du projet PRAGMALEX, propose dans son travail de thèse une définition que se situe dans la droite lignée de travaux de Krzyżanowska, Grossmann & Kwapisz-Osadnik (2021). Elle définit les formules expressives de la conversation comme des énoncés entiers, intégrables dans d'autres énoncés, généralement non compositionnels, parfois polyfonctionnels et figés sur le plan pragmatique en raison de leurs modalités spécifiques d'emploi en contexte. Ces énoncés sont généralement employés en situation de communication familière et impliquent une dimension sociocognitive. Chez Gharbi (2020), les questions de l'acte de langage réalisé et de la dimension affective dans l'emploi des formules sont fondamentales et sont étroitement liées. Elle citera ainsi l'exemple de l'acte de menace qui est souvent associé à de l'indignation ou de la tristesse. Par ailleurs, l'auteure propose plusieurs critères pour identifier les *formules expressives* (Gharbi 2020 : 49-55) :

- **Exclamativité et expressivité :** trait caractéristique permettant au locuteur d'exprimer son affect et son ressenti concernant son dire. L'expressivité renvoie également aux traits prosodiques saillants caractéristiques d'un affect en particulier.

- **Propriétés lexicales :** les formules expressives sont des expressions polylexicales toutes faites, mémorisées en bloc, généralement lexicalisées et sujettes à un blocage combinatoire sur l'axe syntagmatique et paradigmatique. Ces formes sont généralement conventionnelles et excluent d'autres constructions concurrentes.

- **Compositionnalité :** les formules expressives sont non-compositionnelles à des degrés variables, souvent en relation avec le figement morphosyntaxique.

- **Lien avec le contexte :** les formules expressives sont étroitement liées à un cadre de communication, mais non contraintes dans leur emploi par la situation. Ce sont des réactions automatiques en réponse à des stimuli linguistiques ou extralinguistiques.

- **Autonomie syntaxique :** les formules expressives constituent généralement des structures autonomes ayant le statut d'énoncé. Pourtant, certaines formules peuvent s'intégrer à une autre proposition si la syntaxe l'autorise.

Ces deux perspectives autour du projet PRAGMALEX sont originales et présentent de fortes similitudes avec la classe des PhPex. Elles contribuent grandement à une meilleure compréhension des phrases expressives et donnent des outils utiles pour la description de ce phénomène. Les paramètres proposés sont aboutis et semblent opératoires. Nul doute que ce projet a ouvert la voie à toute une nouvelle perspective de recherche centrée sur l'analyse empirique et la prise en compte des paramètres interactionnels, sociolinguistiques, culturels, prosodiques et multimodaux (gestuelle, attitude, etc.). Par ailleurs, la perspective contrastive avec le polonais et l'italien rejoint le travail de Kauffer sur le français et l'allemand. Toutefois, notre travail de thèse diffère de cette approche sur certains points. Tout comme les ALS, les formules expressives constituent une classe large incluant les structures averbales (*bon courage !, ça alors !*) voire certaines formes contractées monolexicales (*courage !*), ce qui n'est pas le cas des PhPex. Par ailleurs, à la différence des formules expressives, nous nous concentrons également sur la prise en compte des formes

phraséologiques très familières, voire vulgaires. Cela s'avère d'autant plus pertinent que les résultats des locuteurs natifs au test lexicopragmatique ont montré une forte fréquence d'emplois relevant du registre très familier.

4.3. Les approches interactionnistes

Cette section traite des approches interactionnistes mobilisées dans la description des séquences préfabriquées fréquemment observées lors de l'interaction orale. Ces différentes approches ont permis d'élargir notre compréhension des phénomènes expressifs aux dynamiques sociales et interpersonnelles.

4.3.1. Routines conversationnelles, Coulmas (1981)

Les linguistes interactionnistes semblent être les premiers à avoir traité des séquences préfabriquées, pragmatiquement figées, sélectionnées en bloc par le locuteur et s'inscrivant dans les rituels d'interactions de la vie de tous les jours. Ces travaux se sont peu intéressés à la description formelle, mais ont surtout décrit les fonctions pragmatiques et les spécificités interactionnelles et culturelles associées à l'emploi de ces formules propres à la conversation. La notion de *routine conversationnelle* chez Coulmas renvoie à des mécanismes linguistiques qui sous-tendent l'usage des expressions et des énoncés formulaires utilisés de manière systématique dans les rites d'interaction de la vie de tous les jours. Ainsi, Coulmas (1981) définit la routine conversationnelle « non pas simplement comme une expression ou une stratégie, mais une expression dédiée à une situation bien particulière et liée à un certain type de stratégie ayant pour vocation de servir un projet communicatif »[34] (Coulmas, 1981 : 16).

La routine conversationnelle constitue une forme linguistique préfabriquée, employée par le locuteur pour accomplir un échange dont les principes sont admis de manière conventionnelle. Elle constitue souvent un acte de langage initiateur ou réactif. La routine

[34] « A routine is not an expression or a strategy, but rather an expression which is appropriate to a situation of a certain kind of strategy which is appropriate relative to certain communicative ends » (Coulmas, 1981 : 16).

conversationnelle se subordonne au principe de *préférence*, c'est-à-dire pour un instant T et un type d'interaction bien délimité, le locuteur va opérer un choix préférentiel et quasi réflexe d'une formulation stéréotypée et conventionnelle qui *de facto* écartera les autres formulations concurrentes. Un exemple de ce principe de préférence est l'emploi en rentrant dans l'espace privé de quelqu'un de la routine *Je te dérange ?,* formule à laquelle on aura tendance à répliquer automatiquement par *non pas du tout entre !* si l'on autorise l'entrée[35]. L'usage d'une autre réplique que le type de réponse attendu (permission ou refus d'entrer), *ça va ?* par exemple, entrainera un *bris de rituel* (Dostie 2018) qui aura tendance à questionner le récepteur, montrant ainsi sa forte ritualisation. Ces séquences préfabriquées de nature récurrente et stéréotypée sont prévisibles dans leur emploi. Elles s'inscrivent dans un usage systématique et standardisé de formes stables au sein d'une communauté linguistique sous forme d'un accord tacite passé entre tous les membres d'une même communauté linguistique et supposé partagé de tous[36] (Coulmas 1981 : 4). Il apparait que le niveau de registre de la routine est fortement conditionné par la *nature* et la *proximité* de la relation interpersonnelle. Coulmas remarquera que plus la relation et le contexte d'interaction tendent vers la formalité, plus le locuteur aura tendance à utiliser des formules avec un très fort degré de figement.

Fait intéressant, Coulmas s'est beaucoup inspiré des phraséologismes pragmatiques du japonais pour nourrir sa réflexion sur la notion de routine conversationnelle. Il donnera ainsi l'exemple dans *Prepatterned speech and conversational routine* (Coulmas, 1981 : 80-81) d'un cas fréquent en japonais où lors de la réception d'un présent de la part de l'interlocuteur, le locuteur va instinctivement employer la routine conversationnelle d'excuse[37] *sumimasen* (litt. : 'ce n'est pas fini'). Coulmas explique que l'expression d'excuse dans ce contexte ne vise pas un acte de repentance, mais un remerciement impliquant la gratitude morale du locuteur bénéficiaire et met l'accent sur les efforts ménagés par l'interlocuteur bienfaiteur dans la sélection et l'achat du présent. Il montrera également que certaines routines peuvent avoir plusieurs fonctions qui s'actualisent en contexte ce qui par exemple se différencie des pragmatèmes (Mel'čuk 1995) :

---

[35] Ou par « une seconde ! », « pourrais-tu repasser plus tard ? » en cas de refus.

[36] « […] tacit agreements, which the member of the community presume to be shared by every reasonable co-members » (Coulmas, 1981: 4). Notre traduction : « […] un accord tacite supposé être partagé par tous les membres de la communauté ».

[37] A noter que cet emploi est moins fréquent dans le registre familier, en situation familiale ou avec ses amis.

> « Par exemple, une formule comme *sumimasen* [...] peut être utilisée comme une structure d'abordage pour débuter une conversation, une expression pour attirer l'attention, une prise de congé, [...] et, fait intéressant, pour exprimer sa gratitude » (Coulmas, 1981 : 82)[38].

Cet exemple illustre bien la contrainte sociale à l'origine du déclenchement spontané de l'usage des routines conversationnelles. Nous constatons ici que la routine conversationnelle constitue une catégorie d'expressions spécifiques à un schéma d'interaction donné et répond à un besoin pragmatique et communicationnel spécifique. Le caractère relativement contraint de l'emploi de la routine se remarque notamment lorsqu'elle est omise, ce qui peut engendrer un dysfonctionnement dans l'interaction. Ces expressions sont sélectionnées spontanément par le locuteur du fait de leur capacité à accomplir un rituel d'interaction spécifique. Le sens de l'expression ne sera donc pas propositionnel, mais pragmatique, ou *formulaire* selon la terminologie de Tamba (2000).

4.3.2. Routine d'interaction, Traverso (1996)

Traverso définit la routine d'interaction comme un ensemble d'expressions récurrentes intrinsèquement liées à un cadre de communication stéréotypé et donc hautement prédictible par les membres d'une même communauté linguistique (Traverso 1996, 2016). Traverso montre que le processus de standardisation des routines dans une acception conventionnelle tend à provoquer un phénomène d'érosion sémantique permettant l'apparition de la fonction pragmatique de l'expression (Coulmas 1981 ; Aijmer 1996 ; Traverso 2016). Ainsi, les routines conversationnelles peuvent perdre leur sens d'origine exprimé par l'addition des sens des constituants qui la composent pour acquérir un sens particulier en contexte d'usage. Un exemple de ce basculement sémantique et fonctionnel donné par Traverso (2016) est la perte de la signification votive de *bonjour* ou de la signification religieuse d'*adieu*. Traverso (1996) attribue deux spécificités aux routines. Elles sont *économiques*, car répondant de manière conventionnelle à une situation, elles permettent d'éviter la charge cognitive liée à la génération d'une séquence libre ou le risque de commettre un impair communicatif. Elles sont aussi *intégratives*, car elles constituent des séquences conventionnelles permettant de

---

[38] « For instance, a formula such as sumimasen [...] can be used as general conversation opener, attention getter, leave taking, [...] and, notably, gratitude formula » (Coulmas, 1981 : 82).

s'intégrer à un groupe social et d'en réaffirmer les valeurs. Enfin, elles constituent des structures « toutes-faites » qui facilitent les interactions en raison de leur préfabrication (Traverso 1996 : 41). Il s'agit notamment d'expressions telles que *ce sera tout ?*, *c'est à qui ?* ou *à vos souhaits !*. Afin de délimiter les spécificités des routines, Traverso propose quatre niveaux de description (Traverso 2016 : 55) :

- Le premier niveau est le ***niveau situationnel***. Ce niveau définit le contexte déclencheur de la routine conversationnelle. Par exemple, bousculer une personne déclenche une formule d'excuse.
- Le second niveau est le ***niveau positionnel***. Ce niveau définit la distribution de la routine dans la séquence interactionnelle. Par exemple, souhaiter un anniversaire se retrouvera le plus souvent en début d'interaction.
- Le troisième niveau est le ***niveau organisationnel*** portant sur la structure et l'enchainement de l'échange et définit l'organisation des échanges en paires initiatives et/ou réactives lors de l'emploi des routines. Certaines routines peuvent impliquer un enchainement spécifique (par ex. : *À vos souhaits !* après un éternuement), d'autres n'impliquent à l'inverse pas de réaction et seront seulement initiatives (par ex. le cas de l'excuse en entrant dans un bureau). Le cas des salutations montre également une organisation sous forme de paire (*Bonjour !* échangé entre deux personnes).
- Le quatrième niveau est le ***niveau linguistique***. Ce niveau concerne la prise en compte des propriétés formelles de la routine employée : figement, type de phrase, polylexicalité, etc.

En outre, Traverso (2016 : 56) fait l'hypothèse que l'usage des routines conversationnelles est conditionné par le statut des relations interpersonnelles. En effet, selon la proximité relationnelle entre les participants à l'interaction, elle remarque que le degré de formalité des échanges varie en fonction du degré de proximité et de la nature de l'interaction. Lorsque la proximité relationnelle est forte, l'usage des routines conventionnelles devient facultatif ou peut prendre des formes plus familières ou compactes[39]. Dans le cas d'une proximité formelle, voire hiérarchique, l'emploi des routines devient incontournable. Enfin, Traverso (2007) observe que certaines routines peuvent être regroupées en paradigmes représentatifs

[39] Par exemple *: je vous en prie* (formel) => *de rien* (registre courant voire familier)

d'un type de comportement interactionnel prototypiquement associé à un rôle social. Elle donner notamment l'exemple du rôle du « petit commerçant » auquel on peut associer des routines du type *c'est à qui ?, madame ?, alors et la petite dame qu'est-ce qu'elle voulait ?, et avec ça ?, il vous fallait autre chose ?* (Traverso 2007 : 26). Ces rôles peuvent également s'affirmer au travers de l'intonation, du volume et du débit de parole[40], des actions récurrentes qui lui sont associées ou du comportement social adopté.

Cette approche des routines par Traverso, proche de celle de Coulmas (1981), apporte selon nous une perspective d'analyse intéressante, celle de la structure interactionnelle et des relations interpersonnelles dans lesquelles s'inscrit l'usage des routines conversationnelles. Ces deux perspectives sont peu traitées dans les travaux sur les phraséologismes pragmatiques. L'approche de Traverso se base sur des données issues d'un corpus d'interactions orales authentiques. Cet appui sur des données orales authentiques ne se retrouve pas nécessairement dans les autres approches qui, faute de données orales authentiques, privilégient les dialogues de romans par exemple. À l'instar de Coulmas, le processus de routinisation se définit essentiellement par les critères de conventionnalisation et de réitération. Toutefois, l'approche de Traverso semble se distinguer de notre théorie des PhPex au niveau de la forte spécialisation des routines conversationnelles impliquées dans l'accomplissement d'un rituel d'interaction fortement codifiée. L'ajout de Traverso sur ce dernier point est que les relations interpersonnelles peuvent néanmoins faire varier la contrainte d'emploi. Ces différents aspects interactionnistes (voire socioculturels) sont des perspectives qui nous ont été très utiles lors de l'analyse des résultats de notre étude. En effet, nous avons pu observer lors des entretiens semi-directifs de la phase 3 que ces deux composantes jouaient un rôle important dans la sélection et la compréhension des PhPex.

4.4. Les approches contextuelles

Cette section traite des approches descriptives centrées sur la notion de contexte dans la description du fonctionnement des phraséologismes pragmatiques. Ces différentes approches se sont essentiellement interrogées sur l'interaction entre contexte et usage, et

---

[40] Des travaux sont actuellement menés par Laura Abou-Haidar à propos des saillances prosodiques observables sur des routines discursives associées au rôle de l'enseignant en classe de FLE. Ces travaux (Abou-Haidar 2022) semblent aller dans le sens de l'hypothèse de Traverso (2007) concernant le lien entre routines et prosodie.

notamment sur la manière dont les spécificités du contexte appliquent une contrainte sur l'emploi des PhP.

4.4.1. Énoncés liés, Fónagy (1982) et Marque-Pucheu (2007)

Fónagy (1982) décrit la catégorie des *énoncés liés* comme des actes d'énonciation « directement et globalement liés à une situation » (Fónagy, 1982, p. 4), c'est-à-dire des séquences phraséologiques dont l'emploi stéréotypé est activé par une situation prototypique d'emploi, ou comme le précise Náray-Szabó (2009) : « *Directement*, car on les prononce au moment même où la situation se produit, et *globalement*, car l'énoncé est indécomposable sémantiquement et pragmatiquement »[41] (Náray-Szabó, 2009 : 49). Les énoncés liés constituent des réactions verbales routinières et spontanées face à une situation de communication stéréotypée, ce que précise à nouveau Náray-Szabó (2009) dans la citation suivante :

> Autrement dit, si l'on réagit dans une situation de dialogue face à un évènement courant dans la vie quotidienne, il semble plus rare de vouloir se borner à un simple constat. Il y a bien plus de chances qu'on souhaite exprimer une réaction plus directe, telle qu'une appréciation positive *c'est le pied !* ou négative *ça ne casse pas la baraque !* la confirmation ou l'infirmation de ce qui vient d'être dit *et comment !* ; *tu penses !*, des sentiments *ah bon !* ; *mais je rêve !*, etc. (Náray-Szabó, 2009 : 1)

Les énoncés liés s'emploient dans le cadre d'un dialogue oral dyadique[42]. À l'inverse des énoncés libres considérés comme transparents et à valeur référentielle, les énoncés liés constituent des séquences idiomatiques généralement associées à une dimension expressive et/ou évaluative. Pour illustrer cela, nous pouvons observer l'énoncé lié *tu vas t'en mordre les doigts !* qui exprime à la fois un jugement négatif porté sur la situation et notamment sur ce que s'apprête à faire l'interlocuteur, jugement qui résulte d'une évaluation subjective du comportement de l'interlocuteur.

---

[41] Mise en italique de notre fait

[42] Toutefois, il nous semble que certains emplois considérés comme énoncés liés, et notamment ceux à fonction expressive, par exemple « mais je rêve ! », peuvent être employés dans le cadre d'un soliloque.

Marque-Pucheu (2007) va poursuivre les travaux de Fónagy sous la forme des *énoncés liés à une situation*. Selon elle, l'*énoncé lié* se caractérise avant tout par une forte récurrence de l'usage conventionnalisé au sein d'une communauté linguistique et par un lien fort avec un contexte prototypique d'emploi. Marque-Pucheu évoque par ailleurs la notion d'opacité sémantique pour caractériser l'énoncé lié qu'elle va rapprocher de la compositionnalité : « un énoncé réputé lié à une situation signifie rarement ce qu'indique la suite des mots qui le composent » (Marque-Pucheu, 2007 : 27). Elle proposera les exemples de *Tu parles !* pouvant être utilisé pour quelqu'un n'ayant pas parlé ou de *tu es tombé du lit ?* qui n'interroge pas sur une chute du lit, mais sur le fait que la personne dont il est question s'est levée étonnement tôt. Elle montrera également que tous les énoncés liés ne sont pas non-compositionnels et que parfois, le sens des constituants permet d'accéder à une partie du sens du tout (par exemple les expressions *ne t'inquiète pas*, *pas de problèmes* ou *quelle bonne surprise !*).

En résumé, l'énoncé lié constitue une séquence contrainte par le contexte dans son emploi et son interprétation. Cependant, l'énoncé lié n'est pas strictement spécialisé à ce contexte, d'autres expressions pouvant remplir la fonction communicative attendue. L'énoncé lié se caractérise également par son ancrage fort aux dynamiques de l'interaction orale. Toutefois, Fónagy ne distingue pas les expressions fortement contraintes à une situation ritualisée (Par ex. : *toutes mes condoléances*) de celles qui ont une application plus large et moins contraintes sur le plan fonctionnel (Par ex. : *la belle affaire !*). Enfin, il est utile de mentionner que la classe des énoncés liés admet une large gamme d'énoncés allant de la phrase à l'interjection alors que la sous-classe étudiée des PhPex se spécialise essentiellement aux énoncés phrastiques dotés d'un verbe conjugué.

4.4.2. Énoncés situationnels[43] (*situation bound utterances*), Kecskés (2000)

Cette classe renvoie à des unités idiomatiques à fonction pragmatique fortement conventionnalisées et préfabriquées, dotées d'une visée sociale explicite, et dont l'apparition est liée à des situations de communication stéréotypées (Kecskés 2018 : 219). L'usage des énoncés situationnels contribue à véhiculer l'identité culturelle d'une langue. L'énoncé

[43] Les traductions proposées pour les termes initialement en anglais proposés dans cette section sont les nôtres.

situationnel possède donc une fonction sociale particulière permettant d'intégrer le locuteur à son groupe langagier d'appartenance et de favoriser la reconnaissance mutuelle avec ses pairs, lui évitant ainsi de commettre un bris de rituel ou un impair communicatif. Pour exprimer cet usage conventionnel, Kecskés (2018 : 220) propose l'exemple de l'expression *Welcome aboard !* (*Bienvenu à bord !*) employé par l'équipage d'un avion de ligne pour saluer et souhaiter la bienvenue aux passagers. L'énoncé situationnel appartient au même paradigme que les routines conversationnelles, mais se différencie de par son périmètre fonctionnel plus restreint, en excluant par exemple les marqueurs discursifs. L'usage des énoncés situationnels est déclenché, voire contraint[44] par un contexte en particulier. Ils constituent la réponse conventionnelle la plus pertinente et la plus acceptable dans un contexte de communication stéréotypé et sont des emplois sélectionnés en bloc, employés de manière récurrente et économique sur le plan cognitif. Selon l'approche de Kecskés (2000 : 613), le noyau fonctionnel n'est pas directement encodé dans l'énoncé en lui-même, mais c'est le contexte qui « charge » ce noyau fonctionnel lors du déclenchement de l'usage. Par cela, il entend que l'énoncé situationnel pris en tant que tel n'actualise que le sens de ses constituants. Ce n'est que lorsqu'il est employé en contexte que le noyau fonctionnel est activé.

Sur le plan sémantique, ce sont des énoncés non compositionnels qui, à force d'emploi, acquièrent un sens différent du sens initial pour devenir des unités fonctionnelles pures, dont le sens ne peut s'exprimer que par une fonction issue d'un usage régulier (Kecskés 2018 : 221). Les fonctions pragmatiques de l'énoncé ne sont pas encodées dans les constituants de l'expression, mais s'actualisent uniquement en contexte. Sur le plan sémantique, les énoncés situationnels sont généralement fondés sur des métaphores conceptuelles ou des implicatures conventionnelles propres à une langue-culture. Leur interprétation nécessite d'avoir accès à la connaissance culturelle partagée au sein d'une communauté linguistique. En outre, leur dimension fortement pragmatique implique une classification non pas en fonction de l'entrée lexicale, mais en regard de sa situation prototypique d'emploi. Kecskés (2000 : 611) identifie trois dimensions de compréhension des énoncés situationnels : le *sens situationnel* (*situational meaning*), la *fonction pragmatique* (*pragmatic function*) et le *cadre d'usage* (*frame*). Les énoncés situationnels sont considérés comme des énoncés indexicaux dans la

---

[44]« […] their use is highly predetermined by the situation » (Kecskés 2000 : 606)

mesure où ils évoquent les caractéristiques spécifiques d'un cadre interactionnel prédéterminé auquel ils sont génériquement associés. Pour un même cadre d'usage, il peut exister tout un paradigme fonctionnel synonymique d'énoncés situationnels qui constituent plusieurs alternatives pragmatiques (*pragmatic access route*, Kecskés 2000 : 612) sélectionnées en fonction de critères socioculturels (relation, rôle, distance, thème de la conversation). Pour illustrer cela, nous pouvons prendre l'exemple de l'énoncé lié *you are welcomed !*. Cet exemple peut se décliner en différentes *pragmatic access route,* donnant ainsi le paradigme suivant activable dans le cadre d'usage « se dit après avoir été remercié » :

- *You got it*
- *Don't mention it*
- *No worries*
- *Not a problem / no problem*
- etc…

Le choix de l'énoncé lié le plus adapté se fera en fonction de la norme interactionnelle impliquée par le contexte, du niveau de proximité relationnelle, de l'intention de s'engager dans l'interaction, de la disponibilité cognitive du locuteur ou du niveau de politesse. Soulignons que les recherches de Kecskés s'orientent vers une optique didactique et acquisitionnelle. Il montre que les énoncés situationnels constituent des unités phraséologiques qui représentent souvent une difficulté concrète pour les apprenants. En outre, son approche se situe à la croisée de celle de Coulmas et de Fónagy (Kecskés 2000 : 606) avec lesquels il partage cet ancrage à la fois aux énoncés stéréotypés et contextuels et à la notion de récurrence. Ses travaux font également écho aux travaux de Traverso (2007) sur les aspects interactionnels. Nous distinguons en revanche notre théorie de la sous-classe des PhPex de Kecskés sur cette notion de contrainte forte du contexte. En effet, selon notre approche, le PhPex n'est pas contraint de manière unilatérale par le contexte. En revanche, pour un schéma conversationnel donné, on peut retrouver un paradigme de PhPex susceptible de remplir la fonction communicative attendue.

4.4.3. Phrases figées illocutoires, Náray-Szabó (2009)

Les *phrases figées illocutoires* constituent « une classe d'expressions idiomatiques phrastiques, utilisées en réaction à une situation conversationnelle donnée. » (Náray-Szabó 2009 : 1). Cette catégorie inclut « les phrases simples ou complexes, elliptiques ou non, à valeur illocutoire non déductible à partir de la seule interprétation du sémantisme de leur élément » (Náray-Szabó 2009 : 155). La situation joue un rôle de première importance dans l'interprétation de ces énoncés. Le locuteur va ainsi mobiliser lors du processus interprétatif à la fois ses connaissances de la situation en cours (contexte) et ses connaissances générales et culturelles du schéma interactionnel auquel il est confronté (cotexte). En outre, les *phrases figées illocutoires* constituent des expressions récurrentes socialement stéréotypées et associées conventionnellement à un contexte prototypique d'emploi. Cette stéréotypie culturelle ou *idiomaticité illocutoire* (Náray-Szabó 2006 : 103) semble être à l'origine du caractère intraduisible de certaines des expressions.

En outre, Náray-Szabó (2009) montre que les phrases figées illocutoires sont généralement non-compositionnelles, l'addition des sens des constituants de la séquence ne permettant pas de déterminer la fonction discursive ou la fonction pragmatique de l'expression hors contexte. Cette non-compositionnalité est selon Náray-Szabó (2009 : 155) de niveau pragmatique, dans la mesure où même si l'énoncé semble transparent, il n'en demeure pas moins que sa valeur illocutoire ne peut être déduite de la somme des sens de ses constituants. La structure actancielle des phrases figées illocutoires est la suivante : [X], [Y], [Z] où X = le locuteur, Y = l'interlocuteur et Z = l'élément réactif. Ces différents éléments sont latents, car sous-entendus dans le discours en raison de l'idiomaticité des phrases figées illocutoires (Náray-Szabó 2009 : 158).

Ces travaux se situent dans la droite lignée de ceux de Fónagy (1982) et mettent en avant le caractère circonstancié des phrases figées illocutoires. Il nous a semblé toutefois que cette approche n'apportait pas de changement théorique significatif par rapport à ce qui avait été fait chez Fónagy, Marque-Pucheu ou Kecskés avec qui son approche partage beaucoup de points communs. Se pose alors la question de la différentiation de cet objet linguistique par rapport à la tradition théorique dans laquelle il s'inscrit. D'autre part, tout comme dans le cas de l'approche de Kecskés (2000, 2018), nous distinguons notre approche de celle de Náray-Szabó sur le plan du lien univoque entre le contexte et l'emploi du phraséologisme.

Nous conclurons cet inventaire en mentionnant que malgré l'engouement actuel des chercheurs japonais pour la phraséologie de l'oral, notamment en didactique des langues, il ne semble pas exister, tout du moins à notre connaissance, de travaux portant spécifiquement sur le traitement linguistique et la typologisation les phraséologismes pragmatiques à fonction expressive de la conversation quotidienne (excepté les rituels linguistiques). Il existe quelques études éparses ciblées sur des emplois particuliers, notamment l'étude de Cài (2019) qui observe l'expression linguistique des sentiments de manière très globale, ou l'étude d'Ishida (2005, 2008) qui se penche sur les expressions de la colère en japonais. Partant de ce constat, nous avons jugé ici pertinent d'observer le phénomène des phraséologismes pragmatiques en considérant à la fois ses réalisations en français et en japonais, et ce de manière à préparer d'autres études que nous souhaitons mener à l'issue de ce travail de thèse.

Tout au long de cette revue théorique des différents objets associés aux phraséologismes pragmatiques, nous avons pu constater que ce domaine a donné lieu à un foisonnement de travaux menés dans une grande diversité de perspectives. Cette étape du cadrage théorique nous a permis de faire le point sur les différentes spécificités communes à la classe des phraséologismes des interactions à valeur expressive. Nous avons également pu discuter les différences et les similarités entre les PhPex et d'autres types de phraséologismes pragmatiques. Nous allons dans la partie qui suit définir la catégorie des PhPex qui sera observée dans notre étude de la compétence lexicopragmatique des locuteurs japonophones.

Chapitre 3 - Les phraséologismes pragmatiques à fonction expressive : forme, sens, usage

Ce chapitre est dédié à la définition de notre objet d'étude qui fera l'objet dans la suite de ce travail de recherche d'une étude appliquée à la compétence lexicopragmatique des locuteurs japonophones de niveau avancé en français. Dans un premier temps, nous allons discuter le choix de sa dénomination, puis nous traiterons tour à tour ses différentes spécificités syntaxiques, sémantiques, pragmatiques et interactionnelles. Nous aborderons enfin la liste des PhPex qui seront utilisés dans notre protocole expérimental.

1. Choix de dénomination

Dans la section consacrée à l'extension du domaine de la phraséologie (Chap.1, section 3), nous avions rapidement évoqué le fait que la grande transversalité disciplinaire des études récentes sur les phraséologismes avait donné lieu à un grand foisonnement terminologique, ce qui n'est pas sans avoir engendré une certaine confusion dans la classification de ces différents objets entre eux. Il peut paraitre paradoxal ici que nous souhaitions ajouter une dénomination de plus à cet inventaire déjà dense. Toutefois, c'est justement dans un souci de clarification terminologique que nous avons choisi de parler ici de *phraséologisme pragmatique à fonction expressive* (abrégé *PhPex*).

Afin d'homogénéiser la dénomination de notre objet d'étude avec le champ théorique dont elle relève, nous avons ici choisi comme dénomination le terme de *phraséologisme.* Ce terme rend notamment compte du caractère phrastique, séquentiel et préfabriqué de l'objet étudié. Il s'inscrit par ailleurs dans une tradition terminologique déjà consacrée par les chercheurs du domaine et il fait l'objet d'un consensus fort. À ce noyau dénominatif, nous ajoutons l'adjectif catégorisant *pragmatique* qui rend quant à lui compte de la dimension actionnelle et communicationnelle de notre objet d'étude. Enfin, le dernier terme renvoie à la fonction *expressive* de ces phraséologismes pragmatiques. Cette dénomination a l'avantage de garder explicite l'ancrage notionnel avec le sous-type de lexème et le périmètre fonctionnel dont elle relève.

Parmi les travaux recensés, peu d'entre eux se sont attelés à définir clairement les raisons du choix terminologique impliqué dans la dénomination de leur objet d'étude, et ce en raison

du caractère isolé et minutieux de l'analyse mise en œuvre. Ce n'est qu'au cours de la dernière décennie que, face à un flou terminologique croissant et une forte inflation des travaux de recherche sur des objets présentant de nombreuses similarités, certains auteurs se sont mis à systématiser leur objet d'étude au regard des spécificités des autres objets apparentés au champ des phraséologismes pragmatiques (Kauffer 2017, 2019 ; Dostie et Sikora 2021 ; Krzyżanowska, Grossmann et Kwapisz-Osadnik 2021). D'autres ont proposé des typologies fonctionnelles permettant de hiérarchiser les différents objets étudiés jusqu'à présent en se fondant sur les propriétés formelles, communicationnelles et actionnelles (Lopez Simo 2016, Tutin 2019 ; Dostie 2019). Enfin, certaines propositions d'articulation et d'homogénéisation des concepts ont aussi émergé de travaux lexicographiques d'ampleur comme c'est le cas avec la notion de *pragmatèmes* qui s'inscrit elle-même dans une typologie des phrasèmes élaborée pour le *lexique explicatif et combinatoire* dans le cadre de la *théorie Sens-Texte* (Mel'čuk 2013). La figure 2 ci-dessous permet de situer le niveau de granularité de notre objet d'étude et l'ensemble terminologique dans lequel il s'inscrit :

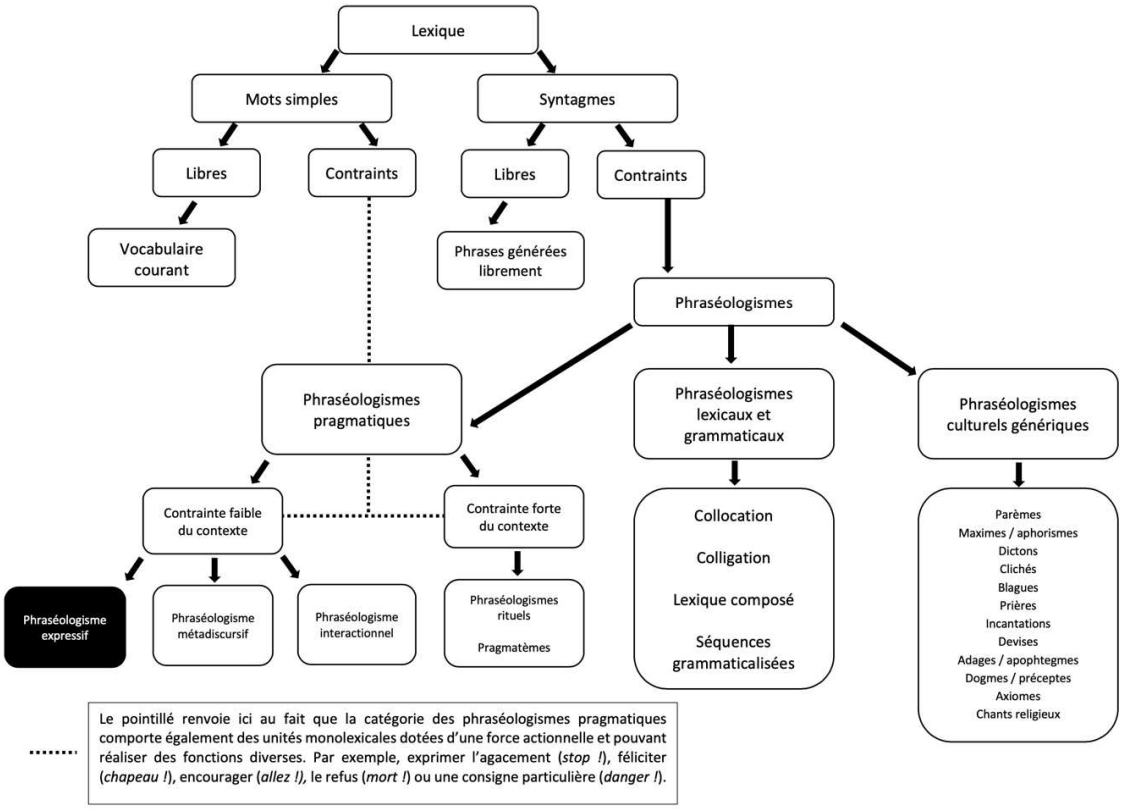


Figure 2 : Typologie simplifiée des phraséologismes[45]

[45] Cette typologie n'est pas exhaustive et ne se veut pas représentative de la diversité des phénomènes associés à la phraséologie.

Le diagramme ci-dessus montre l'organisation de la catégorie des phraséologismes au sein du lexique. Reprenant la définition des phrasèmes de Mel'čuk (1995), la catégorie des phraséologismes renvoie aux séquences polylexicales contraintes sur le plan morphosyntaxique, sémantique et pragmatique. Au sein de cette catégorie, nous distinguons 3 sous-types :

- **Les phraséologismes culturels génériques :** comprend toute séquence polylexicale contrainte sur le plan morphosyntaxique ou sémantique et fortement imprégné du tissu culturel de la langue dont il est question. Ces expressions présentent souvent une iconicité et une généricité forte et sont associées à des activités culturelles diverses et variées. Une dernière caractéristique est la valeur gnomique que peuvent avoir certaines des classes ce sous-type de phraséologismes, c'est-à-dire la propension d'un énoncé à exprimer une vérité ou un enseignement général à valeur morale.

- **Les phraséologismes lexicaux et grammaticaux :** comprend toute séquence polylexicale contrainte sur le plan morphosyntaxique ou sémantique et ayant une fonction grammaticale ou une fonction lexico-référentielle. Cette catégorie constitue avant tout des éléments de la langue qui n'occupent pas de fonction communicative particulière, mais dont le figement sémantique et syntaxique actualise un sens ou une fonction grammaticale qui contribue à améliorer la qualité informationnelle et la fluidité de l'énonciation. Ces séquences résultent souvent d'associations préférentielles de constituants présentant une certaine affinité (ex. *peur bleue, colère noire, fort comme un turc, un vibrant hommage*, etc.) ou des séquences ayant acquis une fonction grammaticale particulière par figement (*d'ores et déjà, petit à petit, à peu près*, etc…).

- **Les phraséologismes pragmatiques :** comprend toute séquence polylexicales contrainte sur le plan morphosyntaxique, sémantique ou pragmatique. Comme précisé plus haut, le phraséologisme pragmatique est intrinsèquement actionnel et interactionnel. Il s'agit avant tout d'énoncés qui visent à agir et faire agir tout en répondant à un besoin communicatif particulier. La composante interpersonnelle et socioculturelle est ici décisive, car les phraséologismes pragmatiques ont cette particularité de pouvoir réguler le comportement des participants au sein de la situation d'énonciation.

Notre objet d'étude, les PhPex, se situe dans la catégorie des *phraséologismes pragmatiques* (pour rappel, abrégé *PhP*) en tant que sous-type dont la fonction première est expressive. Il s'agit notamment de séquences lexicopragmatiques préfabriquées permettant au locuteur d'exprimer son état affectif ou émotionnel, ainsi que son attitude en réaction à un état des choses, un comportement ou une action se réalisant dans la situation d'énonciation. Cette catégorie englobe des expressions particulièrement productives du type *c'est mort ! Tu parles !* ou bien *Tu m'en diras tant !*. Dans ces trois exemples, on remarque assez facilement que le sens global dérivable de l'addition du sens des sous-constituants ne permet pas de rendre compte du sens effectif de l'expression dans son contexte d'emploi. Nous pensons que cette ambiguïté est liée au sens de l'expression qui ne se situe non pas seulement au niveau lexical, mais également au niveau pragmatique et contextuel, niveau dont il est relativement difficile de rendre compte sans une réflexion métapragmatique approfondie.

Comme précisé dans la dernière section du chapitre précédent, nous distinguons notre théorie de la sous-classe des PhPex des *ALS* de Kauffer (2019) et des *formules expressives de la conversation* de Krzyżanowska, Grossmann et Kwapisz-Osadnik (2021). En effet, bien que proches des PhPex au niveau des spécificités, ces classes incluent les unités monolexicales et les structures qui n'ont pas le statut de phrase. Dans le cadre de cette étude, les PhPex seront définies comme des unités phrastiques, dotées d'un verbe à temps conjugué et ayant le statut d'énoncé autonome[46]. En outre, la notion d'acte de langage n'est pas centrale dans notre approche. Nous lui substituons la notion de *schéma conversationnel* qui a le mérite de prendre en compte l'illocutoire visé, mais aussi les différents paramètres extralinguistiques impliqués. Nous aborderons cette notion plus en détail dans ce chapitre. Enfin, notre observation du phénomène des PhPex se centre sur les données orales et médiées (réseaux sociaux notamment) authentiques et non élicitées et prend en compte les variations morphologiques d'usage ainsi que tous les registres de langue, le registre vulgaire inclus. Nous prenons en compte dans la caractérisation des PhPex à la fois les dynamiques interpersonnelles, mais aussi les aspects culturels liés à la perception du sentiment dans la langue-culture et aux modalités d'usages conventionnellement associées à l'expressivité.

---

[46] Il s'agit ici d'un choix méthodologique que nous avons opéré en amont de l'analyse, et ce en raison de la variabilité des structures linguistiques pouvant intégrer la sous-classe des PhPex. Cette définition est vouée à évoluer à mesure que nous approfondiront notre analyse des PhPex et que nous observerons différents types de données orales.

2. Définitions des spécificités des PhPex

Les PhPex se définissent sous quatre niveaux d'analyse que nous allons développer dans les sections qui suivent et qui sont le produit de plusieurs analyses exploratoires faites sur corpus et détaillées dans plusieurs travaux (Ladreyt 2020, 2021, 2022a, 2022b). Cette définition est aussi le fruit de réflexions initiées lors de notre contribution au projet *Pragmalex* Krzyżanowska, Grossmann et Kwapisz-Osadnik (2021).

2.1. Spécificités morphosyntaxiques

Nous définissons les PhPex comme des unités de communication apparentées aux *clausatifs* de Polguère (2011), c'est-à-dire des énoncés phrastiques complets[47] et autonomes sur le plan syntaxique[48] (qui ne sont pas soumis à des phénomènes de rection externe par les éléments linguistiques environnants). En tant que structures faisant partie du stock lexical d'une langue sélectionnées en bloc par le locuteur, les PhPex se définissent comme des séquences soumises à un certain degré de figement qui se manifeste sous la forme d'un degré de contrainte sur l'ordre des unités (Blumenthal 2016) et d'un blocage des propriétés combinatoires et transformationnelles de certains des constituants de l'expression (Gross 1996 : 54). À l'instar de Bolly (2007), de Klein et Lamiroy (2016) et de Kauffer (2017), il nous semble pertinent de décrire le figement des PhPex en parlant de figement à la fois *graduel* et *multiniveau*. Il est graduel dans la mesure où ce figement peut bloquer totalement (*Ça va pas la tête !)* ou partiellement (par exemple le patron [X *rêve*] : *je rêve ! => on rêve ! => on croirait rêver ! => je dois rêver !*) les propriétés distributionnelles ou transformationnelles du PhPex. On pourra dans certains cas parler de *matrices phraséologiques* (Anscombre, 2011)[49] dans le cas de paradigmes de PhPex ayant en commun un même moule, mais ayant des parties variables sur l'axe paradigmatique (par ex. : la matrice [C'est + ADJ / N] => *c'est énorme ! c'est le feu ! c'est mort ! c'est la dèche !*

---

[47] Dotés d'un verbe à temps fini et d'un sujet grammatical.

[48] Ces expressions peuvent toutefois être expansées dans certains cas : « C'est le comble de voir tout ce foutoir ! »

[49] Il serait également intéressant de citer ici les *modèles constructionnels* (Fillmore, Kay, O'connor 1988 ; Goldberg 1995 ; Hoffmann et Trousdale 2013 ; Legallois 2017) pour rendre compte des moules phraséologiques fréquemment observés lors de l'observation des PhP des interactions.

c'est foutu !, cf. Wojciechowska, 2019). Le figement est multiniveau dans la mesure où il n'affecte pas seulement le niveau syntaxique, mais peut également avoir une incidence sur les niveaux sémantique et pragmatique. Ces expressions se définissent également par le principe de *congruence* (Mejri, 2020) qui se caractérise notamment par le fait que les unités qui constituent le PhPex entretiennent une spécificité d'association lexicale ou sémantique qui permet d'actualiser un sens ou une fonction prédéterminée dans un contexte donné. Il est intéressant de noter que les PhPex, en tant qu'expressions permettant de véhiculer une part d'affect, sont souvent employées en modalité exclamative à l'écrit ou avec une intonation montante et expressive à l'oral. C'est d'ailleurs ce dernier trait acoustique qui permet entre autres, dans le cas PhPex ayant plusieurs emplois, de désambigüiser[50] ces différents usages.

2.2. Spécificités sémantiques

Les PhPex se définissent comme des séquences dotées d'une forte idiomaticité[51] (Greciano 1983 ; Hausmann 1997). Ils constituent des expressions fortement ancrées dans la pratique quotidienne du locuteur et sont « le reflet de la conception que les usagers ont du monde » (Gonzalez Rey, 2002 : 68), d'une pratique à la fois itérative et idiosyncrasique. L'idiomaticité qui va de pair avec la fixité et la congruence désigne l'opacité sémantique résultant du processus de fixation de la séquence préfabriquée dans un emploi stéréotypé. Ce processus de fixation donne lieu à l'apparition d'un sens plus complexe étroitement lié aux spécificités socioculturelles de la langue dont il est question. Ces expressions sont en grande partie non-compositionnelles sur le plan sémantique, c'est-à-dire que le sens global de l'expression n'est pas dérivable de l'addition du sens de ses parties[52]. Ce critère est souvent discuté dans les théories observées précédemment et plusieurs chercheurs mentionnent le fait que ce critère demeure instable et peut dans certains cas ne pas être

---

[50] Il nous semble toutefois nécessaire de prendre quelques précautions concernant cette observation. En effet, nous avons pu constater lors d'une discussion avec Agnès Tutin (LIDILEM) concernant ce paramètre que pour une même fonction, il semblerait possible d'observer plusieurs patrons prosodiques possibles. Cette question mériterait à elle seule une recherche plus aboutie afin de déterminer une procédure pour cartographier ces différents patrons prosodiques avec la fonction qui lui ai associée.

[51] « L'idiomaticité n'est pas à la périphérie des faits de langue, elle est au cœur. Elle est l'essence même de la langue. L'idiomaticité des locutions, des collocations, des constructions, cela signifie – au lieu d'une combinabilité libre et illimitée des mots – toutes sortes de figements, de semi-figements, de restrictions, de préférences contextuelles. » (Hausmann 1997 : 284)

[52] Ce critère reste cependant sujet à caution car nous gardons à l'esprit qu'il existe dans la catégorie des PhPex plusieurs expressions dont les constituants permettent d'inférer tout ou partie du sens de l'expression.

pertinent. Nous restons cependant convaincus que dans le cas des PhPex, ce critère demeure utile à la description. En effet, même si l'opacité sémantique (ou la non-compositionnalité) n'est pas totale et qu'une partie des constituants lexicaux permet au locuteur de prédire une partie du sens, il n'en demeure pas moins que dans la majorité des cas, les constituants ne permettent pas d'accéder pleinement à la fonction ni à toutes les nuances d'usage en contexte. En outre, le PhPex se définit par sa structure sémantique synthétique. En plus d'être compacte dans sa forme, sa teneur sémantique est condensée. Elle concentre plusieurs traits qui s'actualisent en fonction du contexte d'emploi et qui convoquent à la fois des informations socioculturelles, mais aussi pragmatiques. Néanmoins, le locuteur non natif qui n'est pas tout le temps en mesure de repérer le caractère préfabriqué et non compositionnel d'une expression ni d'avoir accès aux informations socioculturelles permettant l'accès au sens, va avoir tendance à procéder par analyse pour décomposer le sens du PhPex, ce qui pourra l'induire en erreur[53]. Nous verrons dans la partie de l'analyse dédiée aux entretiens post-test que cette stratégie est récurrente chez les locuteurs non natifs de niveau avancé.

Enfin, les PhPex opposent bien souvent un *sens phraséologique* à un *sens littéral*. La différenciation entre les deux sens se fonde principalement sur l'interprétation du contexte d'emploi et sur la prise en compte des paramètres multimodaux (prosodie, attitude, gestuelle, expressivité du visage) et extralinguistiques (savoir partagé, inférence contextuelle, savoir culturel). Par exemple, l'expression *c'est pas possible* peut exprimer tantôt l'incapacité de réaliser une action, tantôt une réaction émotionnelle forte et négative mêlée à de l'indignation si elle est associée à une prosodie expressive[54]. Cette dernière spécificité constitue parfois une difficulté de compréhension rencontrée par les locuteurs lorsque le contexte ne suffit pas à distinguer le sens dont il est question. Nous tenterons d'observer cette difficulté plus en détail dans l'analyse du test lexicopragmatique (2e phase de notre protocole expérimental).

---

[53] Bally (1909) avait été à nouveau très clairvoyant à ce sujet : « Aussi les unités ne sont pas aussi difficiles à caractériser que les séries ; un étranger est moins tenté de les traduire par leurs éléments ; le sens total s'impose avec plus de force et le groupe se reproduit tel quel à chaque répétition ; si en outre l'unité est d'un usage très fréquent, il est impossible de ne pas reconnaitre que le groupe équivaut à un mot unique. C'est le cas par exemple de locutions adverbiales comme *tout à l'heure*, *tout de suite*, *à peu près*, *tout à fait*, *sans doute*, *sans cesse*, etc. » (Bally 1909 : 75). Bien d'autres après lui ont remarqué cette tendance analytique dans la compréhension des phraséologismes pragmatiques chez les non-natifs, tendance qui constitue aujourd'hui l'une des nombreuses questions de recherche en acquisition de la langue étrangère.

[54] C'est-à-dire une intonation expressive marquée par un débit soutenu et un volume sonore supérieur à la normale.

2.3. Spécificités pragmatiques

Les PhPex se définissent par une fonction intrinsèquement pragmatique. Leur emploi est actionnel et vise à agir sur différents paramètres de la situation de communication ou sur l'interlocuteur. En outre, ils constituent des énoncés fréquemment associés à un contexte d'usage spécifique et conventionnel[55]. Cet ancrage au contexte résulte du processus de routinisation, processus diachronique et itératif au cours duquel un énoncé se fixe dans un usage pour devenir une norme dans une situation de communication récurrente, liant ainsi une forme fixe et conventionnelle à un emploi spécifique et identifiable par les membres d'une même communauté linguistique. Ce faisant, ces expressions sont bien souvent associées à des situations d'interactions socialement stéréotypées et sont très imprégnées du tissu culturel de la langue véhiculée, rendant leur sens et leur emploi d'autant plus complexes[56]. Toutefois, cet ancrage au contexte n'applique pas de contrainte forte d'emploi comme ce serait le cas avec les pragmatèmes (Mel'čuk 1995). Ainsi, la fonction du PhPex s'actualise dans un contexte précis, mais ce même contexte peut activer au niveau cognitif tout un paradigme de PhPex idoine pouvant remplir la fonction communicationnelle attendue avec divers degrés de nuances (par ex. : incrédulité => *j'y crois pas !, mon œil !, tu te fiches de moi !, c'est une blague !, tu plaisantes !, arrête !, tu parles !*, etc..).

Par ailleurs, les PhPex, en tant qu'expressions relevant de la conversation quotidienne, apparaissent préférentiellement dans le registre familier. Cette notion de registre mérite toutefois d'être explicitée, car son maniement est délicat. Il est délicat d'une part parce que comme nous l'observons dans la vie de tous les jours, les frontières entre les différents registres sont poreuses. L'alternance des registres est fréquente à l'oral et s'observe dans une même situation de communication. D'autre part, cette notion peut varier grandement d'une langue à l'autre, notamment en japonais où le registre ne dépend pas uniquement des rapports interpersonnels. Deuxièmement, il est important de préciser que même si une association préférentielle en contexte d'interaction familière est observée chez les PhPex, il n'en

---

[55] Ce contexte est conventionnel dans la mesure où les principes qui le définissent sont connus et admis de tous les membres de la communauté linguistique dont il est question.
[56] Nous avions développé ce lien fort entre dimension actionnelle, usage conventionnel et dimension culturelle dans Ladreyt (2019) concernant les phraséologismes pragmatiques de salutation.

demeure pas moins que ce critère ne contraint pas l'usage pour autant. En effet, il semble possible d'employer des PhPex en contexte plus formel lorsque les spécificités de la situation de communication et la relation interpersonnelle avec l'interlocuteur l'autorisent (ex. : plaisanterie ou étonnement lors d'un speech) si d'une part les constituants lexicaux le permettent (lexique non grossier et n'entrant pas en contradiction avec le décorum du contexte), d'autre part, si l'éthos communicatif de la langue en question admet ce type de spontanéité interlocutive [57]. En outre, notre approche se différencie de celle de Krzyżanowska, Grossmann et Kwapisz-Osadnik (2021 : 27) dans la mesure où nous traitons également le registre *très familier*[58]. En effet, le lexique vulgaire est fréquemment associé aux PhPex du corpus récolté dans la deuxième phase de notre étude. Il nous a semblé que ce dernier niveau de langue avait un haut degré d'expressivité et était tout à fait intéressant pour caractériser les différences et les spécificités entre la variété native et non-native. La notion de registre sera approfondie dans la section méthodologique de la deuxième phase de notre étude (Chapitre 10, section 3.3.2, p. 237).

Les PhPex constituent des réactions *expressives* (Searle, 1982 : 54) dont l'emploi permet au locuteur d'exprimer son état psychologique ou son attitude en réaction à un événement, une situation particulière, un dire ou un comportement se déroulant dans la situation d'énonciation. Cette réaction expressive est étroitement liée au processus d'évaluation et au jugement subjectif. Notre approche de l'expressivité constitue un élargissement de l'approche restreinte de Searle (1982 : 54) dans la mesure où selon nous l'expressivité ne se réduit pas à la dimension émotionnelle ni à des emplois de verbes expressifs, mais renvoie également au positionnement subjectif et explicite du locuteur vis-à-vis de son énonciation ou de la situation de communication. On pourrait ainsi dire que l'expressivité renvoie plus largement à toutes les opérations linguistiques et acoustiques mobilisées dans la mise en valeur du positionnement affectif, psychologique et intentionnel du locuteur. Partant de cela, une expression de joie (*je suis aux anges !)* constituera tout autant un procédé expressif, que

---

[57] Nous exprimons toute notre gratitude à Gaétane Dostie (Prof. Univ. Sherbrooke, CRIFUQ) avec qui nous avons eu une discussion très stimulante lors d'une conférence sur nos travaux que nous avons donnée en mars 2022 dans le cadre des séminaires annuels du CRIFUQ. Les précieux conseils prodigués par Gaétane Dostie lors de cette discussion nous ont permis d'affiner cette notion de registre familier et d'entrevoir tout le potentiel de la prise en compte des variabilités inter-registre.

[58] En cela, nos travaux se différencie de ceux de Krzyżanowska, Grossmann et Kwapisz-Osadnik (2021 : 27) qui ont écarté les emplois vulgaires, notamment en raison de la visée didactique de leurs travaux, ce qui n'est pas le cas de nos travaux ancrés dans la description de l'usage tel quel dans une perspective acquisitionnelle.

le marquage de l'incertitude (*c'est pas gagné !*), ou du désintérêt tinté de grivoiserie (*je m'en tamponne le coquillard !*).

De surcroit, les PhPex n'ont pas spécifiquement une fonction référentielle, mais sont étroitement liés à un schéma conversationnel et à une portée discursive. Ce faisant, ils possèdent la faculté de modifier la nature de la relation qui unit les participants ou d'influencer certains paramètres dans la situation d'interaction. Mentionnons que les PhPex sont fréquemment *polyfonctionnels.* Une même forme peut être employée pour plusieurs fonctions différentes. Un bon exemple de cette polyfonctionnalité est le PhPex *c'est mort !* qui peut en fonction du contexte dans lequel il s'actualise exprimer 4 sens[59] :

- Le refus catégorique : *c'est mort ! j'irai pas à sa soirée !*
- L'incapacité de réaliser une action : *bon bah c'est mort ! le meuble est trop lourd, on va avoir besoin d'aide.*
- Le fait de rentrer dans un lieu peu animé ou l'ennui : *c'est mort ! le dimanche y'a rien à faire à Grenoble*.
- Une situation vouée à péricliter : *c'est mort ! l'équipe en face est trop forte…*

L'usage des PhPex met donc en avant un certain nombre de spécificités pragmatiques liées aux participants, à la situation et au cadre socioculturel régissant l'acte de communication. Il est donc tout à fait légitime de penser que cet usage se retrouve fortement conditionné par la maitrise et la compréhension du cadre de communication, mais également par des spécificités pragmatiques qui constituent le contexte d'énonciation. Ce dernier point sera notamment développé dans l'analyse des résultats des parties dédiées à la phase 2 (Test lexicopragmatique) et 3 (Entretiens métapragmatiques) de notre protocole expérimental.

2.4. Spécificités interactionnelles

À l'instar de Traverso (2007), nous pensons que la prise en compte du niveau interactionnel dans la description des phraséologismes pragmatiques est nécessaire puisque ces derniers

---

[59] Le lecteur pourra trouver un examen plus approfondi des emplois de *c'est mort* ! dans Hinaï et Ladreyt (2022).

s'inscrivent le plus souvent dans le cadre d'une interaction orale. Par *interactionnel*, nous entendons ici tout ce qui concerne l'acte d'interaction entre 2 ou plusieurs interactants, la manière dont ils se distribuent la parole et co-construisent l'acte de communication, mais également la dimension sociocognitive qui intervient dans leur attitude d'interaction et dans le choix des stratégies communicatives employées. Nous traitons plus spécifiquement ici le genre de la *conversation quotidienne*, c'est-à-dire un type d'interaction orale qui se définit principalement par :

- La non-obligation d'un enjeu de communication particulier (discussions banales, triviales)
- La réplication d'un cadre de communication dont les paramètres sont stables
- Un affaiblissement de la distance interpersonnelle
- Un registre de langue généralement familier ou tout du moins autorisant les marques de connivence
- Une place plus grande laissée à l'expression de la subjectivité et de l'affect
- L'effacement momentané des rôles et des places hiérarchiques

Il s'agit donc d'un genre interactionnel qui se caractérise principalement par une forte routinisation et une certaine banalité des enjeux communicationnels visés. Le PhPex est par nature réactif, il constitue une réaction émotionnelle et subjective face à une action, une énonciation ou un événement rencontré dans la situation d'énonciation en cours. Cette réaction est le produit d'une évaluation subjective du déclencheur de la réaction (une personne, un événement, une action, un comportement, etc..) qui n'est pas nécessairement *in praesentia*[60] (dans l'interaction quotidienne médiée par exemple). Ce faisant, l'emploi d'un PhPex ne sera que très rarement initiatif ou conclusif (Bidaud 2002), c'est-à-dire qu'il ne constitue pas à lui seul une ouverture ou une clôture de la conversation. Le critère de réactivité constitue donc une caractéristique centrale des PhPex qui justifie notamment la soudaineté possible de son emploi pouvant être :

- Soit intercalé entre deux tours de parole
- Soit en position de transition du tour de parole vers un autre tour sans changement

[60] Le déclencheur peut être par exemple un événement qui n'a pas lieu directement dans la situation d'énonciation, mais dont on évoque les faits.

de locuteur

- Soit directement en position de chevauchement

De plus, de par son autonomie syntaxique, son sémantisme complexe et sa compacité, le PhPex a la possibilité de constituer à lui seul un tour de parole. Les PhPex peuvent être employés selon plusieurs modalités interactives et s'utilisent autant dans un dialogue ou un polylogue (à 3 personnes ou plus). On observe également que le caractère expressif et spontané des PhPex pourrait permettre leur emploi dans une situation de soliloque où le locuteur réagit émotionnellement sans pour autant qu'un interlocuteur soit présent. Cela semble notamment vrai pour les PhPex n'incluant pas de fonction intersubjective et présentant une forte exclamativité[61] tels que *j'y crois pas* ou *ça me prend la tête !*.

Notons que les PhPex, en tant qu'expressions subjectives et idiomatiques, sont bien souvent le reflet d'un éthos communicatif[62] et d'une coloration culturelle particulière. En ce sens, le matériel linguistique employé pour réaliser le PhPex est sélectionné en fonction de manières de communiquer et d'implicites socioculturels propres à une langue. En outre, ces expressions sont bien souvent associées à des situations d'interactions socialement stéréotypées. Nous désignerons ces situations stéréotypées par le terme de *schéma conversationnel* qui rend compte selon nous de la dimension conventionnelle et prédictible de certaines interactions sans pour autant qu'elles soient contraintes comme le serait des rituels très codifiés tels que l'entretien d'embauche, l'interaction juridique ou l'interaction médicale. Le schéma conversationnel renvoie à des situations stéréotypées de conversation dont certains paramètres semblent rester stables et induisent l'emploi d'un paradigme d'expressions pouvant répondre au besoin communicationnel lié à cette situation de communication (ex. : la dispute ou la plaisanterie entre amis, rassurer quelqu'un, refuser une proposition, etc.). Ces schémas impliquent donc l'emploi de réponses conventionnelles, préconstruites et permettant d'exprimer un sens et une fonction reconnus de tous.

---

[61] Par cela, nous entendons ici une réaction émotionnelle verbale, gestuelle et comportementale présentant un degré d'intensité et de spontanéité supérieur à une réaction émotionnelle banale telle que l'amusement ou l'admiration.

[62] C'est-à-dire un ensemble traits culturels et sociaux propres à la façon de communiquer et associés à une représentation du locuteur idéal d'une langue-culture donnée que le locuteur projette sur sa manière de communiquer et d'interpréter ce qu'il entend, et ce de manière à se conformer aux standards communicatifs institués dans sa langue maternelle.

3. Les PhPex observés dans cette étude

La définition proposée dans la partie précédente nous sert de base lors de l'analyse des résultats qui sera présentée dans la deuxième partie de cette thèse. Les PhPex que nous avons retenus lors de la conception de notre protocole expérimental sont les suivants :

1. C'est le pied !
2. C'est pas possible !
3. Faut pas pousser !
4. C'est abusé !
5. J'y suis pour rien !
6. J'y crois pas !
7. C'est pas vrai !
8. C'est chaud !
9. C'est pas gagné !
10. Tu plaisantes ?
11. c'est génial !
12. c'est le bouquet !
13. c'est une horreur !
14. ça tombe bien !
15. Tu crains !
16. Tu parles !
17. T'inquiète !
18. Tu l'as dit !
19. Tu es sérieux ?
20. Les bras m'en tombent !
21. C'est une honte !

Ces PhPex ont été sélectionnés en fonction de la non-prédictibilité de leur emploi et de manière à exprimer un large panel d'actes illocutoires tels que l'indignation, la satisfaction, la désapprobation ou le doute. Nous avons sélectionné en priorité des PhPex à polarité négative, car il nous a semblé qu'ils étaient les plus propices pour tester la compétence lexicopragmatique des locuteurs japonophones observés dans ce travail de thèse[63]. En outre, nous avons volontairement choisi des PhPex dont l'interprétation est double et permet un emploi phraséologique ou littéral. Nous voulions ici tester l'aptitude des locuteurs non natifs à différencier ce type d'ambiguïté. Par ailleurs, nous voulions observer quel type de stratégies étaient mobilisées lors de l'interprétation, notamment dans le cas des stratégies de traduction d'appui sur la langue maternelle. La plupart des PhPex employés dans notre test ont été validés au niveau de leur productivité et traités dans le cadre du projet *Pragmalex* (Krzyżanowska, Grossmann et Kwapisz-Osadnik 2021).

Au cours de ce premier volet dédié au cadrage théorique des PhPex et du champ disciplinaire dans lequel ils s'inscrivent, nous avons pu faire le point sur les spécificités des PhPex et du contexte scientifique et notionnel dans lequel ils s'inscrivent. Nous allons à présent passer

[63] Une explication plus détaillée de ce choix est proposée dans la partie dédiée au cadre méthodologique.

au deuxième volet de notre cadre théorique concernant l'acquisition des phraséologismes de la conversation quotidienne chez des locuteurs non natifs, en mettant notamment l'accent sur la notion de *compétence lexicopragmatique* qui est centrale dans ce travail de thèse.

Deuxième volet du cadre théorique

-

Des phraséologismes pragmatiques à fonction expressive à l'acquisition de la compétence lexicopragmatique

En guise de transition vers le deuxième volet théorique

Notre examen du champ théorique de la phraséologie et de la notion de *phraséologisme pragmatique* a montré que la compréhension et la description des mécanismes d'usage de ces objets posaient de nombreuses difficultés, et ce en raison du caractère multidimensionnel de leur figement et de la forte idiomaticité de leur emploi. Il n'est donc pas surprenant que l'acquisition du lexique préfabriqué des interactions pose des difficultés au LNN, et ce même si celui-ci présente une expertise élevée en langue cible (désormais LC). Plusieurs études ont ainsi montré que malgré un niveau d'expertise élevé les LNN tendent à connaitre et à utiliser une faible proportion d'expressions préfabriquées comparé aux locuteurs natifs (Blum-Kulka et Olshtain 1986 ; House 1996 ; Wray 2001 ; Bardovi-Harlig 2018). Cette tendance peut s'expliquer pour plusieurs raisons : linguistique (différence de structure entre la LM et la LC, complexité inhérente à la LC), contexte d'acquisition (longueur, qualité et variété d'exposition, expérience de l'usage, opportunité d'interagir) ou individuel (état de développement de la compétence linguistique, facteur émotionnel, facteur identitaire, implication et motivation). Bien que remarquée très tôt chez Bally (1909 : 73), cette difficulté propre aux phraséologismes pragmatiques n'est que peu abordée de manière systématique dans le domaine de la phraséologie, principalement dans le champ récent de la *phraséodidactique* (Gonzalez-Rey 2015, 2019).

Ce deuxième volet de notre cadre théorique a pour objectif d'apporter quelques éléments d'éclairage concernant les différentes notions spécifiques à l'acquisition de la compétence pragmatique en L2 qui seront nécessaires à la compréhension du cadre de notre expérimentation. Cette deuxième phase se découpe en 4 parties : une première partie contextualisera le champ de l'acquisition de la compétence pragmatique. Une seconde partie abordera les approches qui seront utilisées lors de l'observation des données expérimentales et qui s'inscrivent dans le champ de la pragmatique de la langue seconde (*second language pragmatic* ou SLP). La troisième partie évoque la notion de *compétence pragmatique* chez le locuteur non natif (désormais LNN). Dans cette même partie, nous définirons notre approche de la *compétence lexicopragmatique*. Enfin, nous aborderons dans une quatrième partie les dispositifs d'élicitation et d'évaluation de la compétence (lexico)pragmatique. Nous allons procéder dans les pages qui suivent à quelques rappels historiques pour contextualiser les questionnements sur la compétence de réalisation et de compréhension des phénomènes pragmatiques chez les LNN.

Chapitre 4 - Contextualisation du champ de la pragmatique de la langue seconde

Ce chapitre nous permet d'introduire le champ théorique de l'acquisition L2 dans lequel s'inscrit cette étude. Dans un premier temps nous détaillerons le contexte dans lequel s'est développé le champ d'études de la compétence pragmatique en langue seconde. Dans un second temps nous préciserons les trois approches qui sont à la base de notre méthodologie d'analyse des données du protocole expérimental. Enfin, nous tenterons de mettre en évidence le recoupement de ces approches avec le champ de la phraséologie pragmatique des interactions.

1. Tour d'horizon historique et théorique

1.1. L'émergence de l'étude de la compétence pragmatique

L'étude la compétence pragmatique chez le locuteur non natif est un domaine relativement récent. On doit les premiers travaux d'envergure sur le développement de la compétence pragmatique en langue seconde à Thomas (1983) et à Kasper & Dahl (1991). Thomas (1983) propose les notions de *pragmatique transculturelle*[64] et de *compétence pragmatique* pour désigner à la fois la compétence permettant au LNN de mener à bien une tâche communicative et la capacité de compréhension des différents paramètres du contexte (Thomas 1983 : 92). À la suite de Thomas, Kasper et Dahl (1991) introduisent le terme de *pragmatique interlangagière*[65] pour rendre compte de la compétence de compréhension et de production des actes langagiers chez le LNN, ainsi que les interactions possibles avec sa langue maternelle. À partir de ces théories fondatrices, de nombreux travaux vont essaimer, s'axant principalement sur deux questionnements : comment les LNN perçoivent-ils et produisent-ils un contenu linguistique actionnel en contexte ? Comment les LNN développent-ils leur compréhension des phénomènes pragmatiques dans la langue cible ?

---

[64] *Cross-cultural pragmatics* (Thomas 1983)
[65] *Interlanguage pragmatics* (Kasper & Dahl 1991)

1.2. Les années 1980 à 1990 : émergence des études comparatives et translinguistiques

S'inspirant des travaux en pragmatique d'Austin (1970 [1962]) et Searle (1972) sur la théorie des actes de langage, ainsi que des travaux de Brown & Levinson (1987) sur la politesse linguistique, et à la suite des travaux fondateurs de Thomas et Kasper & Dahl, un grand nombre d'études ont émergé, principalement dans le monde anglo-saxon. Ces études partent du principe que chaque langue mobilise des moyens linguistiques différents pour exprimer une fonction pragmatique lors de la communication. Ces travaux se sont ainsi focalisés sur l'acquisition de la compétence pragmatique, mais aussi sur les phénomènes de compréhension et de production chez le LNN au regard des différences entre la langue maternelle du locuteur et la langue cible. Ces études s'orientent principalement sur la prise en compte des paramètres contextuels dans la production et la compréhension des formes linguistiques employées. Nous pourrons citer par exemple l'étude comparative de Blum-Kulka, House et Kasper (1989) sur l'acte de requête ou d'excuse, les travaux de Bardovi-Harlig et Dörnyei (1998) sur des LNN hongrois et italiens ou l'étude de Fukuya, Reeve et Christianson (1998) sur des locuteurs japonais, chinois, coréens et russes. À partir de là, les différentes études proposées vont tenter de faire varier plusieurs paramètres parmi lesquels le niveau d'expertise en LC, la durée d'apprentissage de la LC, la durée et la fréquence d'exposition à la LC, la durée de séjour dans un contexte ou la LC est parlée ou le type de contexte d'interaction en LC. En parallèle, partant du principe que la compétence pragmatique peut être enseignée (Taguchi 2015 ; Culpeper, Mackey, Taguchi 2018) et de fait, évaluée, plusieurs chercheurs se sont intéressés à l'élaboration de dispositifs d'évaluation et notamment aux critères sur les lesquels se fonde cette évaluation de la compétence pragmatique (nous développerons ces protocoles d'évaluation dans la troisième partie de ce cadre théorique).

Il est à noter qu'une majorité des travaux sur la compétence pragmatique se sont principalement orientés vers :

- L'usage des actes de langage (excuse, requête, justification, félicitation, etc.)
- L'usage d'expressions propres à un rituel interactionnel (salutation, présentation de soi, condoléances, etc.)
- L'usage des marqueurs discursifs (marqueurs phatiques, fillers, marqueurs métadiscursifs, etc.)

Ce sont en effet des phénomènes généralement observables dans plusieurs langues et propices à la caractérisation des compétences linguistiques des LNN. Toutefois, à l'exception de la catégorie des marqueurs discursifs, peu de travaux à notre connaissance se sont penchés sur les phraséologismes pragmatiques, et encore moins sur les phraséologismes pragmatiques propres à la conversation quotidienne. Cela s'explique en partie par le fait que les objets linguistiques évoqués précédemment sont relativement stables et cadrés au niveau de leur condition d'emploi. Or, ce n'est pas forcément le cas avec les PhPex pour lesquels on observe plusieurs schémas d'usage pour une même forme, ou à l'inverse, plusieurs formes possibles pour un même contexte. À peu près à la même époque (années 1990), on assiste à l'apparition de travaux portant sur l'évaluation des connaissances pragmatiques en L2 (Hudson et *al.* 1995, Roever 2005) sur le savoir-faire et le savoir-être nécessaires à la sélection de la forme linguistique adéquate en contexte. Durant cette période, l'étude se centre principalement sur le locuteur et sa production.

1.3. À partir des années 1990 : études longitudinales et généralisation des mesures statistiques

À mesure que les études transversales et interlangagières ont gagné en popularité dans le champ de l'acquisition, on observe l'apparition d'approches longitudinales dont la préoccupation première est l'observation du processus d'acquisition dans le temps et dans un contexte donné. Ce type d'étude part du principe que l'acquisition de la compétence pragmatique se situe dans une temporalité longue. L'expérience que le LNN acquiert en contexte, ses tentatives d'utilisation de structures linguistiques particulières ou les hypothèses analytiques formulées en contexte de communication dans la LC sont autant de variables qui ne peuvent être analysées que sur le temps long. Ce type de protocole prévoit plusieurs mesures standardisées de la production à plusieurs étapes de la période d'observation, suivies par des entretiens de routine mis en œuvre pour recueillir des données sociobiographiques. Étant donné la nécessité d'un suivi à la fois sur le temps long et dans un cadre d'exposition stable à la LC en milieu endolingue, une grande majorité de ces études ont été menées sur des étudiants en échange scolaire s'étalant de plusieurs mois à plusieurs années (cf. *study abroad research*). En outre, cette abondance d'études longitudinales a également donné lieu à l'apparition des mesures statistiques inférentielles et corrélatives dans l'analyse de la production des LNN. Ces mesures permettent notamment de vérifier si

les résultats observés et les évolutions de la compétence constatés sur le long terme sont le produit d'une influence des variables contextuelles ou dépendantes du locuteur. Ces approches sont tout à fait intéressantes et ont notamment permis d'observer des phénomènes impossibles à caractériser lors du travail qualitatif d'analyse (ce point sera développé dans la partie statistique de la deuxième phase de notre protocole expérimental).

Par ailleurs, Kasper et Schmidt (1996 : 150) puis Bardovi-Harlig (1999) et Kasper et Rose (2002) constatent que les études menées à cette époque se concentrent essentiellement sur l'usage, mais peu sur le processus d'acquisition de la compétence pragmatique en lui-même. Ils pointent ainsi le manque de données extensives sur le processus de construction de la compétence pragmatique mobilisée dans l'usage de productions analysées. La revue des travaux longitudinaux proposés par Taguchi (2010) montre également une forte augmentation de ces travaux dans les 15 dernières années, en lien avec le perfectionnement des outils de recueil et de traitement des données. Ces études ont permis de réexaminer certains mythes qui ont la vie dure en SLP, en montrant par exemple que la simple exposition ne constituait pas nécessairement un paramètre favorisant l'acquisition de la compétence pragmatique, mais que l'observation répétée, le guidage en cours de langue ou par des LN et l'apprentissage par essai/erreur contribuaient fortement au développement pérenne de la compétence pragmatique sur le temps long.

1.4. Des années 2000 à aujourd'hui : apport des théories de la SLA à la SLP

La SLP, longtemps jugée trop descriptive et centrée sur les réalisations linguistiques plutôt que sur les processus sous-jacents de cette réalisation (Kasper et Rose 2002 :13) se fédère de plus en plus dans les années 2000 autour des travaux réalisés en acquisition de la langue seconde, parmi lesquels l'*hypothèse du repérage conscient* (*noticing hypothesis,* Schmidt 2001) ou la *théorie de la socialisation langagière* (Cook 2008). Les travaux actuels tendent à définir de manière consensuelle la compétence pragmatique comme une compétence construite au contact de l'expérience sociale et de la communication en contexte, et non pas émergeant du locuteur en lui-même. Ce revirement théorique a donné lieu à une plus grande prise en compte des facteurs contextuels et de socialisation dans le développement de la compétence pragmatique. Le LNN est perçu comme un acteur social, en contexte de communication situé et motivé par un besoin communicationnel particulier. L'accent est

également mis sur les processus d'interprétation et de compréhension des formes pragmatiques.

En outre, la forte internationalisation des échanges et des flux de savoir que connaissent les années 2000 a là aussi joué un rôle important dans l'évolution des pratiques d'acquisition et sur les objectifs des travaux portant sur le développement de la compétence pragmatique. Avec l'apparition de notions telles que le multilinguisme et le multiculturalisme, le LNN n'est non plus seulement perçu comme un acteur social, mais également comme un vecteur interculturel et interlangagier. Son expérience de locuteur dans la langue cible est le résultat de divers types d'interaction (tension, coopération, interférence, etc.) entre la langue cible et les langues de son répertoire. Le locuteur apprend de manière autonome et s'expose par la consommation culturelle à une grande diversité de phénomènes linguistiques dans diverses langues. Cette extension a notamment donné lieu à l'apparition d'approches telles que la *compétence pragmatique multilingue* [66] (Alcón-Soler 2013) ou la *pragmatique interculturelle*[67] (Kecskes 2014).

Toutefois, comme le remarque Taguchi (2015) dans sa méta-analyse de 58 études sur l'enseignement et l'évaluation de la compétence pragmatique, en dépit de cette forte mondialisation des flux langagiers, on observe la forte prépondérance des études menées sur l'acquisition de l'anglais (38 études sur 58). Elle ajoute qu'en comparaison, les études menées sur le japonais, l'allemand, le français, le chinois ou l'espagnol langue cible sont relativement moins représentées. Cette centration sur l'anglais du fait de son statut de *lingua franca* ou de langue utilitaire tend à réduire le champ de cette approche interculturelle et interlangagière qui gagnerait pourtant à faire varier les LC étudiées. Une autre constatation de Taguchi (2015) est que la majorité des travaux se concentrent sur des cas de collégiens, lycéens ou étudiants de niveau débutant à intermédiaire (notamment en raison de la grande proximité avec le chercheur et de la commodité de rassembler des locuteurs dans le cadre de leur cursus scolaire), mais plus rarement sur le cas des LNN plus âgés, hors circuit académique et de niveau avancé. Il semble donc exister un enjeu majeur quant à la diversification des données au sein des études sur la compétence pragmatique. Après avoir évoqué le contexte épistémique général dans lequel ont été menées les études sur la

---

[66] *Multilingual pragmatic competence* (Alcón-Soler 2013)
[67] *Intercultural pragmatics* (Kecskes 2014)

compétence pragmatique, nous allons aborder le courant théorique qui constitue le terreau dans lequel se sont développées les approches que nous allons développer par la suite, la *pragmatique de la langue seconde*.

2. Le champ de la pragmatique de la langue seconde

Dans cette section, nous allons préciser le champ disciplinaire dans lequel s'inscrit le volet acquisition de ce travail de recherche, à savoir la *pragmatique de la langue seconde*. Nous aborderons notamment ses fondements théoriques et les objectifs de cette approche.

2.1. Définition

La pragmatique de la langue seconde est un champ d'étude qui regroupe à la fois les théories de l'acquisition de la langue seconde et de la pragmatique. Cette approche part du principe que le langage est de nature indéterminée et ambigu. Les unités qui constituent l'énoncé du locuteur ne suffisent pas à comprendre de manière satisfaisante le contenu exprimé. En d'autres termes, la connaissance seule du sens des éléments linguistiques employés ne suffit pas pour interpréter une énonciation ou produire un énoncé approprié. Le contexte et les différents paramètres qui le constituent sont aussi déterminants pour la compréhension et la production. Le concept est introduit pour la première fois par Kasper et Dahl (1991 : 216) comme l'exploration de la compréhension et de la production des actes langagiers et des connaissances pragmatiques chez les LNN en phase d'acquisition de la langue cible. Il s'agit donc d'un domaine visant « l'étude du développement et de l'utilisation des stratégies mobilisées par le locuteur non natif dans la poursuite d'actions linguistiques[68] » (Kasper et Schmidt 1996 : 150, notre traduction). La SLP étudie de quelle manière le LNN associe une forme linguistique donnée à un emploi répondant à un besoin communicationnel spécifique (Bardovi-Harlig 2010). En d'autres termes, la SLP se donne pour objectif principal l'étude de la manière dont les LNN comprennent comment et quand utiliser une expression donnée, dans un cadre donné, avec un type de locuteur et un objectif donné (Bardovi-Harlig 2013 : 68[69]). La SLP cherche également à comprendre de quelle manière les LNN s'approprient et

---

[68] « study of the development and use of strategies for linguistic action, by non-native speakers » (Kasper et Schmidt 1996 : 150).
[69] « how learners come to know how-to-say-what-to-whom-when » (Bardovi-Harlig 2013 : 68)

co-construisent le sens et leur capacité d'inférence en contexte de communication (Culpeper, Mackey, Taguchi 2018 : 1). Il s'agit donc avant tout de comprendre les mécanismes linguistiques et acquisitionnels qui sous-tendent l'usage de constructions linguistiques actionnelles par un LNN engagé dans un contexte d'interaction sociale.

2.2. Quels sont les objectifs de ce champ disciplinaire ?

Ce champ disciplinaire vise l'étude de la capacité du LNN à mettre en œuvre des fonctions communicatives dans une situation de communication donnée, tout en utilisant ses connaissances sociopragmatiques de manière à sélectionner une forme linguistiquement appropriée à l'activité sociale dans laquelle il est engagé. Les compétences socioculturelle et stratégique sont donc ici des composantes cruciales de la compétence pragmatique. De même, l'attitude du LNN envers la langue cible, sa motivation à l'acquisition et le contexte d'appropriation constituent tout autant d'éléments qui entrent en compte dans l'étude de la compétence pragmatique. L'un des enjeux de cette dernière décennie a été de développer des méthodes expérimentales empiriques et rigoureuses pour permettre à la fois un traitement statistique et outillé des données récoltées. Toutefois, la nature même de la compétence pragmatique rend la tâche difficile, car certains paramètres entrant en compte dans son fonctionnement sont invisibles et relèvent des processus cognitifs du LNN, phénomènes très complexes à observer. Un autre enjeu est de comprendre dans quelle mesure les langues du répertoire linguistique du LNN interviennent dans le développement de la compétence pragmatique, et le cas échéant, de déterminer quels sont les différents types de contacts entre ces différentes langues. Enfin, la SLP vise à développer des moyens d'évaluer le degré d'expertise de la compétence du locuteur en compréhension et en production, mais aussi sa capacité à adapter sa stratégie communicative en fonction du contexte de communication rencontré. Ces travaux visent sur le plus long terme à établir des critères formels et rigoureux permettant de déterminer les spécificités de la production des LNN et non pas seulement leur adéquation. Dans le chapitre qui suit, nous allons aborder les 3 méthodologies d'analyse qui ont fondé notre recherche.

Chapitre 5 - Une approche interculturelle, interlinguistique et contrastive

Après avoir situé le champ disciplinaire de notre étude, nous allons ici détailler les trois approches qui ont fondé la méthodologie employée dans l'analyse des données issues de notre protocole expérimental. Ces trois approches complémentaires concernent chacune une des perspectives qui seront traitées en filigrane dans notre analyse : la zone de contact entre la LC et les différentes langues du répertoire linguistique du LNN (*Pragmatique interlangagière*), la zone de contact entre les normes contextuelles et conventionnelles qui sous-tendent l'emploi des PhP dans la langue cible et dans les différentes langues du répertoire linguistique du LNN (*Pragmatique transculturelle*) et l'effet des différences typologiques entre la LM et la LC sur l'acquisition de la compétence pragmatique (*Analyse contrastive de l'interlangue*).

1. La pragmatique transculturelle (Cross-Cultural Pragmatics)

Approche initiée par les travaux de Kasper et Blum-Kulka (1993), la pragmatique transculturelle (parfois désignée par *Transcultural Pragmatics,* Taguchi et Roever, 2017) vise à comparer l'emploi d'objets linguistiques à fonction pragmatique chez plusieurs locuteurs de langues et de contextes d'apprentissage et sociaux différents. Cette approche se fonde sur la constatation que les réalisations linguistiques et non linguistiques mobilisées pour exprimer un acte langagier, ainsi que les conventions et les structurations sociales qui encadrent ces réalisations sont très variables selon les cultures. C'est cette variabilité qui constitue le centre d'attention de cette approche, en essayant d'une part de déterminer les différences entre les réalisations pragmatiques d'une même fonction communicative dans la LM et la LC. D'autre part, l'approche permet d'observer comment ces différences peuvent intervenir dans le développement de la compétence pragmatique du LNN. En outre, la pragmatique transculturelle observe les phénomènes pragmatiques à l'aune des présupposés sociaux, des valeurs et des conventions véhiculées par certaines pratiques linguistiques au sein d'une communauté de locuteurs donnée. Ce système socioculturel tacite, à la fois en langue source et en langue cible, est ainsi pris en considération dans l'analyse. Cette approche implique la comparaison systématique d'un ou plusieurs groupes tests (LNN de la langue cible) et un groupe contrôle (LN de la langue cible).

La pragmatique transculturelle s'intéresse également à la notion d'*échec pragmatique* (*pragmatic failure,* Thomas 1981 ; Blum-Kulka et Olshtain 1986), intervenant généralement lorsque deux interlocuteurs échouent à comprendre l'intention et l'acte illocutoire véhiculés par l'un d'entre eux, ou lorsque la fonction pragmatique attendue n'est pas réalisée ou exprimée de manière adéquate. Cet intérêt vient notamment de l'observation que ce type d'échec[70] peut intervenir fréquemment lors de la conversation exolingue. Cette différence entre les LN et les LNN a pour origine un contact entre les habitus en langue maternelle et l'acte de langage à exprimer en langue cible. En effet, même s'ils possèdent une connaissance suffisante de l'emploi pragmatique, le LNN n'a pas nécessairement accès aux informations socioculturelles ni aux restrictions qui peuvent peser sur l'emploi. C'est notamment ce dernier cas qui peut favoriser l'apparition d'usages pragmatiques pouvant mener à un échec communicatif (Bardovi-Harlig 2001, Kasper 2001, Kasper et Roever 2005).

Le travail d'analyse propre à l'approche de la pragmatique transculturelle implique l'intégration de la méthodologie de l'analyse contrastive qui peut concerner deux langues et plus. Cette approche est donc particulièrement adaptée au travail de comparaison que nous souhaitons effectuer dans notre travail de recherche, c'est-à-dire entre différentes langues (le français et le japonais) et différents profils de locuteur (LN et LNN). Enfin, il parait utile de mentionner que cette approche s'intéresse également à l'impact des variations au niveau interculturel, mais aussi au niveau intra-culturel dans l'acquisition des formes pragmatiques. En effet, un groupe de locuteurs natifs peut très bien pratiquer une variation de la norme attendue, variabilité à laquelle le LNN n'est bien souvent pas préparé (Taguchi 2018). Pour illustrer cette variation intra-culturelle, nous pouvons prendre le cas du français. Pour exprimer par exemple l'acte de salutation, on peut observer une variation des expressions linguistiques employées à la fois au niveau diatopique (variations régionales) et au niveau diastratique (en fonction du groupe social). La norme n'est jamais homogène pour tous les locuteurs d'une même communauté. Cette variation peut être source de difficulté pour le LNN lors de l'interprétation.

---

[70] Le terme *échec* est ici pris non pas dans son sens négatif, mais dans son sens fonctionnel, c'est-à-dire comme l'absence de réalisation ou la réalisation partielle de la fonction attendue. Le même constat est fait pour l'échec au niveau grammatical où nous ne parlerons pas de *faute*, terme connoté négativement, mais d'*erreur*, terme qui a l'avantage de mettre l'accent sur la dimension systémique du phénomène observé.

2. La pragmatique interlangagière (*Interlanguage Pragmatics*)

La pragmatique interlangagière (ou ILP) (Kasper 1981 ; Kasper et Blum-Kulka 1993) constitue un domaine transdisciplinaire à l'interface de la pragmatique linguistique et de l'étude l'acquisition en langue seconde (Felix-Brasdefer, 2013). Cette approche explore le développement de la compétence pragmatique en relation avec le développement de la compétence générale en L2, traitant ainsi les relations entre expertise en L2 et compétence liée à la réalisation d'un acte de langage. Elle part du principe que la compétence du LNN est transitoire, c'est-à-dire qu'elle constitue une *interlangue* (Selinker 1972) à mi-chemin entre sa langue maternelle, les langues de son répertoire et la langue cible. Ce système transitoire qui est en perpétuelle évolution entre phases d'instabilité et phases de fixation est un lieu de contact entre différentes normes linguistiques, interactionnelles et culturelles tout à fait éclairant pour l'étude de la production du LNN. L'étude de l'interlangue est donc très intéressante pour le chercheur dont l'objectif est de comprendre l'influence translinguistique que peuvent avoir les langues du répertoire du LNN entre elles, mais également pour comprendre comment les différences systémiques entre ces différentes langues peuvent causer des difficultés d'acquisition, ou au contraire, être facilitatrices. Une deuxième dimension prise en compte par cette approche est l'évolution de la compétence dans le temps. L'approche vise ainsi à analyser la production du LNN et à évaluer à plusieurs étapes de progression la compétence pragmatique en langue cible. Cette prise en compte extensive permet d'observer le processus de passage d'une maitrise inadéquate à un usage adapté d'une fonction pragmatique donnée, et les différents éléments qui ont permis cette progression.

La pragmatique interlangagière s'intéresse donc aux caractéristiques pragmatiques du développement de l'interlangue, c'est-à-dire le développement de la compétence impliquée dans la réalisation des fonctions pragmatique en lange cible. Elle étudie la manière dont les LNN comprennent et produisent les actes de langage, ainsi que la façon dont ils acquièrent les connaissances pragmatiques nécessaires à l'interprétation et l'usage de phénomènes pragmatiques particuliers (Kasper et Dahl, 1991 : 216). Dans les années qui suivent l'apparition de cette approche, le champ d'observation s'est élargi avec une ouverture aux stratégies communicatives et à la gestion des paramètres interactionnels et interpersonnels (Kasper et Schmidt 1996 : 150).

Plus récemment, Kasper et Rose (2002), puis Bardovi-Harlig (2010) élargissent à nouveau le champ de l'ILP en la définissant comme une approche qui prend en compte de manière globale la relation entre le système de la langue et l'usage, et ce de manière simultanée et complémentaire. La fonction pragmatique ne peut s'étudier sans la forme qui l'exprime et vice versa. L'ancrage entre fonction et réalisation linguistique est donc fort et constitue la matière empirique de leur analyse. L'analyse de la production et de la compréhension du LNN se fonde sur des réalisations concrètes qui permettent d'étudier les interfaces culturelles et linguistiques potentielles entre la LM et la L2 mobilisés lors de l'usage en langue cible. C'est notamment en raison de cet ancrage fort entre description de la forme, structure de l'interlangue du LNN et spécificités systémiques de la situation d'interaction que nous avons appuyé en partie notre analyse sur cette approche.

3. Analyse contrastive de l'interlangue (*Contrastive Interlanguage Analysis*)

L'*analyse contrastive de l'interlangue* (désormais ACI) a été développée par Granger (1996), impulsée par l'émergence des recherches sur corpus d'apprenants au début des années 1990. Cette approche propose un cadre de comparaison entre deux ou plusieurs langues dans l'exploration de la compétence linguistique du LNN. L'avantage ici est de pouvoir déceler des différences ou similarités qui auraient été invisibles si chaque langue était traitée séparément. C'est cette vision d'ensemble et cette transversalité qui fait de cette approche une approche particulièrement adaptée à l'analyse contrastive de la production de LNN. En outre, l'approche propose un cadre de comparaison à deux dimensions : une première dimension qui concerne la comparaison entre la variété native et la variété non native, et une seconde dimension qui concerne la comparaison entre différentes productions de LNN possédant des langues maternelles différentes, mais partageant la même LC. L'objectif principal est de faire contraster la production de LN et de LNN dans une situation comparable de production (Granger 2015 : 9). Bien que cette approche se prête particulièrement à l'analyse de la compétence écrite, le potentiel théorique pour l'étude des expressions mobilisées à l'oral semble possible à la condition que des données de production organisées en corpus soient produites de manière comparable entre le corpus test et le corpus contrôle. Ce faisant, le corpus d'apprenants est décisif dans la mise en place de l'étude contrastive et doit être construit de manière cohérente avec le phénomène observé.

L'ACI constitue une approche qui s'est longtemps focalisée sur le lexique préfabriqué à l'écrit, notamment sur les colligations, les collocations, les mots composés et les locutions verbales. Ce type d'analyse a notamment été possible à l'aide d'une approche rigoureuse et empirique fondée sur l'utilisation systématique des données de corpus. Cette approche se fonde également sur des outils d'exploration très performants permettant de modéliser des phénomènes complexes autrement invisibles lors du travail d'analyse qualitative. De plus, en réponse à l'émergence des préoccupations actuelles des chercheurs en didactique et en linguistique concernant la production à l'oral, on assiste à l'émergence de corpus multimodaux (vidéo et son) de locuteurs non natifs visant l'étude de phénomènes pragmatiques propres à l'interaction orale (Granger 2015 : 10). L'un des objectifs de la démarche est d'observer des sous ou des sur-représentations d'emplois liées soit à la non-maitrise de la structure attendue, soit à des différences entre la langue cible et la langue maternelle.

L'ACI vise à observer de manière empirique les spécificités de la variété non native au regard des données des LN de manière à identifier des caractéristiques particulières telles que des points de contact entre les langues du LNN et la langue cible, ou des difficultés récurrentes pouvant être systématisées. Enfin, contrairement aux deux autres modèles théoriques présentés, l'ACI possède un ancrage fort à l'enseignement. L'un des buts ultimes de ces approches comparatives et de produire des éléments factuels permettant à l'enseignant d'intervenir plus facilement en classe de langue, ou de développer des outils d'apprentissage proche des besoins des LNN. Voici un schéma synthétisant l'articulation entre les différentes approches méthodologiques évoquées précédemment :

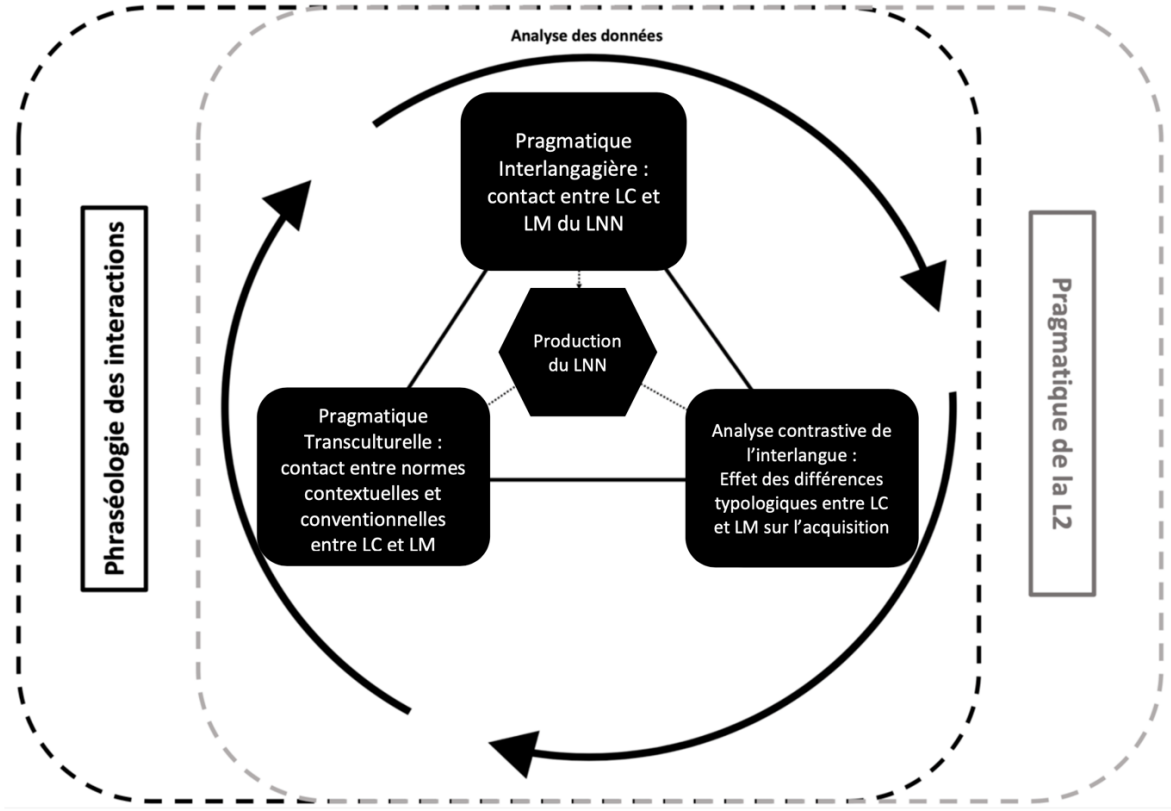


Figure 3 : Articulation des trois approches d'analyse de la production des LNN et des deux champs disciplinaires mobilisés dans notre travail de thèse

4. À propos des travaux sur l'acquisition des phraséologismes pragmatiques de la conversation

Les questionnements sur l'acquisition des phraséologismes pragmatiques de la conversation (ou *routines pragmatiques* selon les approches) sont très récents et n'ont donné lieu à des travaux qu'au cours des 15 dernières années (Forsberg 2010 ; Kecskes 2012 / Bardovi 2012, 2019 ; Roever 2012 ; Edmonds 2013 ; Yang 2016 ; Osuka 2017 ; Barron 2019). Ces études se concentrent essentiellement les expressions à fort caractère rituel ou fortement conventionnalisées. L'une des raisons de cet attrait modéré pour ce champ de recherche réside notamment dans le fait que la pragmatique est un sous-domaine de la linguistique encore jeune et que l'étude de la langue parlée fondée sur les corpus oraux n'en est qu'à ses premiers développements. En outre, la compréhension et l'usage des phraséologismes pragmatiques sont le fruit de la combinaison de plusieurs compétences convoquées

simultanément. La compréhension ne semble pas directement corrélée au degré d'expertise en langue cible, mais c'est en revanche l'expérience répétée en contexte qui participe grandement au processus d'acquisition (Roever 2012). Mentionnons que toutes les routines pragmatiques ne sont pas du même niveau de complexité, et ce faisant les formules ritualisées très utilisées comme les salutations ou les excuses auront tendance à être plus facilement acquises que les structures un peu plus opaques et moins contraintes par le contexte telles que les PhPex. De même, la forte récurrence, la relative stabilité des phénomènes ritualisés et l'appairage quasi systématique du contexte d'emploi avec un ensemble restreint de formes conventionnelles dont ils font l'objet sont l'une des raisons pour lesquelles un grand nombre de travaux se sont focalisés sur les formules rituelles.

Bardovi-Harlig dont les travaux ont grandement contribué à développer le domaine de l'acquisition des phraséologismes pragmatiques a abordé le sujet dans une longue série d'études récentes sur les *expressions conventionnelles*[71] (Bardovi-Harlig 2009, 2010, 2012, 2014, 2018a, 2018b, 2019, 2021, 2022, notre traduction). Bardovi-Harlig définit les expressions conventionnelles comme des séquences récurrentes et stéréotypées d'au moins deux mots, dont l'emploi est préférentiel chez les locuteurs natifs (exclusion de toute forme concurrente en sens) et dont la fonction illocutoire est associée à un contexte prototypique. Ces expressions sont fortement récurrentes et constituent un enjeu d'acquisition important pour le LNN qui évolue en immersion dans la communauté de la langue cible. Une des observations récurrentes dans ses travaux concerne le fait que l'association d'une forme pragmatique avec ses conditions d'emplois et le cadre socioculturel dans lequel elle s'inscrit n'est jamais un processus linéaire. Il nécessite à la fois une exposition extensive et la réunion de plusieurs conditions telles que l'immersion dans la langue cible, la variation du type d'exposition, l'étayage métapragmatique et l'expérience de l'usage et de l'écoute. Le LNN doit être en mesure de déceler et de comprendre les spécificités contextuelles et socioculturelles de l'acte de langage attendu dans le contexte de communication où il se trouve avant de procéder à la sélection de la forme la plus adaptée. Bardovi-Harlig (2019) observe que la maitrise des expressions conventionnelles implique le développement de la compétence pragmatique, mais également le développement d'une connaissance des spécificités grammaticale et morphosyntaxique des formes employées. Elle a ainsi remarqué

---

[71] *Conventional expressions* (Bardovi-Harlig 2009)

une corrélation entre le niveau d'expertise en langue cible et le niveau de développement de la compétence pragmatique.

Bardovi-Harlig (2019 : 50), mentionne qu'il existe actuellement peu de travaux sur l'acquisition des formules préfabriquées ayant une fonction pragmatique dans une perspective linguistique et acquisitionnelle, et les langues traitées sont également peu variées : anglais (Bardovi-Harlig 2009, Bardovi-Harlig et Bastos 2011, Bardovi-Harlig et *al.* 2017), chinois (Taguchi, Li et Tang 2017, Yang 2016, Bardovi-Harlig et Su 2018), français (Forsberg Lundell et Erman 2012, Edmonds 2014), russe (Furniss, 2016), japonais (Tateyama 2001) et allemand (Barron 2003). Concernant le traitement des expressions réactives à fonction expressive, il n'existe à notre connaissance aucune étude approfondie à la fois en compréhension et en production. Seuls quelques auteurs mentionnent de manière très parcellaire l'existence de ce type d'expressions.

Nous concluons ici la première phase de ce cadre théorique concernant la contextualisation du domaine de la pragmatique de la langue seconde et des différents cadres théoriques qui ont nourri notre approche méthodologique. Nous allons dans cette deuxième phase du cadre théorique discuter la notion de *compétence pragmatique* qui est cruciale pour cette étude. À la suite de cet examen des différentes approches de la compétence pragmatique, nous proposerons une définition de la *compétence lexicopragmatique*.

Chapitre 6 - De la compétence communicative à la compétence lexicopragmatique

Le but de ce chapitre est double. D'une part, il s'agit de proposer un aperçu épistémologique du socle théorique sur lequel repose notre approche de la *compétence lexicopragmatique*. D'autre part, il s'agit de proposer une définition rigoureuse de cette approche qui servira de base à notre analyse dans la deuxième partie de cette thèse. En effet, il nous apparait essentiel, pour mieux comprendre le domaine d'étude de la compétence pragmatique, de comprendre les difficultés de théorisation des différents niveaux impliqués dans l'acquisition dont ce domaine a fait l'objet.

1. Évolution de la définition de la compétence pragmatique

Dans cette section, nous allons aborder l'évolution des théories de la compétence pragmatique en lien avec la compétence de communication dont elle découle. Nous considérons la compétence de communication comme dépassant le niveau linguistique et incluant des paramètres comportementaux, culturels et un savoir-faire spécifique au mode de communication choisi. Nous partons du principe que la compétence pragmatique est une sous-compétence centrée sur la production du matériel linguistique et des fonctions communicatives. Nous rappellerons l'articulation entre ces deux compétences afin d'éviter toute confusion.

Souvent opposée à la compétence grammaticale qui renvoie à la maitrise des aspects formels de la langue cible (morphosyntaxe, phonologie, lexique), la compétence pragmatique constitue un objet relativement complexe à définir dont la modélisation a beaucoup évolué au cours des 3 dernières décennies. La compétence pragmatique sera pour la première fois théorisée par Dell Hymes sous la notion de *compétence communicative* (Hymes 1972 : 281-282). Hymes définit cette compétence comme la capacité du locuteur à mobiliser le matériel linguistique approprié en fonction de l'intention qu'il souhaite transmettre. Cette compétence se divise en trois sous-compétences complémentaires : a) une compétence renvoyant à la connaissance linguistique et grammaticale, b) une connaissance socioculturelle de la langue cible et c) une connaissance permettant d'associer un énoncé à une fonction communicative tout en respectant le principe de cohérence discursive :

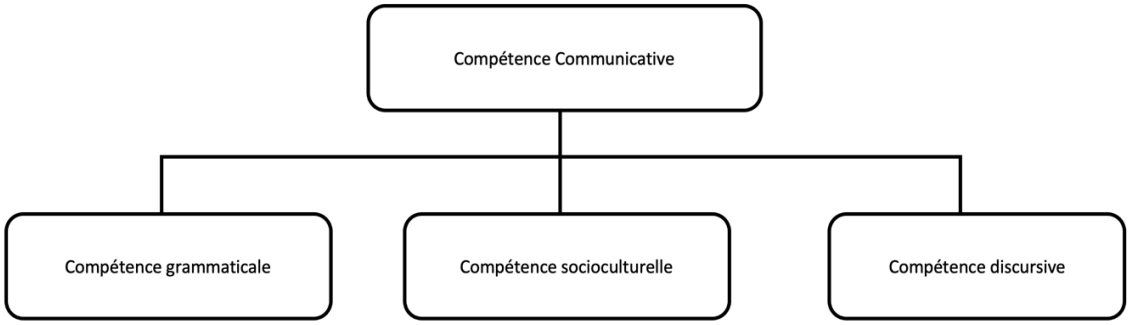


Figure 4 : Le modèle de la compétence communicative (Hymes 1972)

Canale et Swain (1980) puis Canale (1983) vont s'appuyer sur le modèle de Hymes afin de proposer un modèle permettant d'observer le développement de la compétence impliquée dans la réalisation d'un acte de langage chez le LNN. Pour réaliser un acte communicatif réussi, le LNN doit en principe mobiliser 4 sous-compétences : la compétence grammaticale (structure formelle du système linguistique), socioculturelle (connaissance sociale communicative), discursive (cohérence et cohésion du texte) et stratégique (savoir-faire permettant de prévenir les impairs communicationnels) :

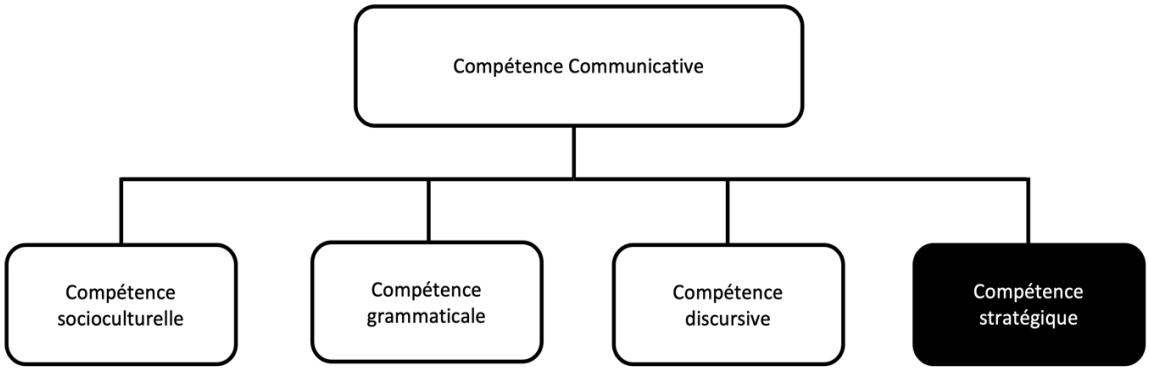


Figure 5 : Le modèle de la compétence communicative (Canale et Swain 1980)

À peu près à la même époque Leech (1983 : 10) va proposer un modèle de la pragmatique fondé sur les travaux de Thomas (1981, 1983). Ce modèle met en évidence la complémentarité de deux niveaux : le niveau linguistique (les ressources linguistiques mises à disposition pour réaliser des fonctions communicatives) et sociolinguistique (connaissance du contexte et de ses contraintes sociopragmatiques) :

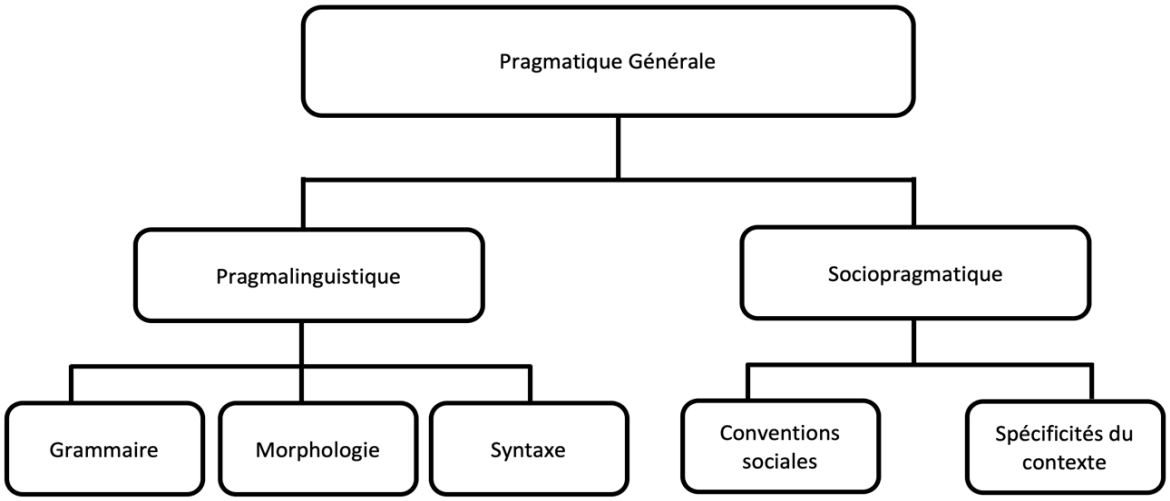


Figure 6 : Les différents champs de la pragmatique selon Leech (1983) et Thomas (1981)

La *pragmatique générale* renvoie aux principes généraux et universels qui organisent un acte de parole. Le champ de la *pragmalinguistique* relève de l'étude du matériel linguistique mobilisé dans la réalisation d'un acte illocutoire, celui de la *sociopragmatique* relève du contexte de communication et des paramètres socioculturels qui cadrent la réalisation de l'acte illocutoire. Ces trois dimensions sont complémentaires et sont à l'origine de la notion de *compétence pragmalinguistique* (capacité à mobiliser les ressources linguistiques nécessaires à l'accomplissement d'une fonction pragmatique en interaction) et de la *compétence sociopragmatique* (comprendre les spécificités du contexte social et sélectionner la forme la plus appropriée en fonction des critères sociaux impliqués) qui seront par la suite des concepts clés dans le domaine de la SLP et dans l'approche de la pragmatique interlangagière (Bardovi-Harlig 1999 ; Kasper and Rose 2002 : 180). En effet, ce modèle sera notamment mobilisé pour décrire la production de LNN et identifier des raisons possibles à des échecs pragmatiques au niveau de l'association forme-fonction (échec pragmalinguistique) et des échecs sociolinguistiques au niveau de l'ajustement de la forme linguistique aux spécificités du contexte et des interlocuteurs (échec sociopragmatique).

Bachman (1990) puis Bachman et Palmer (1996) développent à leur tour le modèle de la *compétence linguistique communicative*[72] (notre traduction) où la compétence pragmatique est autonome par rapport à la compétence grammaticale et préside aux choix communicatifs.

---

[72] *Communicative language ability*, Bachman (1990)

Dans le modèle de Bachman (1990), la compétence pragmatique implique à la fois une compréhension des mécanismes illocutoires liés à un emploi particulier et une compréhension de niveau sociolinguistique codifiant ce qui est approprié et ce qui ne l'est pas. Bachman et Palmer (1996) décomposent la compétence pragmatique en deux types de connaissances : les connaissances fonctionnelles (interprétation des liens forme linguistique/fonction communicative) et les connaissances socioculturelles (adéquation de l'emploi aux caractéristiques sociales de la situation d'interaction) :

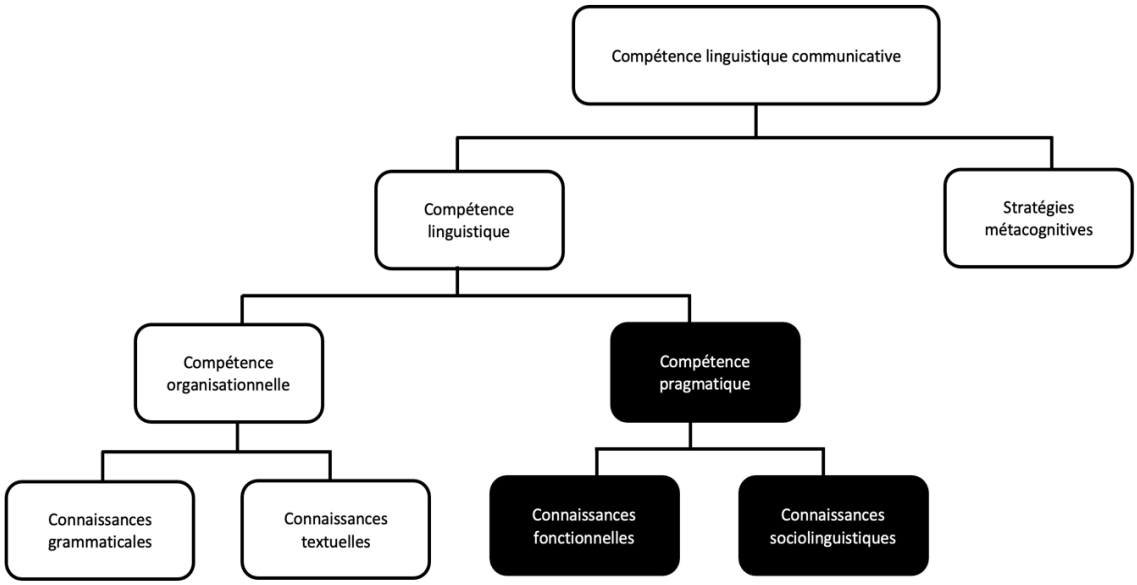


Figure 7 : Le modèle de la compétence linguistique communicative (Bachman et Palmer 1996)

Dans le modèle de Bachman et Palmer (1996), la compétence pragmatique est avant tout perçue comme une connaissance permettant au locuteur de relier une forme linguistique avec sa fonction et son contexte d'emploi. Les critiques principales adressées à ces modèles concernent premièrement la pertinence de la séparation entre compétence discursive et compétence socioculturelle. En effet, l'élaboration du discours semble dépendre également des paramètres socioculturels et s'ajuste en fonction de la personne à qui on s'adresse et du contexte dans lequel se déroule l'énonciation. Par ailleurs, le modèle de la compétence communicative ne détaille pas assez les différents facteurs intervenant dans les différents niveaux de la compétence. Ce faisant, on reproche à ces modèles d'être difficilement opérationnalisables pour mesurer la compétence pragmatique, car ils nécessitent la clarification des critères pour les rendre plus explicites et mesurables de manière empirique. Enfin, ces modèles ont tendance à superposer la connaissance des mécanismes pragmatiques

et linguistiques à la compétence d'usage, mais n'articulent pas explicitement les deux niveaux.

Celce-Murcia et al. (1995) proposent quant à eux un modèle dynamique de la compétence communicative mettant en avant la complémentarité des différentes sous-compétences qui interviennent dans la production du LNN :

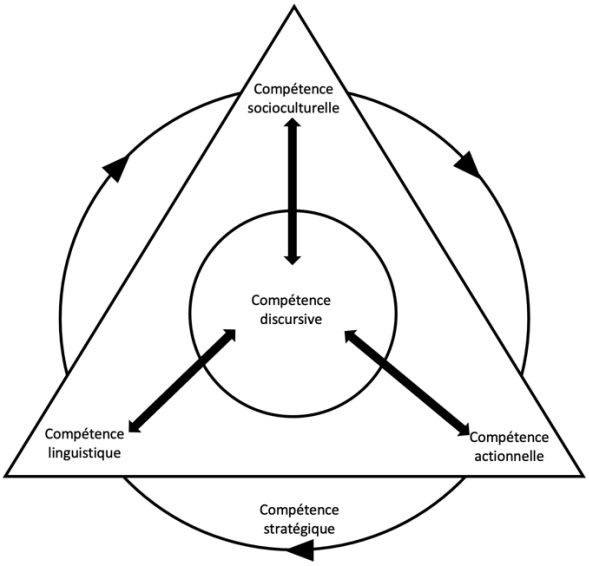


Figure 8 : Le modèle de la compétence communicative (Celce -Murcia et al. 1995)

Celce-Murcia *et al.* (1995) procèdent à une extension du modèle de Canale et Swain (1980). La compétence linguistique ici ne se limite pas à la grammaire et inclut la composante phonologique et lexicale. En outre, les auteurs ont choisi dans ce modèle le terme de *compétence socioculturelle* plutôt que *sociolinguistique* pour mettre en avant que cette compétence implique également une connaissance qui dépasse le niveau linguistique et qui touche aux normes conventionnelles et aux représentations culturelles (Celce-Murcia *et al.* 1995 : 11). Un ajout se trouve au niveau de la *compétence actionnelle* qui constitue la compétence mobilisée pour transmettre et comprendre les intentions communicatives liées à la fonction illocutoire de la forme pragmatique employée (Celce-Murcia *et al.* 1995 : 17-18). La compétence stratégique transite aux trois pôles de la compétence qui eux-mêmes sont en synergie avec la compétence discursive qui concerne l'organisation, la cohérence et la cohésion de l'énonciation. Cette compétence stratégique renvoie à la compétence de planification, d'évaluation et de choix de l'objectif de communication au sein de l'interaction (Celce-Murcia *et al.* 1995 : 26). Ce modèle a le mérite de modéliser la compétence

communicative de manière articulée et complémentaire contrairement aux autres modèles qui semblent différencier les compétences entre-elles.

Cet intérêt pour la compétence communicative ne s'observe pas seulement du côté des experts américains, mais également au sein d'initiatives européennes dont la plus aboutie est sans nul doute le Cadre européen Commun de Référence pour les Langues (CECRL, Conseil de l'Europe 2001). Le cadre CECRL attribue trois niveaux à la compétence langagière : le *niveau linguistique* (morphosyntaxe, phonétique, phonologie, grammaire), le niveau *sociolinguistique* (paramètres socioculturels, registres, connaissances culturelles, conventions, politesse) et le niveau *pragmatique* (fonction communicative de la langue, repérages de situations prédéfinies de communication, acte de parole). Pour chaque dimension, le CECRL définit un ensemble de critères d'évaluation divisés par niveau d'expertise (A1[introductif] à C2 [maitrise]) :

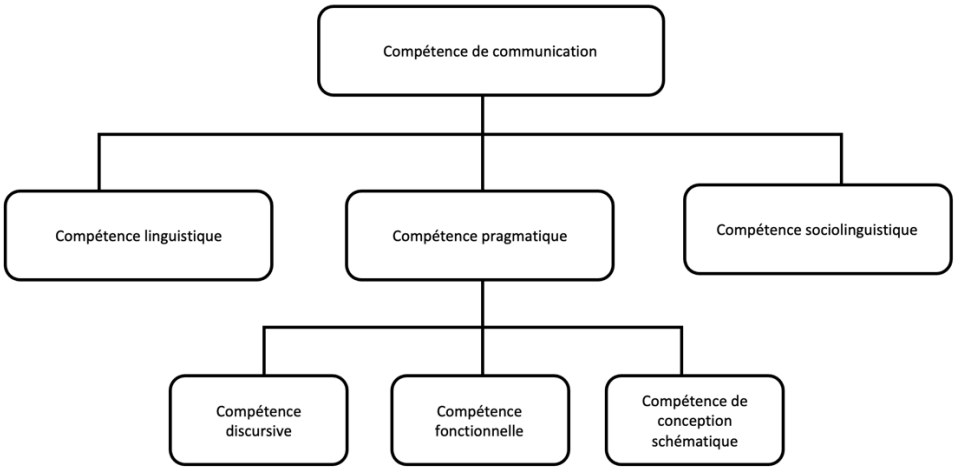


Figure 9 : Le modèle de la compétence de communication selon le CECRL (Conseil de l'Europe 2001)

La compétence pragmatique se subdivise en trois sous-niveaux : la *compétence discursive*, la *compétence fonctionnelle* et la *compétence de conception schématique* (CECRL 2001 : 96). La compétence discursive est définie comme la capacité du locuteur à organiser, structurer et adapter son propos (pertinence, cohésion et cohérence). La compétence fonctionnelle correspond à la capacité du locuteur à comprendre et mettre en œuvre des micro (par ex. : requêtes) et des macro-fonctions communicatives (par ex. : démonstration, explication). Enfin, la compétence de conception schématique renvoie à la capacité du locuteur à utiliser des schémas interactionnels et transactionnels prototypiques du type accord/désaccord ou acceptation/refus et qui structurent l'enchainement du dialogue.

Bardière (2016) montre suite à une analyse critique du CECRL concernant le traitement de la compétence pragmatique que le cadre demeure encore large et ne traite pas assez précisément de la compétence sociopragmatique ou des différents facteurs individuels d'acquisition de la compétence. Les compétences sociolinguistiques, pragmatiques et linguistiques sont traitées de manières distinctes alors qu'elles interagissent lors de l'acte de communication. Lorsque le locuteur choisit une forme pragmatique, il va en principe convoquer au même moment que la compétence linguistique (matérialité et grammaticalité du dire), la compétence pragmatique (stratégie, fonction illocutoire, et construction de l'interaction) et la compétence sociolinguistique (spécificité du contexte, niveau de politesse, registre, conventions sociales impliquées).

Suite aux critiques adressées à son modèle de communication, le CECRL propose un volume complémentaire en 2018 qui vient préciser certaines notions, dont celle de la compétence pragmatique. Il est notamment proposé plusieurs grilles critériées comportant divers descripteurs qui sont appliqués à différents types d'interaction (Volume complémentaire du CECRL 2018 : 86-94). Ces descripteurs sont des ajouts qui permettent de saisir plus précisément certaines microcompétences telles que la *souplesse* (capacité à adapter les acquis langagiers à des situations nouvelles et à exprimer ses pensées de différentes manières) ou la *précision* (capacité à déterminer comment formuler ce que l'on a envie de dire en modulant le niveau de détails en fonction du niveau de détail voulu). Enfin, quelques précisions sont données concernant l'articulation entre les différentes compétences, par exemple entre la compétence fonctionnelle et sociolinguistique.

Avec l'arrivée des approches interactionnistes dans le champ de la pragmatique de la langue seconde courant des années 2000, on assiste à un intérêt croissant pour l'étude des phénomènes interactionnels tels que la co-construction de la séquence interactionnelle chez les LNN. Se fondant sur les premières esquisses de la *compétence interactionnelles*, Celce-Murcia (2007) va revisiter son modèle de la compétence communicative pour intégrer cette dimension :

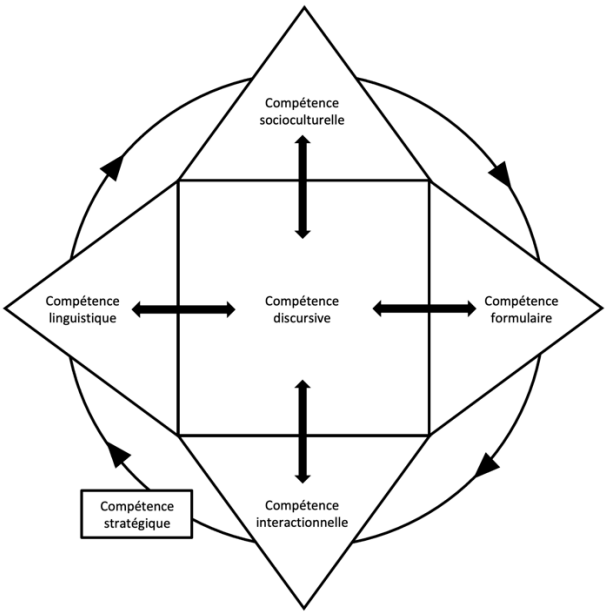


Figure 10 : Révision du modèle de la compétence communicative de Celce-Murcia et al. (1995). (Celce-Murcia 2007)

Cette révision du modèle précédent implémente deux changements majeurs : le remplacement de la compétence actionnelle par la *compétence formulaire* et l'intégration de la *compétence interactionnelle*. Selon Celce-Murcia (2007), la compétence formulaire constitue la connaissance du fonctionnement illocutoire et de l'emploi du lexique préfabriqué utilisé de manière très fréquente dans la conversation quotidienne (routines, collocations, idiomes, *lexical frames* => matrices phraséologiques). La compétence interactionnelle correspond à deux dimensions : la *dimension paraverbale* avec la compréhension des intonations spécifiques, de la gestuelle, la proxémie ou des mimiques spécifiques, et la *dimension structurelle* avec la connaissance des règles d'organisation des tours de parole, des différents types d'interruption, du don ou de la prise de parole, des stratégies de co-construction, du changement de sujet de conversation, entre autres.

À la suite de ces travaux, l'attention des chercheurs en acquisition et en didactique s'est centrée plus spécifiquement sur la notion de compétence pragmatique plutôt que sur la notion de compétence communicative. Cette dernière est alors perçue comme la capacité du locuteur à organiser son discours et les éléments linguistiques qui le composent en fonction de la visée communicationnelle, de l'intention recherchée par le locuteur et de la structure de l'interaction entre les participants engagés dans la conversation. Cette compétence implique la prise en compte de la relation causale entre situation de communication et adaptation du contenu linguistique impliqué par le calcul interprétatif de l'expression le plus approprié au contexte (Taguchi 2009 : 2, Bardière 2016 : 10). L'association

forme/fonction/contexte n'est plus perçue comme un objet stable et préétabli, mais comme le résultat de nombreux ajustements dont le locuteur doit faire preuve pour adapter sa production au contexte d'énonciation (cf. *Flexibilité,* CECRL [2001] 2018). C'est notamment cette flexibilité et cette capacité d'adaptation fondée sur l'analyse des paramètres du contexte et des participants qui dès lors apparaissent être le signe d'une compétence pragmatique développée.

On assiste à la même époque à l'apparition d'études cognitives de la compétence pragmatique, dont celle d'Infantidou (2011, 2014) qui nous a paru tout à fait intéressante du fait qu'elle se centre essentiellement sur la compétence pragmatique sollicitée dans l'usage des structures préfabriquées. Infantidou (2011) part du principe que toute compétence comporte un niveau métacognitif de conscientisation (*awareness*) qui constitue la connaissance nécessaire à la mise en œuvre de cette compétence. Ses travaux posent notamment la question des procédures cognitives indispensables à la mise en œuvre de la communication et des actes illocutoires. Infantidou (2014) établit un lien entre expertise en langue cible, compétence grammaticale et acquisition de la compétence pragmatique. En effet, plus l'expertise est élevée, plus il semble que le LNN multipliera les opportunités de s'exposer à des emplois pragmatiques complexes. De même, plus la compétence grammaticale est élevée, plus le LNN aura de facilité à décoder ou à construire les formes appropriées à son objectif communicatif. Ces trois dimensions perçues comme complémentaires ne se développent pas nécessairement au même niveau au même moment, mais contribuent chacune à l'appropriation de l'autre. L'expertise linguistique dans la langue cible apparait donc comme un facteur facilitateur du développement de la conscientisation des mécanismes impliqués dans les usages pragmatiques. Infantidou (2014 : 130) distingue 3 niveaux de conscientisation dans la compétence pragmatique :

- La capacité d'identifier les instances linguistiques pertinentes : **Conscientisation linguistique** (*Linguistic Awareness*) ;
- La capacité à identifier les effets pragmatiques pertinents : **Conscientisation pragmatique** (*Pragmatic Awarness*) ;
- La capacité à expliquer les relations existantes entre instances linguistiques et effets pragmatiques : **Conscientisation métapragmatique** (*Metapragmatic Awareness*) ;

Cette tripartition permet d'observer les processus métacognitifs qui sous-tendent la compétence pragmatique de manière empirique en se basant sur des critères mesurables. La production du LNN est ainsi évaluée dans ces trois dimensions et permet de mesurer plus précisément l'état de développement de sa compétence pragmatique. La conscientisation pragmatique se centre essentiellement sur le point de vue du locuteur et sur sa production tandis que la conscientisation métapragmatique s'intéresse plus particulièrement à la compréhension et donc à l'interlocuteur qui reçoit le message. La conscientisation métapragmatique constitue un ajout par rapport aux théories précédemment évoquées. Ce niveau de conscientisation intervient notamment à un niveau fin d'analyse des mécanismes d'usage enchevêtrant à la fois des variables linguistiques, contextuelles, discursives et socioculturelles.

L'objectif de l'approche d'Infantidou est de comprendre comment se construisent les processus métacognitifs chez l'apprenant et sur quels éléments se fondent son jugement interprétatif ou la sélection de la forme appropriée, voire son jugement d'acceptabilité de telle ou telle forme linguistique. L'objectif est également de comprendre comment le LNN développe la capacité à tirer parti de la structure et des informations du dialogue dans son intégralité et non pas de manière ponctuelle et ciblée lorsqu'il produit ou interprète un énoncé à fonction pragmatique. Un des reproches que l'on pourrait formuler à l'égard cette approche est qu'elle a principalement été employée pour traiter les genres textuels en production et en compréhension à l'écrit, alors que peu de tentatives ont été réalisées sur l'oral. En revanche, l'intérêt réside indéniablement dans le fait que cette approche propose un cadre rigoureux qui permet de rendre compte du lien entre production et stratégies de communication.

2. Vers une nouvelle approche de la compétence pragmatique ?

Les travaux récents tendent vers une approche discursive et flexible de la compétence pragmatique. L'hypothèse mise en avant est que la relation étroite entre forme et facteurs contextuels n'est pas univoque. L'adéquation parfaite entre contexte et comportement langagier est perçue comme modulaire en fonction de la visée discursive du locuteur, de son intention communicative envers l'interlocuteur. Cette relation forme-contexte est adaptative et peut varier au sein d'une même situation de communication, notamment pour exprimer

des changements de distance interpersonnelle, de registre, d'attitude ou pour exprimer l'affect. Taguchi et Roever (2017 :20) mentionnent par exemple le cas d'une interaction entre deux locuteurs japonais se rencontrant pour la première fois lors d'une fête. Le premier contact et le début de la conversation débuteront bien souvent par l'emploi de formes de politesse linguistique pour marquer la distance interpersonnelle liée à la non-connaissance de l'autre. Toutefois, à mesure que la conversation se poursuit et que des liens de connivence (points communs, intérêts partagés) se créent, il n'est pas impossible que les locuteurs basculent sur des formes non marquées par la politesse (ou alternent entre politesse et neutralité) pour marquer ces liens de connivence et un effacement de la distance interpersonnelle. Au moment de se quitter, ces deux interlocuteurs pourront revenir à une modalité de politesse pour ajouter une dimension rituelle et conventionnelle appropriée à la prise de congé. Cette prise en compte dynamique entre forme, fonction et attitude communicative est intéressante puisqu'elle permet de sortir du cadre un peu trop holiste des observations générales faites sur des productions isolées de leur contexte d'énonciation. De même, le fait d'employer une forme différente de celle qui est habituellement attendue dans un contexte donné ne constitue pas nécessairement un échec pragmatique, mais peut bel et bien être le résultat d'une stratégie communicative volontaire.

Un autre facteur récemment étudié est l'influence de la subjectivité du LNN dans le choix de la posture communicative et du choix des formes linguistiques qui en découle. Même si le LNN dispose d'une connaissance de l'association forme/contexte approfondie, il peut tout à fait ne pas être à même de pouvoir utiliser cette même forme de manière effective, et ce pour des raisons subjectives. Ce non-usage peut par exemple marquer le souhait d'exprimer une certaine résistance volontaire à la norme partagée au sein du groupe de communication. En outre, ce paramètre subjectif peut être à l'origine de la peur ou de l'inconfort ressenti par le LNN à l'idée d'utiliser des formes entrant en contradiction avec sa subjectivité ou son identité. Cette subjectivité des LNN repose notamment sur leur croyance, leur éthos communicatif et leurs standards culturels. La subjectivité et l'identité représentent des facteurs très importants dans l'analyse de la compétence pragmatique des LNN, notamment dans le cadre de l'usage des PhPex qui constituent des expressions subjectives dépendantes de « façons de communiquer » propres à la langue cible qui peuvent entrer en contradiction

avec la langue maternelle. La volonté du LNN d'exprimer un éthos communicatif[73] particulier est là aussi un facteur déterminant dans le choix des formes produites et acquises. Un LNN qui souhaite donner une image formelle de soi aura des propensions fortes à écarter les formes familières pour se focaliser sur des formes plus polies ou standards. Ce paramètre sera observé dans la troisième phase de notre étude dédiée à la discussion métapragmatique où nous verrons par exemple que certains locuteurs refuseront volontairement d'employer des tournures vulgaires en raison du conflit avec leur manière de communiquer dans la LM.

Ces extensions récentes de la définition de la compétence pragmatique montrent à quel point la notion est complexe et scalaire. Les trois dimensions de la compétence pragmatique, à savoir la connaissance linguistique (pragmalinguistique), la connaissance socioculturelle des conventions et des normes d'interaction (sociopragmatique), et la connaissance du fonctionnement des phénomènes pragmatiques et de leur articulation avec les paramètres du contexte (métapragmatique), agissent de manière complémentaire, mais sont également influencées par la visée discursive et la subjectivité du LNN. Les travaux les plus récents semblent s'accorder sur une définition tripartite de la compétence pragmatique (Taguchi 2019 : 4) :

- Des connaissances linguistiques et socioculturelles permettant de choisir la forme adéquate en fonction du contexte de communication

- Des compétences interactionnelles flexibles permettant de s'adapter au caractère variable de l'interaction et de son organisation

- Une composante identitaire renvoyant à l'agentivité personnelle permettant au LNN de choisir en accord avec ses représentations de locuteur s'il souhaite se conformer ou non aux normes pragmatiques de la communauté linguistique dans laquelle il évolue.

À ces trois composantes s'ajoutent des enjeux interculturels et d'intercompréhension liés au contexte socioculturel d'interaction et aux contacts entre les langues des répertoires des locuteurs.

L'une des difficultés actuelles observable dans les tentatives de définition de la compétence pragmatique est que ce qui constitue une production pragmatique acceptable et habituelle

---

[73] Cette notion sera approfondie plus en détail dans la partie analyse des données.

n’est jamais très clairement défini. Les travaux récents se sont focalisés sur les échecs pragmatiques, mais les modélisations fines de ce qui peut être défini comme le prototype du comportement pragmatique idéal tendent à être rares (Perkins, 2007 : 11). Il apparait ainsi que pour mieux décrire les difficultés éprouvées par les LNN et évaluer leur production de manière rigoureuse, il faut avant tout avoir une modélisation précise et empirique de ce qui peut constituer la norme, ainsi que des variations possibles de cette norme qui conditionne l’acceptabilité de production. Il semble donc y avoir un enjeu de collaboration entre la linguistique descriptive (analyse de la conversation, linguistique interactionniste, phraséologie) et les spécialistes de l’acquisition pour d’un côté décrire finement les phénomènes pragmatiques chez LN et proposer ainsi des modélisations du schéma habituel d’emploi d’usage d’un phénomène pragmatique ; et de l’autre, observer comment le LNN s’approprie ces phénomènes pragmatiques à la fois en production et en compréhension. Ci-dessous, un schéma récapitulatif de l’évolution de la compétence communicative et de la compétence pragmatique à travers les différentes approches mentionnées :

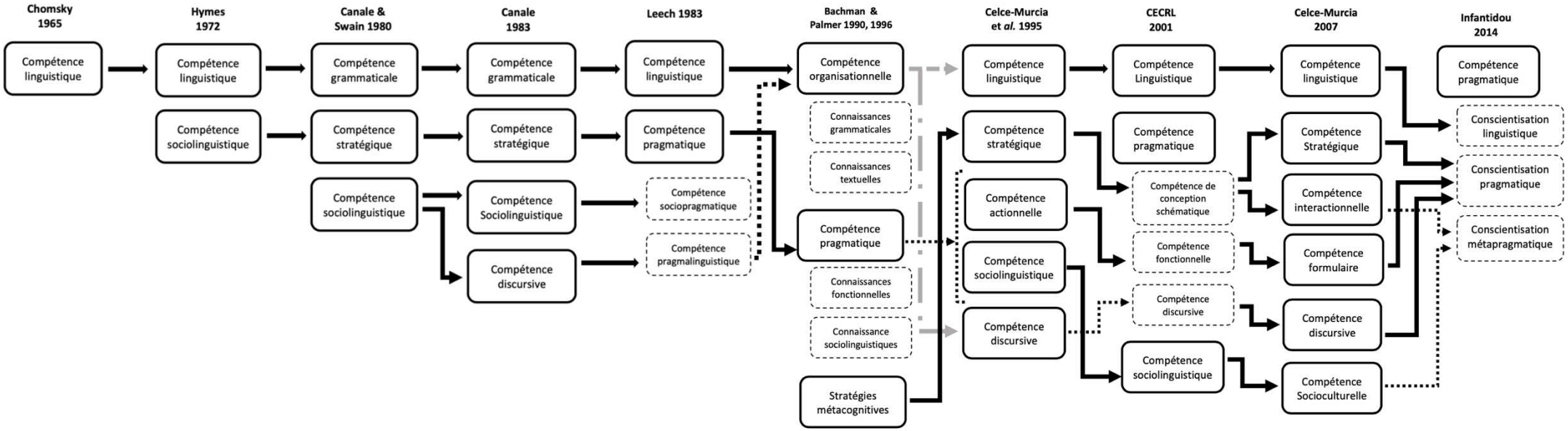


Figure 11: Synthèse chronologique de l'évolution de la compétence communicative et pragmatique (basé sur le schéma de Celce-Murcia (2007))

Après avoir présenté le contexte épistémologique dans lequel s'est développée la notion de compétence pragmatique, nous allons à présent détailler notre modèle de la compétence lexicopragmatique.

3. La compétence lexicopragmatique : définition et modélisation

Dans cette partie, nous allons détailler notre approche de la compétence pragmatique utilisée dans notre analyse pour décrire les différents niveaux de compétence mobilisés dans l'emploi des PhPex. La question de la difficulté d'interprétation des séquences phraséologiques est ancienne, Bally le remarquait déjà dans son *Traité de stylistique française* :

> « *C'est une erreur de croire que, dans la pratique de la langue maternelle, on comprend tout ce qu'on dit ou écrit; il nous échappe en parlant une foule de* ***mots sans signification pour nous****, et* ***nous conservons****,* ***à notre insu****, une quantité de vestiges du passé de la langue dont* ***nous ne saisissons pas le sens, même grossièrement****. En un certain sens, l'étranger qui apprend le français* […] ***il le saisit plus analytiquement que nous****; supériorité peu enviable, du reste, car, dans l'étude de la phraséologie, ce* ***don d'analyse est aussi préjudiciable*** *que dans l'étude des mots isolés* […]». (Bally 1909 : 78, nous soulignons)

Bally explique que les stratégies de décomposition analytique posent particulièrement problème aux LNN, car les phraséologismes sont pour la plupart mémorisés en bloc par les LN et utilisés en tant qu'unité « prête-à-parler » pour leur fonction communicative. La compétence lexicale seule ne semble donc pas suffisante pour la compréhension des phraséologismes pragmatiques, l'analyse des constituants ne permettant pas d'accéder au sens ou à la fonction du phraséologisme.

L'angle d'analyse que nous proposons ici est celui de la *compétence lexicopragmatique*. L'enjeu de cette dénomination est de mettre l'accent sur la composante lexicale et combinatoire dans la compétence pragmatique mobilisée lors de l'emploi des PhPex. En effet, nous postulons ici que même en ayant une connaissance pragmalinguistique et sociopragmatique approfondie, la méconnaissance des potentialités lexicales et combinatoires disponibles lors de l'apparition d'un besoin communicationnel en contexte peut être source de difficulté pour le LNN. De même, nous postulons qu'une connaissance lexicale et combinatoire, aussi approfondie soit-elle, ne permet pas au LNN de contourner certaines difficultés d'emploi liées au PhPex si le développement des compétences

sociolinguistique et pragmalinguistique ne s'opère pas en parallèle. Enfin la troisième hypothèse est celle de la nécessité de la construction d'un niveau métacognitif de compréhension et de prototypage des phénomènes phraséopragmatiques et communicationnels auxquels le locuteur est confronté. Il s'agit d'un niveau de compréhension méta qui permet au locuteur d'identifier les facteurs contextuels et linguistiques qui conditionnent la réalisation et la validité de la fonction communicative dont il est question. Cet angle d'analyse lexicopragmatique nous semble ici indispensable pour comprendre la compétence de niveau actionnel et praxéologique mobilisée par le locuteur afin de sélectionner une forme adéquate sur le plan pragmatique et socioculturel, mais également appropriée du point de vue des spécificités du lexique employé et de ses propriétés syntactico-sémantiques.

Nous partons ici du principe que les facteurs contextuels, les connaissances linguistiques, la compréhension des aspects conventionnels liés à l'emploi et le degré de conscience analytique sur le processus en cours du locuteur sont mobilisés de manière complémentaire et simultanée lors de l'utilisation de la compétence lexicopragmatique. Cette compétence n'est pas une conséquence immanente de l'exposition au contexte d'interaction de la langue cible, mais bel et bien le résultat de la mobilisation de stratégie à plusieurs niveaux et d'ajustements de ces stratégies en fonction des spécificités de la situation d'énonciation. Cette compétence n'est donc pas statique et obéit à des contraintes au niveau du contexte, du système linguistique et du locuteur lui-même. La compétence lexicopragmatique renvoie au savoir-faire permettant au locuteur natif (ou LN) ou à l'apprenant de faire une utilisation intelligible et cohérente de ses ressources linguistiques pour : a) inférer le sens, b) repérer les spécificités lexicales, sémantiques et les traits formels des expressions préfabriquées, c) identifier le fonctionnement pragmatico-discursif du lexique dont l'emploi est régi par des contraintes contextuelles et interpersonnelles et d) répondre de manière conventionnelle à des enjeux sociolinguistiques de communication (Crystal 1997 : 240; Taguchi 2009 : 4; Kecskes 2015 : 420; Sikora et Tsedryk 2019). Dans la figure suivante, un schéma modélisant les différents niveaux de la compétence lexicopragmatique :

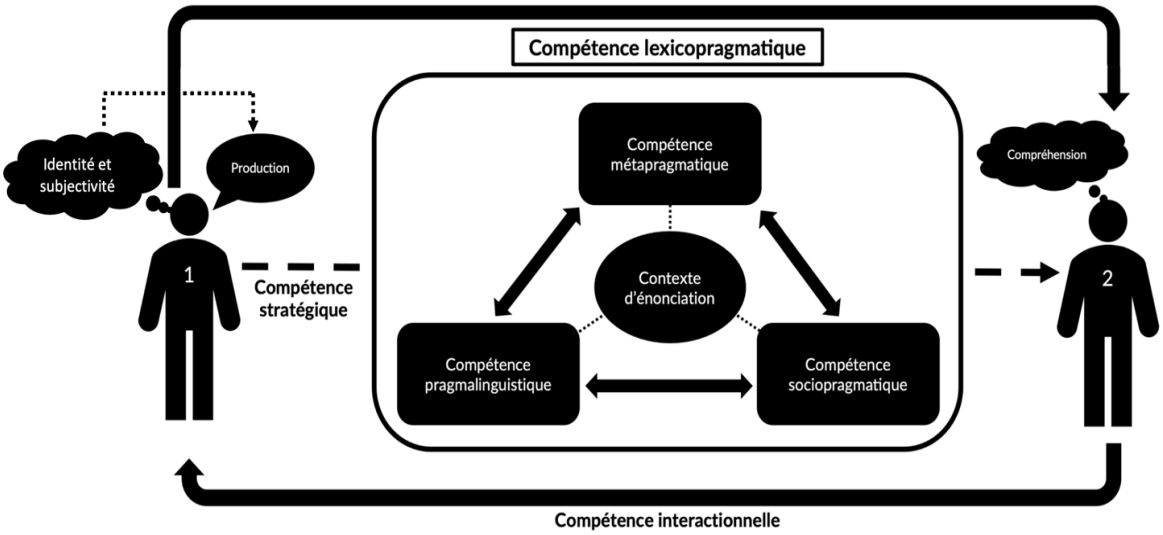


Figure 12 : Modélisation de la compétence lexicopragmatique et son articulation dans l'acte de communication

Cette modélisation repose sur le cadre épistémologique présenté dans la section précédente, et plus particulièrement sur les modèles de Leech (1983), de Celce-Murcia (2007) et d'Ifantidou (2014). Le modèle Leech (1983) est intéressant ici, car il met en avant les différents niveaux de la compétence pragmatique et leur interaction lors de la production et de la compréhension. L'approche de Celce-Murcia (2007) ajoute la dimension dynamique de la compétence pragmatique avec la prise en compte des aspects interactionnels et stratégiques qui nous intéressent tout particulièrement. Enfin, le modèle d'Ifantidou (2014) élargit un peu plus la définition de la compétence pragmatique en intégrant les processus métacognitifs d'interprétation et de sélection des formes appropriées chez les LNN. Toutefois, dans le cadre de notre étude, il nous a semblé que ces modèles trouvaient leurs limites pour décrire la compétence d'usage des PhPex. En effet, ces modèles restent assez généralistes et finalement peu éprouvés sur des phénomènes pragmatiques spécifiques à l'interaction orale quotidienne. Ils ne prennent pas vraiment en compte le statut des locuteurs et l'influence que peuvent avoir les spécificités des locuteurs sur le choix de formes linguistiques. Il nous a donc fallu proposer un cadre d'analyse de la compétence pragmatique qui soit compatible avec notre approche située à l'interface de la phraséologie pragmatique des interactions et de l'acquisition de la compétence pragmatique en L2.

D'une part, nous avons mis l'accent dans cette modélisation sur l'aspect dynamique et complémentaire des sous-compétences entre elles. D'autre part, nous avons voulu représenter l'articulation de chaque sous-compétence avec le reste du système de

communication. Nous prenons en compte dans ce modèle le point de vue de l'émetteur et du récepteur, c'est-à-dire à la fois la production et la (méta)compréhension. De même, nous prenons en compte les dynamiques interactionnelles et stratégiques. L'acte de langage n'est jamais isolé et se situe dans un schéma bidirectionnel d'action-réaction entre le locuteur et son interlocuteur. Les choix linguistiques qui mobilisent la compétence lexicopragmatique sont également le résultat de l'évaluation de ces dynamiques bidirectionnelle et de l'effet que peut avoir une énonciation sur ces dynamiques. Dans la plupart des cas[74], l'énonciation est située et orientée en fonction d'un besoin communicatif ou d'un objectif discursif particulier. La relation et le rôle de chacun des participants sont aussi une dynamique importante. La relation hiérarchique ou l'intimité relationnelle sont par exemple deux paramètres qui peuvent grandement influencer le type de forme qui sera employé pour une même fonction. Enfin, nous avons ajouté la dimension subjective et identitaire du locuteur qui joue également un rôle dans la construction de la compétence lexicopragmatique. Ces différents facteurs impliqués dans la sélection et la compréhension des PhPex seront testés dans notre protocole expérimental. Nous allons à présent détailler les trois pôles de notre approche dans les sections qui suivent.

3.1. Le niveau pragmalinguistique

Nous avons choisi le terme de *pragmalinguistique,* car celui-ci rend compte de l'interface entre le système formel de la langue et ses réalisations pragmatiques en contexte de communication. La compétence pragmalinguistique est une compétence de *niveau microstructurel* (matérialité linguistique) renvoyant à la connaissance du système linguistique, ainsi qu'à la connaissance des possibilités combinatoires et actionnelles qu'il offre pour générer des formes linguistiques adaptées au besoin communicatif rencontré. D'autre part, cette compétence concerne la capacité à générer des emplois structurés, respectant les principes inhérents de référence et d'intelligibilité de la langue cible.

Sur le plan structurel, la compétence pragmalinguistique rend compte de l'habileté à pouvoir mener et organiser le propos selon les standards de cohérence et de cohésion de la langue cible. Il s'agit notamment de la manière dont le propos est organisé et articulé dans la langue

---

[74] A l'exception peut-être du soliloque qui est un cas particulier ici. L'énonciation est orientée vers soi, mais les fonctions demeurent les mêmes. Par exemple, s'auto rassurer (*ça va le faire !*), se résigner (*bon bah pas le choix …*) ou s'exclamer (*Putain !*).

cible selon les règles grammaticales propres à l'écrit ou à l'oral. La compétence pragmalinguistique implique également la compréhension de la relation forme-fonction et de ses effets sur la situation de communication (par ex. : énoncé sentencieux et élaboré syntaxiquement = politesse et formalité, ou en japonais, marquage des relations interpersonnelles au niveau morphosyntaxique). Cette sous-compétence implique ainsi la capacité à solliciter les outils linguistiques disponibles dans la langue cible pour agir sur la situation de communication ou pour agir sur le dire de l'interlocuteur, et ce de manière à ce que cet acte s'intègre de manière dynamique dans le discours (Leech, 1983). En outre, il s'agit de la capacité à utiliser des formes appropriées à la visée communicative de l'interaction et dont la combinatoire et l'actualisation du sens permettent la réalisation d'une fonction pragmatique donnée. Il peut s'agir par exemple de l'usage de constituants périphériques pour renforcer l'aspect expressif des PhPex (ex. : l'usage de, *mais* avec le PhPex *c'est pas possible ! => mais c'est pas possible !*).

3.2. Le niveau sociopragmatique

Nous avons choisi le terme de *sociopragmatique,* car celui-ci rend compte de l'interface entre la dimension sociale de la langue et la réalisation pragmatiques de ces phénomènes sociaux en contexte de communication. La compétence sociopragmatique est une compétence de *niveau mésostructurel* (contexte et environnement proche) renvoyant dans à la connaissance de la structuration sociale et des règles de communication et comportementales propre à la communauté de locuteur dans laquelle la langue est employée. Ainsi, la compétence sociopragmatique constitue « l'interface sociologique de la pragmatique » (Leech, 1983, p. 10), c'est-à-dire la compréhension des phénomènes sociaux qui façonnent l'interprétation et l'emploi des phraséologismes pragmatiques. Cette compétence suppose une systémisation en cours de langue et une pratique régulière et immersive afin de permettre la pérennisation des connaissances sociopragmatiques observées en contexte naturel de communication. En outre, l'acquisition de cette compétence implique de développer un éventail de schémas conversationnels, eux-mêmes composés de scripts pragmatiques déclenchant l'usage d'un paradigme d'expressions préfabriquées et particulièrement adaptées à l'acte de communication visé (Dewaele et Wourm, 2002, p. 130). Les scripts constituent un ensemble de processus cognitifs permettant au locuteur de comprendre les informations indexicales liées aux cadres de communication qu'il

rencontre et de saisir la complexité des différents paramètres qui président à l'usage des PhP, paramètres établis de manière conventionnelle au sein d'une communauté linguistique donnée, sur la base de critères socioculturels.

La compétence sociopragmatique apparait comme la dimension la plus complexe à analyser et à mesurer. La forte conventionnalité de normes sociales associées à une forme donnée est de nature tacite, impliquée lors de l'usage, mais d'importance primordiale. Le LNN a rarement accès à ces informations de manière explicite et c'est par l'expérience de l'interaction authentique et l'observation solitaire ou conjointe (avec l'enseignant et/ou le locuteur natif) qu'il peut développer sa compréhension de cette dimension. En outre, même si certains prototypes propres à une culture sont observables, on distingue à un niveau plus situationnel une grande variation en fonction des communautés de locuteurs d'une même langue culture. Cette variabilité, elle aussi tacite, peut poser certaines difficultés au LNN, car elle s'établit en fonction de paramètres et de marques qui sont bien souvent implicites. Un exemple assez populaire de cette variabilité est le vouvoiement qui, même s'il obéit à des règles d'emploi explorées dans de nombreux travaux et est vraisemblablement enseigné de manière précoce en classe de langue[75], est souvent source de difficulté pour le LNN. En effet, en plus des facteurs prototypiques enseignés, il existe tout un ensemble de micro-facteurs contextuels comme la relation interpersonnelle, l'engagement dans la conversation ou le degré de connivence entre les participants, qui peuvent modifier l'usage du vouvoiement.

Notons en outre que notre modèle prend en compte le profil psychologique et identitaire du locuteur. Les études récentes sur le lien entre profil du LNN et acquisition de la compétence pragmatique ont observé que le facteur identitaire jouait un rôle important dans le choix des formes pragmatiques employées (Iwasaki 2011, Brown 2013, Kim 2014). En effet, ces études montrent que le locuteur n'applique pas à l'aveugle la norme pragmatique établie dans la communauté dans laquelle il évolue. Il a ainsi été observé que le LNN pouvait rejeter certains usages, par exemple l'usage de certaines formes familières ou du langage genré, lorsque ces usages entrent en contradiction avec l'identité de locuteur, les valeurs morales ou la perception de soi qu'a ce LNN. Par exemple, il a été fréquemment remarqué dans les

[75] Il s'agit d'une hypothèse basée sur l'observation de quelques manuels de communication orale.

études portant sur les apprenants coréens et japonais de l'anglais que ce type de LNN ne se sentait pas à l'aise lorsqu'ils devaient accepter un remerciement et remercier l'interlocuteur. Cela a notamment été expliqué par l'éthos communicatif de ces locuteurs fortement axé sur la valeur de la modestie qui entrait en contradiction avec la mise en valeur de soi liée au compliment. Les productions jugées inadéquates au regard de la norme pragmatique sont donc intéressantes à observer en parallèle du profil du LNN. Cette observation permet de comprendre de quelle manière ces conflits d'identité peuvent générer des emplois qui mettent en avant l'agentivité du locuteur dans ses propres décisions pragmatiques, et ce de manière à maintenir leur éthos communicatif (Clyne 2006). Il semble donc pertinent de considérer certaines des réponses ou non-réponses jugées inadéquates non pas comme des résultats déviants, mais comme des formes qui peuvent aussi témoigner de choix conscients et identitaires.

3.3. Le niveau métapragmatique

Nous avons choisi le terme de *métapragmatique,* car celui-ci rend compte du niveau de conscientisation (Ifantidou 2014) des mécanismes communicatifs impliquant le système formel de la langue, ses fonctions pragmatiques possibles et les constructions sociolinguistiques et socioculturelles qui lui sont propres. La compétence pragmalinguistique est une compétence de *niveau macrostructurel* (prototypage communicationnel) renvoyant à la connaissance globale de l'articulation entre la forme linguistique, la dimension actionnelle et praxéologique du système linguistique et les manières dont ses fonctions se manifestent sur le plan social et culturel. La compétence métapragmatique relève à la fois de la discussion épilinguistique et ontologique autour de la méta-analyse des phénomènes communicationnels dans une perspective généralement onomasiologique (de la fonction à la forme) ; et du phénomène de conscientisation par l'observation des phénomènes pragmatiques que le locuteur rencontre et qui s'opère généralement dans une perspective sémasiologique (de la forme à la fonction).

La compétence métapragmatique constitue la capacité du locuteur à déduire le sens, les fonctions ou les conditions de production d'une séquence préfabriquée à fonction pragmatique donnée en se basant sur son observation et sur les connaissances linguistiques issues de son répertoire linguistique. Il s'agit ainsi d'une compétence de reconstitution des spécificités d'emploi d'une expression donnée basée sur l'observation et la déduction

empiriques. Dans le cas du locuteur non natif, celui-ci procède le plus souvent à tâtons, de manière inductive en alternant entre essai, échec et formulation d'hypothèse. Cette posture métapragmatique permet au LNN de comprendre les implications de certaines expressions, d'en comprendre la portée et les effets sur son interlocuteur, sur la relation sociale qui les unit ou sur la situation de communication en elle-même. La compétence métapragmatique ne se limite cependant pas à la seule capacité de description de l'usage. Elle implique également la perception des conditions d'opérationnalité de l'expression ou des répercussions en cas de violation des contraintes d'usage (cf. *Bris de rituel* (Dostie 2018) ou *échec communicatif* (Bardovi-Harlig 2001)). Elle permet également au locuteur d'évaluer l'acceptabilité d'un emploi ou de juger ou non de la possibilité d'un défigement ou d'une remotivation d'une séquence préfabriquée. Cette compétence dépasse le niveau de l'énonciation pour aller dans la structure prototypique de l'usage (ou ce que l'on pourrait appeler le *macrosystème communicationnel*).

La compétence métapragmatique relève également de la perception de l'indexicalité[76] associée à des formes linguistiques à fonction pragmatique, c'est-à-dire des informations contextuelles intrinsèquement liées à l'emploi dont il est question. Il s'agit de la capacité d'inférence d'éléments utiles à la compréhension du PhPex et souvent disséminés dans le cadre d'interaction, ce que Gumperz (1982) appelle des *indices de contextualisation* (*contextualization cues*). On peut mentionner par exemple l'intonation ou le ton qui permet d'inférer l'intention ou l'expressivité de l'interlocuteur ou la proxémie entre les différents participants lors de l'interaction. De surcroit, l'expérience joue un rôle déterminant dans la construction de la conscience métapragmatique. En effet, l'analyse métapragmatique procède souvent de la comparaison avec l'expérience propre du locuteur lors de la conversation. C'est de cette expérience de l'interaction qu'émerge la capacité à associer un style communicatif à un type de situation de communication. Nous pensons qu'il est tout à fait utile et nécessaire de développer chez le LNN la compétence métapragmatique, car elle constitue une habileté complémentaire aux compétences pragmalinguistiques (forme + sens codifié dans la langue) et sociopragmatique (contraintes sociales et contextuelles).

---

[76] « L'indexicalité, ce sont toutes les déterminations qui s'attachent à un mot, à une situation. Indexicalité est un terme technique adapté de la linguistique. Cela signifie que bien qu'un mot ait une signification trans-situationnelle, il a également une signification distincte dans toute situation particulière dans laquelle il est utilisé » (Rogers, M.F., 1983 : 95)

Nous achevons ce chapitre concernant d'une part le développement de la notion de compétence pragmatique au sein de la SLP, d'autre part, la définition de la compétence lexicopragmatique qui constituera notre angle d'analyse dans la deuxième partie de ce travail de thèse. Après avoir situé notre étude dans le champ de la pragmatique de l'acquisition de la L2 et avoir défini la compétence lexicopragmatique qui constitue le cœur de notre étude, nous allons aborder les dispositifs de mesure et d'évaluation de la compétence (lexico)pragmatique.

Chapitre 7 - Méthode d'élicitation et d'évaluation de la production pragmatique en L2[77]

Dans cette section, nous allons aborder la question du recueil des données permettant d'éliciter et d'évaluer la compétence des LNN à produire des énoncés dotés de fonctions actionnelles. Nous aborderons notamment les choix méthodologiques qui ont été faits lors de l'élaboration de notre protocole expérimental. L'exploration de ces différentes méthodes nous a ainsi permis de sélectionner les méthodes les plus appropriées à l'observation de la compétence lexicopragmatique mobilisée dans la production et la compréhension des PhPex des LNN observés dans notre étude. Nous diviserons cet inventaire en deux temps : la première partie sera consacrée aux protocoles permettant d'explorer la compétence lexicopragmatique du locuteur en autonomie, la seconde partie abordera les protocoles mobilisés dans l'exploration de la compétence soit en interaction orale réelle ou simulée, soit par une modalité d'élicitation en groupe. Un tableau de synthèse des différents protocoles abordés est disponible dans l'annexe 2.

1. Les protocoles non-interactifs

Les protocoles non-interactifs sont désignés ici comme les protocoles réalisés en autonomie par l'informateur sans la présence d'un interlocuteur. Ils ont pour fonction principale d'évaluer la production ou la compréhension de manière ponctuelle et non-dynamique. L'avantage premier qu'offre cette catégorie est la rapidité de mise en place et la possibilité de tester plusieurs variables de manière contrôlée. Ce type de protocole a principalement été mobilisé dans l'élaboration du test lexicopragmatique employé dans la deuxième phase de notre protocole expérimental. Nous voulions en effet tester deux aspects de la compétence lexicopragmatique : 1) la capacité à produire un PhPex pertinent du point de vue de la situation de communication et 2) observer concrètement les interactions avec le répertoire langagier de l'informateur lors de la sélection de la forme adéquate.

[77] Les méthodologies de récoltes abordées dans cette section constituent principalement des méthodes de recueil adaptées à l'étude des phraséologismes pragmatique. Au-delà des protocoles présentés ici, il existe tout un ensemble de méthodes plus spécifiques à l'étude d'autres compétences telles que la compétence rédactionnelle ou l'usage de la prosodie à l'oral. Pour plus de détails, on pourra consulter Nguyen (2019) qui constitue l'une des méta-analyse des méthodes d'évaluation les plus récentes.

1.1. Activité de complétion du dialogue (*Discourse completion task* ou DCT)

L'activité de complétion constitue un protocole d'élicitation de la production dont les premiers développements sont attribués à Blum-Kulka (1982). Le principe général de ce protocole est d'inviter le LNN à produire la suite d'un extrait d'interaction oral en prenant en compte les éléments du contexte, mais aussi les éléments linguistiques contenus dans le dialogue. La mise en place de ce protocole implique l'utilisation d'un élément de contextualisation court suivi d'un dialogue suffisamment long pour permettre au LNN de disposer d'assez d'éléments pour répondre. Le protocole implique également une réflexion sur la façon dont est découpé l'extrait à compléter, de manière à ne pas créer une difficulté liée au format de l'activité. Le dialogue à compléter peut comporter plusieurs emplacements vides. La situation de communication présentée doit être assez explicite pour permettre au LNN l'identification de la fonction attendue. Sont précisés en amont : le but de la communication, l'action en cours, la relation interpersonnelle entre les participants et tout autre élément du cotexte. Enfin, il est généralement demandé au répondant de se mettre dans la peau du locuteur du dialogue afin de répondre avec le plus de naturel possible. Cela permet de faciliter la production dans les cas où le locuteur ne souhaite pas répondre en son nom, car il ne semble pas très familier avec la situation proposée. Le projet CCSARP (*Cross-Cultural Speech Act Realization Project*) mené par Blum-Kulka, House et Kasper (1989) est un exemple de projet mobilisant le protocole du DCT pour explorer de manière empirique la compétence pragmatique en L2 à travers plusieurs situations de communication (requête et excuse) et plusieurs langues.

Les avantages du DCT sont multiples. Le protocole est relativement simple à mettre en place et demande peu de moyens. Il permet de moduler les paramètres contextuels à tester de manière à observer précisément une hypothèse concernant un phénomène pragmatique donné. Par ailleurs, les variables contextuelles étant manipulables, ce protocole est particulièrement bien adapté à la comparaison entre langues, permettant ainsi d'observer des éléments de variation entre les différentes stratégies mobilisées par les LNN. De plus, ce protocole offre un cadre de comparaison rigoureux avec le groupe contrôle des LN étant donné la grande reproductibilité de la procédure. En outre, le DCT est particulièrement adapté au recueil rapide et massif de données sur une grande population de répondants. Enfin, l'avantage non-négligeable est la possibilité d'éliciter des phénomènes pragmatiques

habituellement complexes à provoquer en situation naturelle de communication, par exemple les insultes.

Concernant les désavantages, plusieurs critiques ont été adressées au protocole du DCT. La première critique concerne le fait que les DCT sont fabriquées par le chercheur en se basant sur ses propres connaissances de locuteur pour concevoir les dialogues et les situations envisagées, ce qui peut créer un décalage entre ces situations et celles que le LNN est amené à rencontrer. Il peut s'agir par exemple de demander au locuteur des emplois spécifiques au monde de l'entreprise alors que celui-ci n'a jamais été en contact avec ce type d'interaction. La deuxième critique concerne le manque d'authenticité ou le décalage avec la réalité de l'usage des dialogues s'ils sont entièrement fabriqués. La dernière critique, sans nul doute la plus importante, concerne le fait que le DCT se compose d'un dialogue trop court et pas assez élaboré pour permettre au LNN de pouvoir développer une connaissance globale à la fois au niveau illocutoire, interactionnel et paraverbal (Kasper 2008, Bardovi-Harlig 2010). Ce type d'activité ne s'inscrit pas dans la temporalité authentique du dialogue, mais dans un cadre de réponse artificiel laissant un temps de réponse que le LNN n'aurait pas habituellement dans une interaction réelle. En d'autres termes, les DCT permettent d'observer la compétence à produire une forme pragmatique en contexte, mais ne sont pas révélateurs des choix linguistiques en contexte réel de communication articulée entre plusieurs locuteurs (Félix-Brasdefer, 2010). Les conséquences inhérentes aux choix linguistiques faits par le locuteur ne sont pas observables, car la scène de communication est statique. De même, ce type de test ne permet pas d'évaluer la capacité du LNN à développer une conversation longue. Ce type de mesure reste ponctuel et ciblé.

Nous avons choisi ce type de protocole pour plusieurs raisons. Il nous a paru intéressant, car il permet de tester la compétence de production tout en ayant un contrôle sur les variables à observer. Cela est d'autant plus important que nous souhaitions mettre en place des analyses statistiques et quantitatives. La standardisation des conditions de réponse qu'offre le DCT permet de garantir la fiabilité de ces mesures. En outre, le DCT offre la possibilité d'éliciter une grande variété de situations de communication et de faire varier une large gamme de paramètres, ce que ne permettent pas les enregistrements d'interactions non-élicitées. L'élicitation contrôlée du DCT nous a semblé particulièrement adaptée à la mise en œuvre d'un test de production et de compréhension. En complément au DCT habituel, nous avons proposé une variante de ce protocole sous la forme d'une activité d'équivalence où il était

demandé au LNN de remplacer le PhPex dans le dialogue proposé par un équivalent fonctionnel en français et en japonais. Cette contrainte supplémentaire nous a permis de tester la capacité du LNN à repérer les éléments importants de la situation de communication avant de produire la forme appropriée.

1.2. Journal de suivi (ou autobiographie d'apprenant)

Ce type de protocole consiste à demander au LNN de rédiger dans un journal personnel des observations concernant des phénomènes se déroulant dans une situation d'interaction en langue cible. Ces observations peuvent être de l'ordre de réflexions personnelles, d'étonnement interculturel, d'analyse ou d'hypothèses concernant l'usage d'une structure pragmatique ou une manière de communiquer. Les observations formulées par l'informateur peuvent également concerner le développement de sa compétence ou de sa compréhension d'un phénomène en particulier. La rédaction du journal peut être combinée à des enregistrements prélevés en situation naturelle de communication qui seront ensuite commentés par le LNN afin d'expliciter les stratégies communicatives employées lors de l'usage d'une structure donnée. Les entrées discutées dans le journal peuvent être proposées par le chercheur afin de cibler l'attention du LNN sur des évènements communicationnels particuliers. Ce protocole constitue l'une des manières les plus empiriques d'explorer la compréhension métapragmatique des LNN de par sa nature autoréflexive et analytique. Le journal permet également d'obtenir des détails concernant les différentes variables entrainant des variations d'emploi en contexte (Hassall 2006). En outre, ce protocole peut être particulièrement utile pour observer s'il existe un décalage entre la connaissance de la fonction de l'expression et la capacité à réaliser un usage effectif de l'expression en question. En effet, nous l'observerons dans l'analyse, le LNN a bien souvent une connaissance stable du phraséologisme pragmatique qui lui est proposé, mais ne parvient tout de même pas à répliquer son emploi. Le journal constitue donc un protocole intéressant permettant d'explorer les paramètres liés à l'emploi (par exemple des paramètres contextuels comme la relation, le niveau de politesse ou l'enchainement dans la séquence interactionnelle). En outre, les données demeurent authentiques et étayées, avec à la fois la dimension pragmalinguistique et métapragmatique, mais également les paramètres sociopragmatiques impliqués dans la production du LNN. Ce protocole constitue un bon complément pour compléter une étude empirique sur des données prélevées au cours d'échanges authentiques. Le journal de suivi est également un support unique pour l'exploration de la dimension

interculturelle et translinguistique. En effet, l'informateur peut y consigner certaines de ses observations sur les différences d'emploi ou sur les proximités et les différences linguistiques remarquées entre les différentes langues de son répertoire. Ces informations pourront ensuite être traitées par le chercheur afin de construire des hypothèses à vérifier sur des corpus de production ou lors d'entretiens métapragmatiques.

L'un des points faibles de la démarche est le caractère autocentré du protocole qui peut rendre les données recueillies fortement influencées par la subjectivité du locuteur. Ce faisant, on ne sait pas si l'informateur évoque les faits avec détail et sincérité. En outre, le locuteur qui ne possède pas le même regard que le chercheur ni les hypothèses qui sous-tendent l'étude est susceptible de ne pas mentionner des informations cruciales (Kasper et Rose 2002). Le LNN peut également choisir de mentionner ou non certains évènements, voire d'altérer certains éléments, notamment ceux qui pourraient attenter à son image de locuteur. Ce type de données nécessite un traitement très long et qualitatif qui exclut d'emblée les mesures statistiques et la comparabilité intersujet. Enfin, la mise en place du protocole nécessite un cadrage préalable de la démarche auprès du LNN par le chercheur qui peut être lourd et chronophage à constituer. Enfin, il semble également difficile de trouver des informateurs volontaires étant donné le caractère intrusif que peut avoir cette méthode d'élicitation.

Ce protocole étant avant tout à visée longitudinale, nous avons choisi de l'écarter en raison du format de notre étude. En effet, notre protocole vise à étudier les spécificités d'un état de la compétence lexicopragmatique des LNN observés à un instant T. Il ne s'agit pas d'étudier l'évolution de cette compétence dans le temps. Il nous a donc semblé que ce protocole ne serait pas utile à notre étude.

1.3. Questionnaires à choix multiples (QCM)

Ce type de protocole part du principe que la compétence pragmatique implique de manière simultanée la compréhension et la production. Cette compréhension peut intervenir soit en amont de la production afin de sélectionner la forme adéquate, soit en aval afin de comprendre la forme à laquelle le locuteur est exposé. Lorsque le locuteur interprète, il se base avant tout sur les informations contextuelles à sa portée qu'il va décoder en parallèle du contenu linguistique. La communication implique donc le décodage d'indices de

contextualisation qui permettent d'une part de comprendre l'intention de locuteur, d'autre part de choisir la forme adaptée au contexte communicatif rencontré.

Le QCM constitue le premier type de protocole employé par la SLP pour observer la compétence des LNN en réception (Carrell 1979). Le principe du protocole consiste à soumettre au LNN un dialogue dans lequel un phénomène pragmatique en particulier est mis en évidence. Il est ensuite proposé dans le test plusieurs assertions correspondant au sens, à l'un des paramètres d'usage ou à la fonction du phénomène observé dans le dialogue. Le LNN doit alors choisir entre ces diverses affirmations celle qui coïncide avec le phénomène observé. Ce type de protocole est exigeant sur le plan cognitif, notamment pour le LNN, puisqu'il implique, en plus du processus interprétatif du phénomène pragmatique et du cadre de communication, de devoir reconstruire le sens des unités linguistiques concomitantes du phénomène observé (Taguchi 2008). Les QCM permettent également de tester la compétence de reconnaissance pragmalinguistique et sociopragmatique, c'est-à-dire la compétence permettant au locuteur de juger de l'acceptabilité pragmatique d'une forme linguistique en contexte parmi plusieurs propositions. Le test est ainsi construit de manière à observer si le locuteur est à même de repérer les paramètres sociopragmatiques associés à la forme. Le test peut aussi associer au jugement d'acceptabilité du locuteur, une échelle de Likert[78] mesurant le degré d'adéquation de l'énoncé estimé par ce locuteur. Ces tests d'acceptabilité à choix multiple sont tout à fait utiles pour mesurer le jugement d'acceptabilité et les critères qui fondent ce jugement, lorsque couplés avec un protocole réflexif comme le journal d'observation vu précédemment.

L'avantage de ce type de protocole est qu'il est rapide à mettre en place et permet une collecte massive de données. Facilement formalisable, il permet des traitements statistiques corrélatifs et inférentiels. En outre, le test est particulièrement adapté à des analyses comparatives entre différents locuteurs, différentes langues ou différents phénomènes. La grande paramétrabilité du protocole permet au chercheur d'observer un large spectre de phénomènes dans diverses conditions d'interaction. Le protocole peut-être aussi facilement

---

[78] L'échelle de Likert constitue un outil psychométrique permettant de mesurer une échelle d'évaluation du répondant face à un phénomène. Elle consiste en une affirmation au sujet de laquelle l'informateur dois choisir une valeur graduelle. L'échelle contient pour chaque item une graduation comprenant en général cinq ou sept choix de réponse qui permettent de nuancer l'évaluation.

proposé sur papier qu'en ligne via un formulaire, ce qui permet de pouvoir réunir un maximum de locuteurs tout en s'affranchissant des contraintes géotemporelles.

Les principales critiques qui sont faites concernent le caractère fabriqué et artificiel qui ne permettrait pas de mesurer de manière précise la compréhension dans des conditions réelles de communication avec tous les éléments aléatoires et imprévisibles que comporte une situation d'interaction. Le test est aussi très exigeant, car il implique la réalisation simultanée d'une tâche de compréhension, d'analyse et de discrimination pour chaque énoncé proposé. En outre, ce type de protocole ne permet pas de mesurer les stratégies interactionnelles ou les phénomènes de co-construction. Le QCM ne permet également pas de mesurer la fluidité ou l'aisance d'interprétation ou de réponse du LNN. Le format du QCM peut également engendrer des risques de réponses aléatoires ou des phénomènes de précipitation liés à la fatigue cognitive. Un autre problème est la difficulté à intégrer la dimension multimodale (image et son) sans surcharger le test en information à traiter. Enfin, l'élaboration des items de question et de réponse nécessite de prendre certaines précautions pour s'assurer que le contenu est clair et sans ambigüités. Ce processus de vérification peut être chronophage et nécessiter plusieurs tests pilotes qui engendreront de fait de multiples ajustements.

Nous avons sélectionné ce protocole dans le cadre de notre test lexicopragmatique. En effet, les deux dernières activités du test sont axées sur la compétence de compréhension de PhPex proposé dans plusieurs contextes de communication. Nous avons opté pour deux modalités de QCM : 1) un QCM lexical axé sur la compétence de sélection et d'évaluation de la forme linguistique appropriée, 2) un QCM sémantique permettant de tester la capacité du LNN à sélectionner l'interprétation la plus appropriée aux contextes proposés. Toutefois, nous le verrons dans la partie de l'analyse dédiée au test, la mise en oeuvre de ce protocole a été relativement complexe. Nous avons dû mettre en place un test pilote afin de vérifier la pertinence des items sélectionnés. Malgré ce test pilote, il s'est avéré que ce type d'élicitation pouvait présenter un certain nombre de difficultés d'interprétation et de désambigüisation, tant pour le LN que pour les LNN. Ces difficultés n'étaient cependant visibles qu'à l'issue de l'analyse des résultats où nous avons pu remarquer des scores de réussite faible dans les deux cohortes pour certains items.

1.4. Questionnaires à échelle (*scaled-response questionnaires*)

Ce type de protocole exploratoire est particulièrement adapté à la mesure de la compétence métapragmatique. Il consiste à soumettre un dialogue comportant un phénomène pragmatique à observer au LNN. Le LNN est ensuite invité à évaluer le niveau d'adéquation de la réponse en sélectionnant la valeur appropriée sur une échelle de Likert. L'échelle peut aussi concerner le degré d'influence de paramètres contextuels tels que la relation entre les participants ou le niveau de politesse dans la réponse des locuteurs à un test de production réalisé au préalable. Ce type de questionnaire permet au chercheur d'obtenir des informations utiles concernant les paramètres contextuels qui permettent au LNN d'évaluer la pertinence de la réponse. Par ailleurs, le questionnaire à échelle est complémentaire des tests en production puisqu'il permet d'éliciter certaines des stratégies déployées par le LNN pour répondre au test. Ainsi, une des premières applications de ce type de protocole est celle de Maeshiba, Yoshinaga, Kasper et Ross (1996) en tant qu'extension aux données obtenues dans un DCT. Cette complémentarité a notamment permis de mesurer le degré de correspondance entre l'adéquation de la production des LNN (anglais L2) et leur évaluation de l'importance des facteurs contextuels influençant leur choix.

Le questionnaire à échelles possède plusieurs avantages communs avec les DCT. Il est rapide à mettre en place et permet de mener des études statistiques à grande échelle. En outre, ce protocole permet un contrôle optimal des variables, limitant ainsi les biais ou le bruit dans les réponses. La forte standardisation et la reproductibilité du protocole permettent de faciliter la comparaison entre les types d'informateurs, les phénomènes observés et les langues ciblées. Tout comme les QCM, la forte paramétrabilité des items de question permet d'explorer à la fois la compétence pragmalinguistique, sociopragmatique et métapragmatique. Enfin, ce type de protocole a l'avantage de s'intégrer très facilement à d'autres protocoles non-interactifs tels que les DCT ou les QCM, mais également avec des protocoles interactifs lorsqu'ils sont proposés à la suite de la réalisation de la tâche d'interaction. Il joue le rôle de complément afin d'élargir l'analyse des données principales ou confirmer des hypothèses préalables.

La contrainte principale concerne l'élaboration du test. Le choix des items et des échelles est extrêmement important et doit être minutieusement préparé et testé, ce qui est chronophage. En effet, répondre à une échelle peut être abstrait pour certains locuteurs, les items pouvant

être assez ressemblants, notamment dans les valeurs intermédiaires. La réflexion sur ce qui est demandé et la fatigue cognitive que le choix de réponse peut engendrer sont donc ici très importantes. Enfin, on peut reprocher à ce protocole, à l'instar des autres protocoles non-interactifs, de manquer d'authenticité et de ne pas tester la compétence en situation réelle de communication. Ce sont notamment pour ces raisons que nous avons écarté ce protocole de notre étude. En outre, il nous a semblé que cette approche par échelle manquait de précision pour évaluer la compétence métapragmatique, ce pour quoi nous lui avons préféré un protocole d'entretien métapragmatique.

1.5. Verbalisation de la pensée à voix haute (*Think Aloud Protocol*)

Ce protocole consiste à demander au LNN de verbaliser pendant ou après la réalisation de l'activité les processus mentaux qu'il mobilise lors de l'interprétation des formes linguistiques qui lui sont proposées (Rosa et Leow 2004). Les LNN peuvent verbaliser tout ce qui leur traverse l'esprit lors de la réalisation de la tâche. Le protocole se déroule dans la langue dans laquelle le LNN met en œuvre la tâche et réalise son processus cognitif, et implique que le chercheur puisse lui-même comprendre la ou les langues mobilisées lors de la verbalisation. Le but est d'obtenir des informations sur le type de raisonnement sous-jacent à la production et la compréhension des phénomènes pragmalinguistiques et sociopragmatiques. Le chercheur a ainsi accès aux processus mentaux du LNN, ce qui ne serait pas possible dans des protocoles classiques ou en conditions réelles d'interaction.

Parmi les points faibles les plus souvent observés, nous pouvons constater que la mise en œuvre de ce type de protocole est complexe et nécessite la mise en place de conditions expérimentales très strictes et couteuses en ressources matérielles. Un des reproches souvent formulés est que le locuteur n'est pas en mesure d'exprimer en temps réel tous ses processus mentaux, ce qui pose de fait un problème de fiabilité et de précision de la verbalisation. En outre, la moindre sollicitation externe du chercheur lors du processus de verbalisation peut altérer le processus et biaiser les données. Le problème de la mémoire est souvent abordé, car plus la tâche à réaliser est longue, plus le locuteur ne sera plus en mesure de stocker dans sa mémoire toutes les informations qu'il souhaite verbaliser par la suite. La tâche est lourde sur le plan cognitif et nécessite de l'entrainement pour ceux qui ne sont pas habitués à ce type d'exercice. La langue de verbalisation et la difficulté de la tâche peuvent substantiellement altérer la qualité des données.

Ce protocole a été écarté de notre étude en raison de la très grande complexité de sa mise en place et de sa faible compatibilité avec le format de notre expérimentation. En effet, ce type de protocole semble particulièrement adapté à une approche qualitative menée sur quelques locuteurs seulement. Au-delà d'une dizaine de locuteurs, l'approche nous semble extrêmement complexe à mettre en œuvre. En outre, comme nous le verrons dans le chapitre de la partie analyse dédié à la méthodologie du test lexicopragmatique, le test étant passé à distance pour les LNN, il nous a été impossible de proposer des conditions d'expérimentation suffisamment contrôlées pour mettre en place un protocole de ce type.

2. Les protocoles interactifs

Les protocoles interactifs sont désignés ici comme des protocoles réalisés par l'informateur en interaction directe avec ses pairs ou avec le chercheur. Ils ont pour fonction principale d'évaluer la production ou la compréhension de manière dynamique et articulée dans une interaction. L'avantage premier qu'offre cette catégorie est la possibilité d'observer les différents éléments externes qui interviennent de manière simultanée à l'interaction. Ce type de protocole a principalement été mobilisé dans la première phase diagnostique de notre protocole expérimental. Nous voulions en effet tester 3 aspects de la compétence lexicopragmatique : 1) la capacité des LNN de niveau intermédiaire à produire un PhPex pertinent du point de vue de la situation de communication, 2) observer concrètement les stratégies de gestion des fonctions communicatives chez ces LNN et 3) observer la capacité de ces LNN à interagir avec les LN tout en adaptant leur manière de communiquer. Ce type de protocole nous a aussi permis d'observer concrètement les stratégies d'appui sur la langue maternelle en situation de communication réelle. Le deuxième avantage est la possibilité d'interaction directe entre le LNN et le chercheur qu'offre ce type de protocole d'élicitation, notamment dans le cadre de la discussion sur les phénomènes interprétatifs de niveau métapragmatique. Toutefois, de par leur nature spontanée, ces protocoles offrent un faible contrôle des variables à observer. Les données obtenues à l'issue de ces protocoles sont difficiles à standardiser et nécessitent une annotation qui peut être couteuse en temps et en ressource. Ce faisant, ce type de données ne permettent pas tout le temps le même type de traitement statistique ou quantitatif qu'avec les protocoles non-interactifs. Elles sont en

revanche utiles pour l'analyse qualitative de microphénomènes ou pour l'observation des stratégies de compréhension ou de sélection des formes adéquates.

2.1. Jeux de rôle

Le jeu de rôle (désormais JDR) consiste à faire participer les LNN à une simulation de situation d'interaction en attribuant à chacun un rôle social spécifique qui conditionnera la nature de leur production orale ou la fonction communicative attendue. Les JDR sont généralement vus comme de bons compromis entre un protocole standardisé avec un contrôle sur les biais possibles et une certaine authenticité des données. Deux types de configuration sont observables (Kasper et Dahl 1991) : les JDR avec canevas qui vont détailler la trame de l'interaction, le contexte et le type de réponse envisagé, et les JDR « ouverts » avec seulement un contexte général, laissant la liberté aux LNN de se positionner dans l'interaction. Les JDR classiques se déroulent en dyade avec soit un LNN, soit un LN (qui peut être un assistant ou l'enseignant lui-même). Les versions plus élaborées à plus de trois participants peuvent impliquer également des LNN ou des LN. Les développements récents des JDR ont vu l'apparition de pratiques fondées sur l'usage de plateforme de simulation en ligne dans lesquelles les LNN incarnent un avatar qu'ils peuvent déplacer dans un lieu virtuel pour aller à la rencontre d'autres LNN et interagir. En plus d'ajouter une dimension ludique, ces nouvelles approches permettent de contourner l'angoisse de la prise de parole ou la peur de se tromper en diminuant la pression interactionnelle ressentie puisque le locuteur joue généralement un rôle qui n'est pas le sien.

L'avantage premier de ce protocole d'exploration est la possibilité de simuler des interactions en condition réelle d'échange de tours de parole. Le but est de recréer les conditions d'interaction en contexte authentique de communication. Pour le chercheur, ce protocole donne la possibilité d'obtenir des données orales dynamiques permettant d'observer la production du LNN à la fois aux niveaux micro (choix lexicopragmatique, intonation, gestuelle) et macro (articulation de la prise de parole avec les autres participants, stratégies déployées, gestion des relations interpersonnelles, etc.). De plus, les connaissances linguistiques développées au cours du protocole ont de fortes chances d'être réutilisables par le LNN.

Un des désavantages est notamment lié au caractère artificiel et préparé des données. Le contexte n'est jamais vraiment authentique ni ne présente d'enjeux communicatifs particuliers pour le LNN comme ce serait le cas dans une interaction naturelle. En outre, la situation de communication proposée peut parfois s'éloigner grandement des conditions d'interaction auxquelles le LNN est amené à être confronté. Se pose alors la question de quel poids donner aux résultats ? Sont-ils révélateurs de la compétence pragmatique réelle du LNN ? Dans une certaine mesure, on pourrait même se demander si les résultats reflètent les choix propres du LNN où s'ils ne reflètent pas une interprétation du LNN de ce qui serait idéalement attendu dans le contexte et le rôle qui lui sont proposés. Ces rôles peuvent alors ne pas correspondre à l'expérience réelle de locuteur en langue cible du LNN. De même, les données obtenues avec un même protocole ont de grandes chances d'être différentes, ce qui là aussi ne permet pas une pleine comparabilité des données. Enfin, plus les consignes proposées et le guidage sont nombreux, moins la conversation qui en résulte sera susceptible d'être naturelle.

Nous avons écarté ce protocole pour plusieurs raisons. La première raison concerne la nature de notre objet d'étude. En effet, les PhPex sont des expressions de la vie quotidienne qui expriment l'affect du locuteur dans une grande variété de contextes. Leur usage n'est donc pas conditionné par un contexte particulier comme le seraient par exemple les pragmatèmes (Mel'čuk 2015). De même, chaque locuteur réagit d'une manière qui lui est propre et qui n'est pas nécessairement un PhPex. Cela peut être par exemple des réactions vocales, des interjections ou l'usage d'un contenu linguistique qui n'est pas préfabriqué. Il est donc difficile d'établir un prototype de situation impliquant l'emploi d'un PhPex en particulier comme on pourrait le faire avec un énoncé rituel de salutation ou avec un acte de langage bien défini (par ex. la requête). En outre, le caractère préparé d'un JDR rend plus difficile l'usage spontané et « sincère » d'une réaction émotionnelle. Certains informateurs pourraient trouver cela ridicule ou enfantin et ne pas jouer le jeu. La deuxième raison concerne le fait que ce type de protocole était très difficile à mettre en place dans le cadre de notre étude. En effet, nos informateurs non natifs étaient dispersés entre la France et le Japon et étaient tous soumis aux aléas de leur activité professionnelle ou académique. Il nous était donc impossible de les réunir pour effectuer un JDR ou de mettre en œuvre ce protocole à distance. La troisième raison est qu'au moment de l'élaboration de notre expérimentation, nous souhaitions observer les stratégies d'usage lexicopragmatique en contexte faiblement

élicité de manière à obtenir des données les plus authentiques et représentatives de la variété non-native des LNN.

2.2. Enregistrements non-élicités / semi-élicités

Un enregistrement élicité renvoie à tout type de conversations provoquées et enregistrées dans le cadre d'une collecte de données (Kasper et Rose 2002 : 84). Ce protocole est très complexe à mettre en place, mais offre la plus grande authenticité pour l'analyse des stratégies communicatives mobilisées par le LNN et son interlocuteur. Il s'agit d'un recueil effectué en l'absence du chercheur, celui-ci n'a peu ou pas d'influence sur ce qui est dit. Le LNN impliqué dans le recueil assure par lui-même l'enregistrement des données à l'aide d'un micro ou d'une caméra. Le protocole se veut le moins intrusif possible pour rendre la situation de communication authentique. Les seuls éléments pouvant influencer la production des locuteurs sont le matériel d'enregistrement et le fait de savoir que leur interaction sera étudiée, ce qui peut parfois déclencher des phénomènes d'auto-restriction ou de sur-correction de la production. Dans le cas de l'enregistrement de l'interaction semi-élicitée, un support déclencheur est proposé aux interactants (sujet au choix, une vidéo, un article de journal, une image, etc.), mais aucune autre intervention du chercheur n'est possible. Ce protocole diffère grandement du JDR dans la mesure où le LNN n'est pas amené à s'imaginer dans le rôle d'un locuteur tiers, il communique lui-même et occupe le rôle qu'il a habituellement en contexte de conversation naturelle.

Ce type de recueil a le mérite de permettre d'observer un large spectre de constructions discursives et pragmatiques, ainsi que les phénomènes de co-construction de l'interaction et de la relation interpersonnelle avec tous les imprévus qu'implique une interaction non préparée et négociée au fil de l'eau. Par ailleurs, la spontanéité qui s'installe rapidement dans ce type de protocole donne lieu à des comportements communicatifs non contrôlés ni anticipés fort intéressants pour la description de la variété orale non native en langue cible. L'observation de la créativité linguistique, ainsi que les stratégies de compréhension et de production mise en place par les LNN, apparait plus pertinente, car elle se situe dans une situation de communication spontanée, mettant en avant la capacité d'adaptation et d'anticipation des besoins communicatifs. En outre, ce type de protocole permet d'observer la réalisation de l'acte illocutoire et ses effets perlocutoires sur l'interlocuteur. Ce faisant, l'accès à ce type de données permet également d'observer les échecs de réalisation de l'acte

communicatif et la stratégie mobilisée pour y remédier. L'observation porte sur des données extrêmement proches de ce que produirait le locuteur en situation naturelle de communication.

Toutefois, ce type de protocole est plutôt complexe à mettre en œuvre en raison de ses nombreux points faibles méthodologiques. En premier lieu, ce type de recueil est bien souvent chronophage et couteux à mettre en œuvre, il faut trouver le lieu, les personnes qui correspondent aux critères de la recherche et mettre en place un matériel d'enregistrement. Parfois, il peut arriver que les données soient volumineuses, mais avec peu d'occurrences du phénomène étudié. C'est notamment un problème que nous avons rencontré dans la première phase diagnostique de notre protocole expérimental. Il apparait donc parfois difficile de mener ce type de protocole en ayant une hypothèse de recherche sur un phénomène trop précis. En outre, un autre problème est l'impossibilité de contrôler les variables ou de limiter les biais, car les données sont imprévisibles et dépendent de la dynamique de l'interaction. On peut renforcer l'élicitation afin d'augmenter les chances de production du phénomène observé, mais on perd d'un autre côté le caractère authentique.

De surcroit, le traitement des données est long, et on observe parfois un foisonnement d'informations si bien qu'il faut au préalable procéder à un tri. L'annotation et l'analyse sont elles aussi longues et nécessitent d'avoir recours à des procédures de validation interannotateur. Les contextes d'interaction sont parfois difficilement reproductibles, ce qui implique des questionnements quant à la comparabilité des données sources avec celles d'un corpus contrôle. L'absence du chercheur implique un risque de mauvaise qualité des données et ne lui permet pas de sélectionner comme il le désire ce qu'il souhaite recueillir. Le recueil est la simple discrétion du LNN et on ne peut pas vraiment obtenir toutes les informations sur les conditions de recueil ni sur les évènements se déroulant à ce moment-là. Cela pose également un problème éthique lorsque l'enregistrement est fait de manière dissimulée. Il faut donc prévoir des formalités juridiques lourdes pour s'assurer du caractère éthique et de la diffusibilité des données.

Dans le cadre de cette thèse, nous avons choisi cette méthode d'élicitation pour la première phase diagnostique de notre protocole expérimental. Pour les raisons citées plus haut, nous avons toutefois adapté ce protocole sous la forme de l'interaction libre semi-élicitée. Cette variante faiblement élicitée apparait comme un juste compromis permettant de limiter au

mieux les facteurs aléatoires et d'augmenter les chances d'observer le phénomène ciblé. De plus, ce protocole nous est apparu tout à fait approprié pour observer la compétence lexicopragmatique des LNN en milieu écologique et proposer ensuite un diagnostic des difficultés. Comme nous le verrons dans la partie méthodologique dédiée au diagnostic, l'élicitation était minimale de manière à reproduire au maximum les conditions optimales d'une interaction exolingue habituelle.

2.3. Entretien métapragmatique

L'entretien métapragmatique (Kinginger et Blattner 2008) consiste en une discussion guidée par le chercheur sur le lien entre formes linguistiques actionnelles, cadre sociopragmatique et organisation de l'interaction. Il vise à obtenir des traces observables du raisonnement et des processus mentaux qui sous-tendent le choix d'une forme pragmatique ou un comportement interactionnel particulier. Cet échange ne peut se déclencher qu'après une phase de conscientisation chez l'informateur qui sera généralement suscitée par le chercheur. Ce type de protocole exploratoire implique un niveau de réflexion plus intense, le locuteur se questionne sur le fonctionnement des mécanismes observé sur lesquels il émet des hypothèses qu'il soumet à la discussion. Cette discussion peut être générée soit par la discussion entre pairs, soit par la discussion étayée entre tantôt l'enseignant ou le chercheur et le LNN, tantôt le LN et le LNN. Le déclencheur peut être une situation d'interaction, un dialogue ou une expression en particulier. Le principe est de demander aux LNN d'expliciter et de justifier au maximum leur choix d'analyse, de préciser les indices de contextualisation ou les éléments du contexte employés pour fonder leur jugement. L'entretien métapragmatique peut aussi être employé en complément d'une activité d'élicitation interactive ou non-interactive (Van Campernolle and Henery 2014). Le chercheur doit alors coordonner ses questions de manière à guider le LNN dans sa réflexion. Le format de l'entretien métapragmatique est très cadré, les questions sont préparées à l'avance et sont élaborées en fonction des hypothèses du chercheur, de manière à amener l'informateur à dévoiler ses connaissances et ses stratégies face au stimulus qui lui est proposé.

Les inconvénients de la procédure sont multiples. Tout d'abord, le protocole produit généralement des données de faible comparabilité et ne permet pas d'analyse statistique. Par ailleurs, étant donné la nature complexe de la notion de conscientisation, il apparait difficile de mesurer ou d'évaluer quantitativement le niveau de conscientisation ou la capacité à

modéliser des phénomènes pragmatiques chez le LNN. Ce type de protocole est également très difficile à calibrer sans trop guider le processus de réflexion du LNN. En outre, on peut observer un phénomène d'anticipation des réponses attendues par le chercheur ce qui là aussi peut constituer un biais. Comme tous les types de protocoles autoréflexifs, on ne peut pas savoir si l'informateur dévoile tout ou dit la vérité. La perception qu'a le locuteur des phénomènes est très subjective et difficilement généralisable. Il faut donc prendre des pincettes lors de l'analyse. Enfin, il s'agit d'un protocole long à mettre en place, nécessitant beaucoup de préparation en amont pour le guide d'entretien et dont la transcription et l'analyse des données sont très chronophages.

L'avantage de ce protocole est de pouvoir développer chez l'informateur une attitude engagée et active dans la compréhension du sens pragmatique et des mécanismes de la conversation. En outre, le protocole permet au chercheur d'explorer les processus interprétatifs conscients du locuteur et de pouvoir observer les stratégies de reconstitution et/ou d'ajustement de la fonction s'appuyant sur la prise en compte du contexte d'énonciation. Kinginger et Blattner (2008) ont également observé que ce type de protocole permettait d'observer comment l'identité du locuteur peut influencer le choix des formes pragmatiques employées, notamment afin de favoriser ou non la connivence avec son interlocuteur. L'entretien métapragmatique, s'il est bien cadré, constitue un protocole empirique et rigoureux pour explorer chez le LNN les phénomènes de conscientisation et de niveau méta liés à l'usage de structures à fonction pragmatique. C'est notamment pour ces diverses raisons que nous avons choisi d'employer cette méthodologie d'élicitation dans la troisième phase de notre protocole expérimental. Nous souhaitions en effet explorer 3 angles d'analyse : 1) les stratégies mobilisées par le LNN dans la sélection et la compréhension des PhPex, 2) les difficultés d'usage rencontrées lors du test lexicopragmatique du point de vue du LNN, 3) le rôle de LM dans la compréhension ou la production et son influence sur la compétence lexicopragmatique. Bien que complexe à mettre en place et à traiter lors de l'analyse, nous verrons dans la partie dédiée aux entretiens post-test que cette méthodologie d'élicitation a constitué un apport intéressant dans la compréhension de l'usage et de la compréhension des PhPex chez les LNN.

Au cours de ce dernier chapitre consacré aux méthodes d'élicitation et d'évaluation de la compétence pragmatique, nous avons pu faire le point sur les différents protocoles qui ont été mobilisés dans notre démarche expérimentale. L'un des points forts de notre travail de

recherche est qu'il combine de manière complémentaire les approches interactives qui sont particulièrement utiles pour le diagnostic de la compétence lexicopragmatique et les approches non interactives qui offrent un cadre expérimental standardisé et contrôlé propice au test de la compétence et à la mesure de certains paramètres tels que le taux d'usage des séquences préfabriquées ou l'indice de variabilité lexicale des réponses enregistrées. En outre, l'articulation de ces différents protocoles nous permet d'observer la compétence lexicopragmatique en compréhension et en production. L'usage de l'entretien nous a permis d'avoir un retour du LNN sur les stratégies mobilisées lors du test et sur les difficultés particulières éprouvées lors de l'usage des PhPex. L'analyse de la compétence pragmatique mobilisée dans l'usage des PhPex relève de nombreux paramètres et de plusieurs niveaux de compétence. C'est en raison de cette complexité inhérente à notre sujet de recherche que nous avons mobilisé ce protocole en trois étapes et que nous avons articulé ces différentes méthodes d'élicitation. Nous allons dans ce qui suit proposer une synthèse qui permettra de faire le lien entre la partie du cadre théorique dédiée aux phraséologismes pragmatiques et celle dédiée à l'acquisition de la compétence pragmatique.

Synthèse du cadre théorique

Le cadre théorique qui a été présenté se situe à l'interface de deux disciplines : la phraséologie et l'acquisition de la L2. Dans un premier temps, nous avons précisé le cadre relatif à la perspective phraséologique de notre travail de thèse en trois étapes. Nous avons premièrement évoqué le contexte historique et épistémologique dans lequel se sont développés les champs de la phraséologie en France et au Japon. Nous souhaitions montrer que le travail effectué dans cette thèse répond également à un besoin méthodologique et théorique émanant du domaine de la phraséologie japonaise, domaine encore en friche. Nous avons également choisi d'inclure cette comparaison avec le Japon, car nous pensons que l'élargissement du champ de la phraséologie à d'autres langues, d'autant plus si elles sont typologiquement éloignées, constitue un enjeu important sur les plans épistémologique, typologique, acquisitionnel et didactique. Nous avons ensuite défini le champ de la phraséologie pragmatique et notamment d'un sous-domaine qui s'intéresse aux phraséologismes pragmatiques employés spécifiquement dans l'interaction orale ou médiée, et qui constitue le périmètre empirique de notre étude. Enfin, nous avons défini les contours notionnels de notre objet d'étude, les *phraséologismes pragmatiques à fonction expressive*, objet appartenant à la classe des phraséologismes des interactions.

Dans un second temps nous avons défini les orientations théoriques et méthodologiques concernant la deuxième discipline qui sera mobilisée dans ce travail de thèse, l'acquisition de la L2 et plus précisément l'acquisition de la compétence pragmatique. Ce second temps s'est divisé en quatre étapes. La première étape a constitué un tour d'horizon théorique de l'émergence des questionnements en acquisition de la L2 sur la compétence pragmatique. La seconde étape a précisé les contours théoriques du champ de la pragmatique de la langue seconde. Dans une troisième étape, nous avons détaillé les principaux modèles de la compétence pragmatique qui ont contribué à façonner notre angle d'analyse : la compétence lexicopragmatique. Enfin, dans une quatrième étape, nous avons dressé un inventaire des différents protocoles d'élicitation de la compétence pragmatique des LNN. Cet inventaire nous a permis de situer nos choix méthodologiques quant à la conception de notre protocole expérimental.

Que ce soit dans la partie phraséologie ou acquisition de ce travail de synthèse théorique, il nous a semblé important d'introduire brièvement le contexte épistémologique dans lequel se sont développées les notions qui seront utilisées dans l'analyse. En effet, l'analyse de la compétence mobilisée dans l'emploi des PhPex chez les LNN est un domaine encore en friche qui soulève beaucoup de questionnements sur les plans méthodologique et théorique. Ces questionnements sont notamment importants pour comprendre les difficultés que nous avons rencontrées dans la mise en place de l'expérimentation et lors de l'analyse des données. Par ailleurs, il nous a semblé qu'aucune synthèse théorique bi-disciplinaire de ce type n'existait. Nous pensons donc que ce travail sera utile comme point de départ pour tout chercheur désireux de mener des recherches dans le champ de la phraséologie pragmatique des interactions quotidiennes et de l'acquisition de la compétence pragmatique.

Nous concluons ainsi cette première phase de cadrage théorique qui constitue la première partie de cette thèse. Dans la partie suivante, nous allons présenter plus en détail la conception et l'analyse des résultats des trois phases de notre expérimentation.

PARTIE 2
-
PROTOCOLE EXPÉRIMENTAL ET ANALYSE

Avant-propos

Cette étude porte sur la compétence lexicopragmatique mobilisée dans l'usage et la compréhension des PhPex chez des LNN japonophones de niveau avancé du français. Nous précisons ici que lorsque nous employons *locuteur non natif*, nous désignons des locuteurs en phase d'acquisition de la langue cible, qu'ils soient engagés dans un apprentissage en milieu institutionnel ou non. Le protocole qui sera détaillé dans cette deuxième partie de la thèse se divise en 3 étapes comme modélisé dans le schéma ci-dessous :

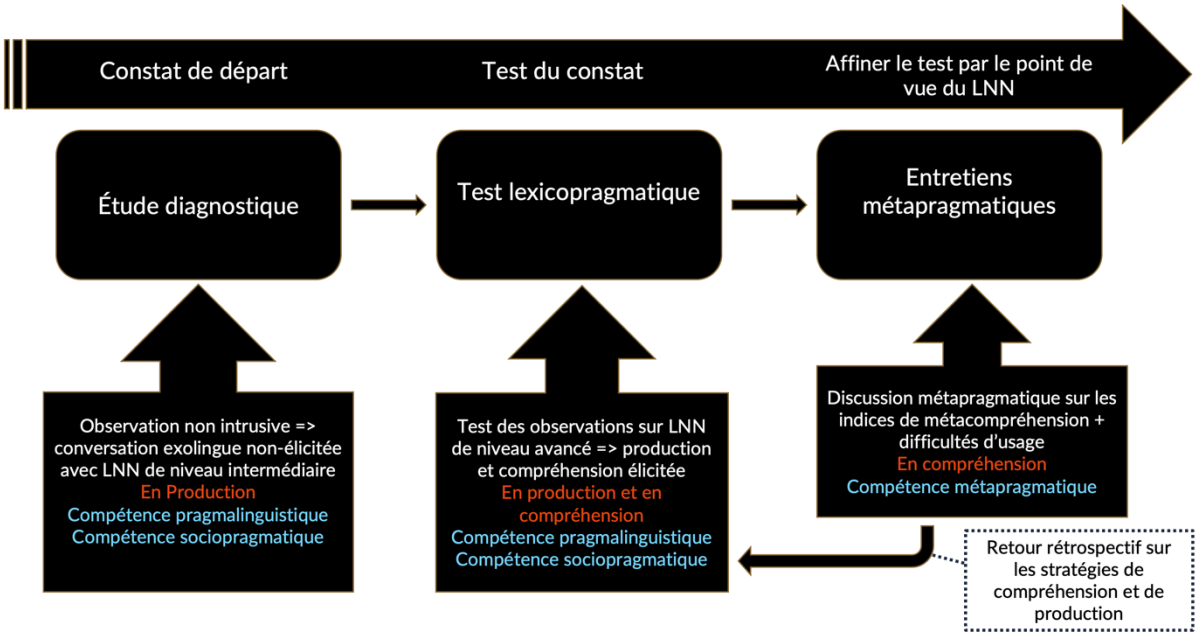


Figure 13 : Modélisation de notre protocole expérimental en trois phases

La première étape est un diagnostic. Cette étape vise notamment à proposer un premier état de la compétence lexicopragmatique des locuteurs japonophones apprenants du français de niveau intermédiaire CECRL A2/B1 en contexte d'interaction informelle et exolingue avec des locuteurs natifs du français. Elle vise également à confirmer l'hypothèse que même à un niveau intermédiaire où l'apprenant est censé disposer d'une compétence conversationnelle qui lui permet d'être autonome lors de la communication (cf. Tableau compétence conversationnelle CECRL, 2001), la maitrise des objets complexes que sont les PhPex (et plus globalement des phraséologismes pragmatiques) ne semble pas acquise de manière stable. Nous observons dans cette première partie du protocole des critères d'ordre lexicopragmatique (emplois des expressions idiomatiques et des marqueurs discursifs et interactionnels) et expressif (emploi des PhPex, des rires et des interjections) et décrivons

certains emplois inhabituels que nous retrouverons dans la production des apprenants japonais du français observée. Cette première étape nous permettra ainsi de constituer une base de comparaison pour observer si les difficultés rencontrées au niveau A2/B1 peuvent également concerner les niveaux très avancés.

La seconde étape est celle du test lexicopragmatique. Elle constitue l'étape principale et originale de notre protocole. Cette étape se situe dans la continuité de l'étude diagnostique et permet d'observer plus spécifiquement la compétence lexicopragmatique sollicitée dans la compréhension et l'emploi des PhPex chez des locuteurs japonophones avancés de niveau fin B2 à C2. Elle permet d'évaluer la capacité du locuteur non natif de niveau avancé à produire un PhPex approprié, à le comprendre ou à l'interpréter en fonction du contexte donné. Nous articulerons donc prioritairement notre réflexion sur les savoirs linguistiques et les savoirs procéduraux mobilisés dans l'emploi des PhPex. Une autre spécificité de ce test est sa dimension contrastive qui nous permet à la fois de repérer les différences entre les deux systèmes de langue à l'origine de difficultés d'emploi, mais aussi de comprendre les stratégies mises en place pour pallier le manque lexical.

La troisième étape est celle de la discussion métapragmatique. Alors que les deux premières étapes étaient axées principalement sur l'observation de la microstructure, le but de cette troisième phase est d'observer la méso- et la macrostructure de la compétence lexicopragmatique à l'aide d'un protocole basé en partie sur l'entretien métapragmatique. Nous partirons ici de l'hypothèse que le travail sur le *niveau métapragmatique* permet à l'observateur d'appréhender les mécanismes qui fondent l'analyse et le jugement permettant aux LNN de sélectionner la forme la plus appropriée en fonction des éléments linguistiques, sociolinguistiques et paralinguistiques à sa disposition. La notion de *discussion pragmatique* renvoie à un réel échange participatif entre le chercheur et le LNN dans le but d'étayer le fonctionnement d'un phénomène pragmatique désigné. Notre but ici est de proposer dans le cadre d'un entretien semi-directif plusieurs situations susceptibles d'activer chez le LNN une posture d'analyse et de mise en hypothèse de ses différentes observations concernant les expressions proposées dans le test lexicopragmatique.

Première phase du protocole expérimental :

Étude diagnostique de la compétence lexicopragmatique en français mobilisée lors de l'interaction orale informelle de locuteurs japonophones de niveau intermédiaire

Point de départ du protocole : le diagnostic

Le but de l'étude diagnostique discutée dans ce chapitre est de proposer un premier état de la compétence lexicopragmatique des locuteurs japonophones apprenants du français de niveau intermédiaire (CECRL A2/B1) en contexte d'interaction informelle et exolingue avec des locuteurs natifs du français. Cette étude diagnostique de la compétence lexicopragmatique et notamment de l'emploi des formules expressives des apprenants de niveau intermédiaire nous sert de première étape expérimentale pour mieux comprendre la compétence de production et de compréhension des PhPex qui sera étudiée dans la deuxième étape. Cette étape expérimentale vise à confirmer l'hypothèse que même à un niveau intermédiaire où l'apprenant est censé disposer d'une compétence conversationnelle qui lui permet d'être indépendant lors de la communication, et donc de pouvoir s'engager dans des interactions quotidiennes, la maitrise des objets complexes que sont les phraséologismes expressifs (et plus globalement des phraséologismes pragmatiques) ne semble pas acquise. Nous nous proposons donc d'observer quelques critères d'ordre lexicopragmatique (emplois des expressions idiomatiques et des marqueurs discursifs et interactionnels) et expressif (emploi des PhPex, des rires et des interjections) et de décrire certains emplois spécifiques que nous retrouverons dans la production des apprenants japonais du français observés. La validation de cette première hypothèse nous permettra d'introduire l'étude principale de cette thèse visant à observer les spécificités de la compétence de production et de compréhension des phraséologismes expressifs de la conversation des locuteurs japonais du français de niveau CECRL C1/C2. Notre objectif sera de montrer que les difficultés rencontrées au niveau A2/B1 peuvent également concerner les niveaux très avancés. Nous tenterons également de montrer qu'une compétence générale élevée en langue cible, même si elle constitue un facteur facilitateur, n'est pas forcément la condition unique pour la maitrise des phraséologismes expressifs et que d'autres paramètres indépendants du niveau d'expertise en langue cible peuvent intervenir dans l'emploi de ces structures. Dans une première partie, nous détaillons le cadrage théorique de cette étude diagnostique. Dans une seconde partie, nous évoquons les différentes étapes de la méthodologie employée pour recueillir, formaliser et annoter les corpus qui servent de base à notre analyse. Dans une troisième partie, nous procédons à l'analyse des résultats obtenus suite à l'annotation de notre corpus de production orale, puis nous concluons ce chapitre avec une synthèse qui pose le diagnostic de la compétence pragmatique des locuteurs non natifs observée dans cette première étape.

Chapitre 8 - Cadrage expérimental de la phase diagnostique

1. Questionnement de départ et description de l'étude

La compétence (lexico)pragmatique est une composante de l'acquisition des langues qui pose de nombreuses difficultés à l'apprenant de langue étrangère, même pour un étudiant de niveau avancé (B2/C1 CECRL) ayant une très bonne maitrise du standard et des aspects formels de la langue cible (Blum Kulka 1982 ; Eisenstein & Bodman 1986). Par ailleurs, la mesure et l'évaluation de cette compétence sont difficiles à opérer. En effet, les phénomènes linguistiques dont l'emploi requiert un certain degré de compétence pragmatique sont dépendants du contexte et les paramètres impliqués dans leur emploi sont multiples (Cavalla *et al.* 2020 : 12). Partant de ces constatations, nous avons tenté d'élaborer un dispositif diagnostic ciblé et non exhaustif qui permettrait d'observer de manière empirique et rigoureuse le degré de développement de la compétence lexicopragmatique des LNN japonophones du français de niveau CECRL B1/B2 en nous basant sur l'étude pragmatique de leur production orale dans contexte d'interaction exolingue[79] avec des locuteurs natifs.

Ce dispositif diagnostique se base sur deux enquêtes[80] réalisées en France et au Japon ayant débouché sur deux corpus d'interaction libre exolingue (LNN japonophones et LN francophones) et un corpus contrôle d'interaction libre endolingue (LN francophones uniquement). Ces trois corpus sont enregistrés en milieu semi-institutionnel[81]:

- Le corpus Exolingue-Grenoble (corpus test en milieu exolingue[82])
- Le corpus Exolingue-Tsukuba (corpus test en milieu exolingue)
- Le corpus Endolingue-Tsukuba (corpus contrôle endolingue)

---

[79] La notion d'exolingue renvoie à un cadre d'interaction où les interlocuteurs ne partagent pas la même langue maternelle. À l'inverse, la notion d'endolingue renvoie à un cadre d'interaction où les locuteurs partagent la même langue maternelle. (Porquier 2003, Bange 1992).
[80] Les données chiffrées concernant ces deux enquêtes sont explicitées dans le cadre méthodologique.
[81] Cadre universitaire, mais discussion libre et hors contrainte d'apprentissage ou de consigne liées à un cours de langue.
[82] Milieu homophone = utilisation du français en France / Milieu allophone = utilisation du français hors de France

Les deux premiers corpus nous servent d'observable sur lequel va porter notre analyse diagnostique. Nous observons notamment de quelle manière les locuteurs non natifs interagissent avec les locuteurs natifs, les spécificités de leur production orale sur le plan lexicopragmatique et structurel et l'apparition de sur ou sous-emplois particuliers (Granger 2015). Enfin nous tentons d'identifier d'éventuelles zones d'erreurs potentielles, notamment dans l'emploi des séquences pragmatiques à fonction expressive. Le corpus endolingue nous sert de corpus de contraste à partir duquel nous comparerons les productions des LNN aux productions de LN dans le cadre de conversations spontanées. Nous pourrons ainsi mettre en évidence d'éventuelles variations dans la fréquence d'emploi des phraséologismes pragmatiques ou des emplois particuliers que nous pourrons explorer de manière approfondie dans le corpus.

Afin d'analyser ces corpus, nous proposons un dispositif diagnostic qui nous permet de repérer des régularités dans les emplois lexicopragmatiques observés lors d'interactions exolingues informelles. Ce dispositif permet notamment de repérer des stratégies palliatives visant à combler les lacunes lexicales liées à une compétence lexicopragmatique encore instable. Ce dispositif diagnostic se compose d'une transcription orale de nos corpus et d'une typologie d'annotation de quelques marques spécifiques de l'interaction orale spontanée, ainsi que des marques d'anomalies[83] lexicopragmatiques retrouvées chez les LNN. La typologie d'annotation est implémentée par l'outil ELAN[84]. Au travers de ce dispositif, nous nous basons sur des traces concrètes de la variété native, des traces empiriques quantifiables et observables dans le matériel linguistique employé, ce qui nous permettra à la fois de faire un travail quantitatif puis qualitatif.

L'élaboration d'un dispositif diagnostique de ce type n'est pas chose simple en raison de la complexité et de la spontanéité du phénomène observé, ainsi que de la difficulté d'obtenir des données représentatives de qualité. En outre, les études sur la compétence pragmatique chez le locuteur non natif montre qu'il existe encore très peu d'études observant les compétences sollicitées dans l'emploi du lexique préfabriqué lors de l'interaction quotidienne : il n'existe pas à notre connaissance d'études similaires se focalisant sur la

---

[83] *Anomalie* n'étant pas ici perçu comme un terme péjoratif, mais comme un écart constaté dans la production des LNN à la norme observée dans la production native, écarts pouvant engendrer des échecs sociopragmatiques ou pragmalinguistiques.
[84] https://archive.mpi.nl/tla/elan

mesure diagnostique de la compétence lexicopragmatique en regard des marqueurs pragmatiques de la compétence conversationnelle et de l'analyse des emplois pragmatiques inhabituels dans la production de l'apprenant. Une autre difficulté se trouve au niveau de l'annotation de cette compétence lexicopragmatique, car il existe peu de propositions de typologie des phénomènes permettant la caractérisation de cette compétence. En outre, nous remarquons qu'il n'existe aucune typologie générale et opérationnelle de l'ensemble des marqueurs de la conversation quotidienne attestés (Traverso 2016, Noda 2011 : 32-33, Pellets 2005), ce qui a rendu difficile la constitution d'une grille d'annotation des phénomènes pragmatiques repérés. Quelques tentatives de taxinomies à visées didactiques ont vu le jour au cours des 20 dernières années (Chanet 2004, Paillard & Vu 2012, Lopes *et al.* 2015) sans pour donner pour le moment de résultats pleinement exploitables sur les plans linguistique et typologique. Par ailleurs, à notre connaissance, peu de littérature semble disponible sur l'évaluation ou la caractérisation de la compétence lexicopragmatique en production chez le locuteur non natif en situation de conversation quotidienne[85] exolingue. Il en va de même pour les protocoles de recueil de données ou d'élicitation de l'emploi des phraséologismes à fonction pragmatique. Enfin, le couple franco-japonais étant très peu doté en données empiriques visant l'étude de la compétence pragmatique (Cavalla et *al.* 2020 : 13), l'annotation des anomalies spécifiques à ce profil d'apprenant a été d'autant plus complexe et a nécessité de nombreux ajustements.

2. Le diagnostic de la compétence lexicopragmatique et la notion d'*erreur*

Notre étude diagnostique s'articule autour de la détection, de la caractérisation et de l'analyse des erreurs lexicopragmatiques des apprenants japonais du français de niveau intermédiaire. Nous partons de l'hypothèse suivante : la compétence lexicopragmatique ne dépend pas seulement du niveau d'expertise global du LNN en langue cible, mais dépend aussi de sa compétence à analyser et à systématiser les phénomènes extralinguistiques sous-jacents à l'emploi des formes spécifiques de la langue orale quotidienne (Taguchi 2011 : 285). Cette compétence est développée lors de l'exposition à la langue authentique (Barron 2019 : 44-45; Warga et Schölmberger 2007 : 243-244) et à l'aide du guidage en classe

[85] Il existe cependant une grande multitude de travaux portant sur la mesure et l'évaluation des compétences sollicitées lors d'actes de communications bien explorés comme la requête, les salutations, l'excuse, l'expression de l'avis, le refus ou les formules rituelles.

(Harada 2007, 2020). L'analyse des productions des LNN nous permet ainsi d'observer le niveau de développement et les spécificités de leur compétence lexicopragmatique au regard des différentes variations d'emploi et erreurs repérées dans les corpus.

Nous avons basé notre analyse diagnostique des données sur les outils méthodologiques développés dans le cadre de l'*analyse des erreurs* (Corder 1980 ; Perdue 1980 ; Marquilló Larruy 2003 ; Porquier 2020). Cette approche s'appuie sur l'analyse systématique des erreurs retrouvées dans les productions des locuteurs non natifs. Le locuteur non natif n'est pas « vierge » de connaissances linguistiques et procède par comparaison avec son savoir expérientiel acquis en pratiquant sa langue maternelle et/ou les autres langues de son répertoire linguistique. Ce savoir peut parfois entrer en interaction avec sa pratique de la langue cible, ce que Corder (1980) appelle *transfert*. Ces transferts constituent des traces empiriques nous montrant que l'erreur est principalement de nature systématique. Ils sont révélateurs des processus d'apprentissage de l'apprenant, de ses difficultés d'emploi, des zones de contacts entre les différentes langues de son répertoire langagier et de son engagement dans la conversation. Porquier (2020 : 23) dans la lignée de Marquillò Larruy (2003) distingue deux types d'erreurs en fonction de leur origine externe ou interne. Le premier type concerne les erreurs *interlinguales* (transferts et interférences) liées à la mobilisation du matériel linguistique dans la langue maternelle et la langue cible. Le second type renvoie aux erreurs *intralinguales* liées à la généralisation de règle d'emploi, à la non-maitrise d'autres règles du système en langue cible ou un emploi inadapté à la tâche de communication. Trévisiol-Okamura et Marquillo Larruy (2020 : 49) relèvent 4 niveaux dans l'erreur :

- L'adéquation au contexte communicatif (pragmatique)
- L'adéquation au système de la langue (grammaticalité)
- L'interprétabilité du message (sens / signification)
- Le bon usage (correct / incorrect)

Dans cette étude, nous nous concentrons essentiellement sur les erreurs interlinguales et intralinguales relevant du niveau pragmatique, c'est-à-dire sur l'adéquation de l'emploi des formes pragmatiques au contexte de communication. Nous formulons l'hypothèse que les habitus communicationnels développés dans la compétence lexicopragmatique de la langue maternelle influencent la compétence lexicopragmatique développée en langue cible et

servent d'appui à la compréhension et à l'acquisition des expressions idiomatiques de la conversation quotidienne. Dans la lignée de Marquilló Larruy (2003 : 10) et de Trévisiol-Okamura & Marquilló Laruy (2020 : 45), nous définirons ici l'erreur comme un écart entre une forme produite par l'apprenant et une forme attendue dans une situation de communication donnée. Il convient cependant de définir sur quel écart se base notre analyse de l'erreur. Dans le cadre de notre expérimentation, nous considérons comme anomalie tout écart entre la production des apprenants japonais A2/B1 CECRL du français qui constituera notre observable et la production spontanée des locuteurs natifs du français enregistrée dans une situation identique de communication qui constituera notre référence. Il s'agira donc de repérer des éléments de variation dans la production d'apprenant par rapport à ce qui serait habituellement produit par les natifs dans des conditions d'interaction similaires. Nous partirons ensuite de ces éléments de variation pour formuler nos hypothèses quant au diagnostic de la compétence lexicopragmatique des apprenants japonais du français de niveau intermédiaire. Notons que la démarche est originale dans la mesure où l'analyse des erreurs a surtout été employée pour analyser des erreurs aux niveaux syntaxique et phonétique, mais rarement dans le cadre de l'analyse de la compétence lexicopragmatique. Notre analyse se situe donc à l'interface du lexique et des paramètres pragmatiques qui président à l'emploi des phraséologismes pragmatiques (registre, relation interpersonnelle, enjeu de la conversation, axiologie, état émotionnel et psychologique, lieu, nombre de participants à l'interaction, etc.). Enfin, afin de limiter les biais de subjectivité liés à la tâche d'annotation, nous avons entrepris de valider notre schéma d'annotation par le biais d'un accord interannotateur qui sera détaillé dans une section ultérieure.

3. Méthodologie de l'étude diagnostique

Dans cette section, nous évoquerons les modalités de collecte du corpus ayant servi à l'étude diagnostique. Nous discuterons ensuite la formalisation et l'annotation des données. Nous aborderons notamment les spécificités de la transcription et de l'annotation de ces données, ainsi que le processus de validation du schéma d'annotation à l'aide d'une mesure d'accord interannotateur.

3.1. Le recueil des données

Afin de fonder notre diagnostic sur des emplois authentiques et spontanés, nous avons opté pour la mise en place d'une discussion libre semi-élicitée (cf. Chap. 7, p 127). Ce format a le mérite de limiter au maximum le guidage du propos tout en offrant un cadre de communication très proche d'une communication quotidienne. Aucune consigne particulière n'a été proposée aux participants, si ce n'est une liste de sujets de discussion facultative en français et en japonais proposés sous forme de cartes thématiques. Parmi ces sujets de conversation proposés aux informateurs, nous avons sélectionné des thématiques interculturelles telles que la nourriture, le système scolaire ou la politique. Ces thématiques font appel à la subjectivité des interlocuteurs et favorisent donc l'émergence de phraséologismes expressifs ou pragmatiques qui sont spécifiques à l'oral et, cela est bien connu, de marqueurs discursifs. Comme évoqué précédemment, les trois volets de ce corpus sont utilisés de manière complémentaire. Les deux premiers volets constituent des corpus de production orale authentique en situation de communication exolingue entre des locuteurs natifs et des locuteurs japonophones. Le troisième volet est un corpus d'interaction orale en milieu endolingue mettant en scène des locuteurs français en situation de discussion libre. Les deux premiers corpus constituent nos corpus diagnostiques que nous contrasterons avec le corpus endolingue, notre référence de contrôle. Ce faisant, nous pourrons ainsi observer d'une part les différences d'emploi, d'autre part les difficultés rencontrées par les locuteurs non natifs ou les phénomènes de sur ou sous-emploi par rapport au corpus de contrôle. Dans les deux sous-parties qui suivent, nous allons expliciter les dispositifs méthodologiques déployés dans le cadre de la collecte du corpus triple utilisé dans cette étude diagnostique.

3.1.1. Le corps Exolingue-Grenoble

La collecte du corpus Exolingue-Grenoble s'est déroulée au sein de l'Université Grenoble Alpes lors d'une session de discussion libre organisée entre des étudiants de la licence Langue Étrangère Appliquée (LEA) en anglais / japonais et des étudiantes japonaises en échange interuniversitaire. Les 2 locutrices non natives enregistrées sont de niveau B1 CECRL[86] certifié. Rappelons que notre objectif était de tester un niveau intermédiaire où l'apprenant est censé, selon le CECRL (2001), commencer à développer tout un panel de

[86] Les étudiants japonais en échange à l'Université de Grenoble doivent en principe attester d'un niveau en français B1/B2 ou valider la certification CLES B1 pour être éligibles au programme d'échange universitaire (https://langues-etrangeres.univ-grenoble-alpes.fr/fr/menu-principal/international/etudier-a-l-etranger/pays-partenaires-pour-l-echange-vers-l-international--45842.kjsp)

compétences d'ordre pragmatique, et donc pouvoir participer de manière plus moins active à une interaction orale. Mentionnons que les 2 locutrices ont été soumises essentiellement à un apprentissage en classe avec une faible proportion d'immersion à la langue française de la métropole et donc peu d'occasions de s'exposer aux spécificités de la conversation quotidienne. Par ailleurs, les locuteurs natifs étant des étudiants en LEA anglais/japonais, ils disposent d'une certaine expertise en japonais. Notons également que les deux locuteurs natifs se connaissaient, et qu'il en était de même pour les locutrices japonaises. En revanche, les locutrices japonaises n'avaient eu l'occasion de communiquer auparavant qu'avec le locuteur natif masculin et ne connaissaient donc pas la locutrice native. Les informateurs étaient installés autour d'une table rectangulaire par groupe de deux face à face. Nous avons aménagé l'espace de discussion pour créer une situation de discussion conviviale et facilitatrice de l'émergence des phénomènes pragmatiques qui nous intéressent. Afin de nous rapprocher au plus d'une conversation exolingue naturelle, nous avons également autorisé en cas de lacunes lexicales l'usage d'une langue tierce autre que le français. En effet, nous souhaitions également observer les stratégies de palliation en cas de manque lexical en langue cible, notamment avec le recours de la langue maternelle. La durée de l'interaction était de 45 minutes, mais les informateurs étaient tout à fait libres de poursuivre après cette limite s'ils le souhaitaient. Nous avons ainsi obtenu un corpus d'environ 70 minutes.

3.1.2. Le corpus Exolingue-Tsukuba

Le corpus Exolingue-Tsukuba a été recueilli à l'Université de Tsukuba (Japon), lors de sessions de discussion interculturelle dans un club universitaire de conversation franco-japonais nommé « Club Francophonie ». Le niveau de l'informateur est le niveau A2 obtenu à l'issue de la validation du parcours d'études françaises dispensé au CEGLOC [87]. Mentionnons que ce locuteur a un profil tout à fait particulier. En effet, bien que cet informateur ait appris le français en milieu institutionnel, celui-ci semble posséder une certaine aisance à l'oral. Les cours dispensés au CEGLOC ne dépassant pas le niveau A2, cet étudiant a poursuivi son apprentissage en autodidacte, ce qui lui a permis de progresser vers le niveau à l'oral jugé B1 au regard des descripteurs du CECRL. Il a également eu l'occasion de s'exposer de nombreuses fois au contexte de la conversation quotidienne lors

[87] *Center of Education of Global Communication* : centre de langues étrangères de l'université de Tsukuba.

du Club Francophonie où il participe très activement à la discussion avec les locuteurs natifs de français. Les locuteurs natifs de français quant à eux ne semblent pas posséder de compétences certifiées en japonais (les 4 sont issus d'un cursus en ingénierie), mais semblent avoir quelques rudiments en communication orale que l'on pourrait rapprocher du niveau CECRL A1 débutant. En outre, il est utile de noter que les locuteurs présents se connaissaient tous et avaient déjà eu l'occasion de communiquer ensemble auparavant. Concernant la configuration de l'espace, nous nous sommes basés sur le protocole Exolingue-Grenoble. La durée de l'interaction était de 60 minutes, mais les informateurs pouvaient poursuivre à leur guise après cette limite, permettant ainsi de recueillir un corpus d'environ 95 minutes.

3.1.3. Le corpus Endolingue-Tsukuba

Le corpus Endolingue-Tsukuba a été récolté au même moment que la constitution du corpus Exolingue-Tsukuba. Nous avons enregistré une discussion endolingue entre quatre locuteurs volontaires dans des conditions d'interaction libre non élicitée. Nous avons employé un protocole identique aux deux précédents corpus en proposant aux informateurs une discussion libre sur une thématique interculturelle liée à leur expérience au Japon. Les données prélevées dans le cadre de cette discussion nous servent de corpus contrôle que nous comparons aux données prélevées en contexte de conversation exolingue. Enfin, mentionnons que les locuteurs se connaissaient depuis quelque temps déjà et avaient eu l'occasion d'interagir ensemble par le passé. Le volume de données recueillies est d'environ 10 minutes. Voici un tableau récapitulatif des spécificités de nos trois sous-corpus :

	Identifiant	**Type de conversation**	**Contexte**	**Nombre d'extraits**	**Durée totale extraits**	**Durée totale enregistrement**	**Population étudiée**	**Date**	**Langue utilisée**
Corpus conversation apprenant 1 (Corpus d'observation)	Exolingue-Grenoble	Conversation libre non-élicitée	Conversation exolingue en milieu institutionnel, Grenoble, dans une salle de cours (gouter organisé)	5	11'45	69'52	2 locuteurs natifs du français 2 locuteurs japonophones apprenants du français	01/2019	Français / japonais / anglais
Corpus conversation apprenant 2 (Corpus d'observation)	Exolingue-Tsukuba	Conversation libre non-élicitée	Conversation exolingue dans un club de discussion, Tsukuba (Japon) dans le cadre Club Francophonie (club de discussion)	7	17'28	94'41	4 locuteurs natifs du français 1 locuteur japonophone apprenants du français	06/2019	Français / japonais / anglais

Corpus conversation natif (Corpus de contrôle)	Endolingue-Tsukuba	Conversation libre non-élicitée	Conversation exolingue dans un club de discussion, Tsukuba (Japon) dans le cadre Club Francophonie (club de discussion)	1	9'10	9'10	4 locuteurs natifs du français	06/2019	Français

Tableau 1: Métadonnées des différents corpus employés dans l'étude diagnostique

Et ci-dessous un tableau récapitulatif des données concernant les informateurs :

Identifiant[88]	Corpus	Âge	Spécialité[89]	Nationalité	Niveau FR	Niveau JP (estimé)	Autres langues connues	Séjour en France / Japon[90]	Relation entre les participants
LocFJP1	Exolingue-Grenoble	19-25 ans	SHS	Japon	B1 DELF	Natif	ENG	FR (échange et tourisme.)	Tous se connaissent sauf LocFFR qui ne connait pas les trois autres participants.
LocFJP2	Exolingue-Grenoble	19-25 ans	SHS	Japon	B1 DELF	Natif	ENG	FR (échange et tourisme)	
LocFFR	Exolingue-Grenoble	19-25 ans	SHS	France	Natif	A2	ENG	NC	
LocHFR1	Exolingue-Grenoble	19-25 ans	SHS	France	Natif	B1/B2	ENG	JP (Tourisme)	
LocHJP	Exolingue-Tsukuba	19-25 ans	SHS	Japon	A2 DELF B1 estimé	Natif	ENG	FR (échange et tourisme)	Tous se connaissent plus ou moins sauf l'observateur qui n'était présent que dans le cadre de l'étude.
LocHFR2	Exolingue-Tsukuba et Endolingue-Tsukuba	19-25 ans	SI	France	Natif	A1	ENG	JP (échange et Tourisme)	
LocFFR1	Exolingue-Tsukuba et Endolingue-Tsukuba	19-25 ans	SI	France	Natif	A1	ENG	JP (échange et Tourisme)	
LocFFR2	Exolingue-Tsukuba et Endolingue-Tsukuba	19-25 ans	SI	France	Natif	A1	ENG	JP (échange et Tourisme)	
Observateur	Exolingue-Tsukuba et Endolingue-Tsukuba	25-30 ans	SHS	France	Natif	C1	ENG	JP (échange, Tourisme et travail)	

Tableau 2 : Métadonnées des locuteurs ayant participé à l'étude diagnostique

3.1.4. Matériel de recueil

Afin de procéder à la collecte de nos données, nous avons privilégié un matériel de bonne qualité pour avoir des données fiables et facilement exploitables (ce qui est crucial pour la qualité de la transcription et l'annotation). Nous avons également veillé à ce que ce matériel ne soit pas trop visible par le locuteur pour ne pas influencer sa production[91] et obtenir des données les plus spontanées possibles. Nous avons opté pour un matériel de taille et de couleur discrète permettant une prise de vue à 360° : la caméra *Kodak Pix Pro SP 360*[92] et le micro *Zoom H2n*[93]. La caméra est nativement dotée d'un micro intégré permettant la

[88] Loc = locuteur / F = femme / H = homme / JP = Japonaise.e / FR = français.e
[89] SHS = Sciences humaines et sociales / SI = Sciences de l'ingénieur
[90] FR = France / JP = Japon / NC = Non connu
[91] Cf. *Paradoxe de l'observateur* (Labov 1973, Gadet 2007 : 38)
[92] https://kodakpixpro.com/docs/manuals/actioncam/sp360/sp360-manual-fr.pdf
[93] https://www.zoom.co.jp/sites/default/files/products/downloads/pdfs/F_H2n_0.pdf

captation des voix dans un périmètre restreint. Toutefois, ce micro intégré semble trouver ses limites lorsque les interlocuteurs sont trop nombreux ou éloignés. C'est pour cette raison que nous avons choisi d'utiliser un micro externe pour avoir une captation des voix plus fine et obtenir ainsi des données dont l'exploitation et la transcription seraient facilitées.

3.1.5. Cadrage juridique des données et anonymisation

Le recueil des données est soumis à un cadrage juridique strict avec signature préalable d'un consentement éclairé et anonymisation complète de toutes les données permettant de remonter à l'identité des informateurs (Baude, 2007). En fonction des demandes particulières exprimées dans les formulaires de consentements éclairés, certaines précautions supplémentaires ont été appliquées. Le consentement éclairé (Annexe 3 et 4) a été signé en deux exemplaires, un remis au chercheur et un autre remis à l'informateur. Il a été clairement signalé aux informateurs que leur exemplaire du consentement éclairé contenait toutes les informations importantes concernant le cadre et le déroulement de l'étude et leur permettait à tout moment de contacter le chercheur pour se retirer entièrement ou partiellement du corpus. De même, les informateurs ont été informés au préalable des risques potentiels que peuvent rencontrer leurs données une fois diffusées, et connaissaient donc tous les tenants et les aboutissants de la démarche de recherche au moment de la signature et du passage du protocole.

3.2. Transcription et annotation du corpus

Cette étape se divise en trois phases. Dans un premier temps, nous détaillons les modalités de transcription de notre corpus. Dans un second temps, nous expliquons la procédure d'annotation des transcriptions. Enfin, nous précisons de quelle manière nous avons validé et amélioré notre typologie à l'aide d'un test interannotateur.

3.2.1. La transcription des extraits

Les échantillons (Annexe 5) ont été annotés selon la charte de transcription ICOR (Groupe ICOR, UMR 5191 ICAR, CNRS Lyon 2 ENS de Lyon, 2013). Nous avons procédé à quelques simplifications de la charte de transcription afin de l'adapter aux besoins de notre étude et à l'outil utilisé pour notre transcription. Dans le cadre de ce travail de thèse, nous

utilisons une transcription orthographique adaptée (Gadet 1989 : 43-49) au plus proche de la représentation habituelle du français standard, car notre étude n'est pas centrée sur les spécificités phonétiques, morphologiques ou intonatives des emplois qui nous intéressent. Toutefois, dans le cas de certaines constructions morphologiques particulièrement productives à l'oral (structures élidées, amalgames), nous nous sommes réservé la possibilité de transcrire ces phénomènes dans une forme qui se rapproche plus de la langue parlée que de l'écrit. Il s'agit notamment de formes telles que *chais pas*, *ouais*, *'fin* et les doubles négations non réalisées. Ces phénomènes constituent des traces de la compétence pragmatique dans la conversation quotidienne et il nous a semblé utile de les garder telles quelles dans la transcription. Les allongements et les pauses (non mesurées dans notre étude) sont également transcrits. Ces phénomènes spécifiques à l'oral étaient intéressants pour l'analyse de nos données dans la mesure où ils peuvent être à la fois révélateurs de stratégies conversationnelles spécifiques (comme la cession ou non du tour de parole ou la reformulation) ou au contraire révélateurs d'anomalies conversationnelles liées à une compétence encore insuffisamment stabilisée en langue cible (longues hésitations, recherche lexicale). Lorsque les locuteurs non natifs emploient des termes ou des marqueurs de la conversation issus de leur langue maternelle, nous avons choisi de retranscrire ces formes en langue japonaise pour rendre plus saillante l'alternance entre les deux codes[94]. Pour ce qui est des marques de ponctuation, à l'instar des auteurs de la convention ICOR qui, s'appuyant sur les travaux de Blanche-Benveniste sur l'oral (1990), revendique la non-pertinence des marques de ponctuation à l'oral, nous n'utiliserons pas non plus de marques typographiques si ce n'est pour marquer des phénomènes comme les disfluences[95] ou les allongements[96].

Cette transcription a été réalisée à l'aide de l'outil ELAN[97]. Outre les aspects techniques, l'avantage d'ELAN réside également dans la manière dont les différentes couches de transcription ou d'annotations sont affichées. Chaque locuteur dispose de son propre canal de communication, ce qui rend particulièrement facile à annoter des phénomènes propres à l'interaction orale. En effet, la conversation quotidienne fonctionne peu sur la base d'une

---

[94] En cas de diffusion de corpus, nous procéderons à l'ajout de la transcription romanisée des alternances en langue japonaise pour faciliter la réutilisation de ces données chez des locuteurs non experts.
[95] Emploi du signe « - », par exemple « il a peu-peut-être trouvé une autre solution »
[96] Emploi du signe « : », par exemple « les carottes euh: je les ai mises dans le frigo »
[97] https://archive.mpi.nl/tla/elan

alternance stricte des prises de parole, mais plus généralement sur une négociation (Kerbrat-Orecchioni 2000, Traverso 2001) et/ou une co-construction (Kerbrat-Orecchioni 2005, Traverso 2007, André 2011) de l'interaction. Ainsi, il n'est pas rare d'observer des phénomènes de chevauchement ou de continuation[98] dans la conversation orale. Pour ce qui est de la segmentation des tours de paroles, nous avons estimé qu'un tour de parole correspondait à une prise de parole continue séparée soit par la prise de parole d'un autre interlocuteur, soit par une pause qui peut être courte ou longue (quelques secondes).

3.2.2. L'annotation des échantillons

Dans cette section, nous allons aborder le processus d'élaboration et d'amélioration de la typologie d'annotation. Ce processus a donné lieu à deux améliorations successives de la typologie originale. La 3ème version qui sera ici présentée est la version définitive qui sera appliquée à notre corpus. En prémices de l'élaboration de notre procédure d'annotation, nous avons d'abord entrepris un premier test sur un extrait de quelques minutes issu du corpus Exolingue-Grenoble. Nous avons entièrement transcrit cet extrait puis nous avons examiné de manière approfondie les productions orales des locuteurs non natifs. À l'aide de ce premier repérage empirique, nous avons dressé une typologie préliminaire des anomalies repérées dans la production orale des locuteurs japonophones du français comme indiqué dans le tableau 3 :

	Sous-catégorie	Définition sous-catégorie
Erreurs pragmatiques	Usage inapproprié du registre	Registre de langue et attitude interactionnelle inappropriés aux spécificités de la communication en cours.
	Expression inadaptée au contexte	Emploi d'une expression qui n'est pas adaptée à la situation de communication ou à la relation interpersonnelle entre le locuteur et l'interlocuteur.
	Expression pragmatique dévoyée / mal employée	Expression ou marqueur pragmatique employé de manière détournée comparé à l'emploi fréquemment admis chez les locuteurs natifs. Il s'agit notamment d'un emploi erratique lié à une mauvaise compréhension de normes d'usage chez le locuteur allophone.
Calques / transferts de	Transfert lexical	Transfert du lexique de la langue maternelle sans adaptation au système phonétique de la langue cible, afin de pallier une lacune lexicale.
	Transfert pragmatique	Transfert d'une fonction pragmatique ou conversationnelle depuis la langue maternelle afin de répondre à un besoin communicationnel auquel le locuteur allophone ne peut répondre par sa compétence en langue cible.

[98] Ce phénomène concerne les moments de la conversation où le locuteur finit la phrase de son interlocuteur qui est en train de parler.

	Transfert morphosyntaxique	Projection de la structure morphosyntaxique de la langue maternelle sur la langue cible, altérant ainsi la structure syntaxique et morphologique de la langue cible en production.
Erreurs systémiques / formelles	Erreur de grammaire	Erreurs de grammaire concernant notamment l'emploi des prépositions, des déterminants et des flexions verbales, nominales et adjectivales. Une faible compétence grammaticale est signe d'une communication souvent moins naturelle, ce qui influe sur la compétence pragmatique et sur la perception qu'a le locuteur natif de la compétence du locuteur non natif.
	Erreur de syntaxe	Erreurs commises au niveau de la structure syntaxique de production orale, notamment sur l'enchainement des constituants et des parties de l'énoncé.
	Erreur de prononciation	Toutes les erreurs concernant le rythme de parole, l'intonation, la prononciation et l'intensité. L'intonation joue un rôle crucial pour exprimer certaines fonctions pragmatiques et pour fluidifier l'interaction, la rendre plus spontanée.

Tableau 3 : Première version de la grille d'annotation basée sur une typologie préliminaire des erreurs observées chez les locuteurs japonais du français B1/B2

Le premier repérage que nous avons fait concerne principalement les anomalies pragmatiques que nous avons essayé de classer en trois catégories qui englobent :

- Des erreurs liées au registre de langue
- Des emplois acceptables, mais très inhabituels en conversation
- Des emplois erronés d'expressions ou de marqueurs spécifiques à l'oral

Nous avons ensuite repéré de nombreux phénomènes de transfert de la langue maternelle vers la langue cible. Ces transferts sont de trois types : lexicaux, pragmatiques et morphosyntaxiques. Par ailleurs, nous avons formulé l'hypothèse qu'une faible compétence en grammaire, en syntaxe ou en prononciation de la langue cible peut influer sur la compétence lexicopragmatique du LNN et sur la perception qu'a le LN de la compétence du LNN. Cette perception de la compétence du LNN peut avoir une influence sur la spontanéité et la complexité de la production du locuteur natif. C'est pour cela que nous avons choisi d'annoter les anomalies systématiques, car elles semblent révélatrices d'une compétence lexicopragmatique en cours de développement.

Après avoir établi une première typologie des erreurs de locuteurs japonophones de notre corpus, nous nous sommes aperçu qu'il était nécessaire d'annoter également les emplois pragmatiques dans les productions orales LN et des LNN pour pouvoir mettre en évidence des phénomènes de sur et de sous-emploi entre les deux groupes de locuteurs. Ce nouveau

volet de notre annotation nous permet d'observer à la fois les anomalies, mais aussi les emplois qui sont caractéristiques de la compétence pragmatique des locuteurs japonophones de niveau A2/B1 CECRL en français. Pour ce faire, nous avons sélectionné dans les deux corpus exolingues quelques courts extraits pertinents[99]. Ces extraits ont été sélectionnés à des moments où la conversation est dense et semble expressive. Nous avons ensuite observé les phénomènes pragmatiques les plus employés puis nous les avons classés en trois catégories : les *marqueurs expressifs*, les *marqueurs discursifs*, les *marqueurs de spontanéité orale* [100] . Nous avons ensuite créé une quatrième catégorie *marqueurs anomalie interactionnelle* dans laquelle nous avons inséré notre typologie des erreurs. Enfin, nous avons rédigé une nouvelle version du schéma d'annotation :

[99] Exolingue-Grenoble : 5 extraits pour 11min45 / Exolingue-Tsukuba : 7 extraits pour 17min28

[100] A ce stade du développement de la typologie, les catégories proposées sont encore en cours d'expérimentation et en attente de validation par l'intermédiaire d'un accord interannotateur. Ces catégories se fondent principalement sur les phénomènes linguistiques observés dans nos données. Nous avons ensuite étoffé notre typologie à la lumière de plusieurs travaux effectués dans le domaine des marqueurs discursifs. Cette étape sera abordée plus bas dans cette partie.

	Description	Sous-catégorie	Définition	Exemple
Marqueurs expressifs	Marqueurs pragmatiques présents à l'oral permettant au locuteur d'exprimer un état émotionnel et mental en réaction à un dire ou une situation à laquelle il est confronté.	PhPex	Phraséologisme à fonction expressive permettant d'exprimer un état affectif du locuteur.	*C'est chaud !* *Sérieux ?*
		Rire	Occurrence d'un rire qui peut avoir ici différentes fonctions comme le marquage de la connivence, l'empathie, l'amusement ou la gêne.	
		Réaction vocale	Ces marqueurs constituent des réactions vocales émises de manière à exprimer un affect. Dans le cas d'un locuteur non natif, ces réactions peuvent être un indice de stratégie palliative lorsque celui-ci ne possède pas la compétence lexicopragmatique suffisante pour exprimer son affect en langue cible.	Toutes manifestations vocales expressives jugées non naturelles en français.
		Interjection	Marque l'usage d'une interjection dans le discours parlé.	*Ah, oh, hé*
Marqueurs discursifs	Marqueurs pragmatiques de l'oral et statistiquement très présent dans le discours du locuteur natif. Ils sont généralement facultatifs, non essentiels à la construction du sens, mais occupent des fonctions discursives et interactionnelles diverses.	Filler	Marqueur permettant de combler entre deux tours de parole ou durant une pause ou permettant de marquer les temps d'hésitation.	*Euh, hm, hum*
		Articulateur	Marqueur permettant d'introduire ou d'articuler les différents éléments du discours du locuteur. Marqueur permettant également au locuteur d'introduire une précision un exemple ou un avis du locuteur vis-à-vis d'un élément de la conversation.	*Mais (euh), du coup, voilà, quoi, déjà, alors* *En fait, genre, justement*
		Régulateur phatique	Marqueur permettant de réguler l'attention ou d'ajuster l'implication du locuteur dans l'interaction en cours. Peut aussi être utilisé comme un moyen de montrer à son interlocuteur une incompréhension liée à la rupture du canal de communication.	*Hm hm, oui (oui oui), okay, d'accord, hein, quoi*
		Marqueur épistémique	Marqueur permettant au locuteur d'exprimer un degré de certitude, de modaliser son dire ou de moduler son assertion.	*Je trouve, je pense, pas tellement*
		Marqueur reformulation	Marqueur permettant d'exprimer une reformulation suite à une idée exprimée, mais jugée inappropriée ou non pertinente.	*Enfin ('fin),*
		Marqueur confirmation	Marqueur permettant au locuteur de confirmer ou valider ce que son interlocuteur dit.	*Ouais, oui, okay*

		Marqueur intersubjectif	Marqueur permettant au locuteur d'impliquer son interlocuteur en sollicitant sa compréhension ou son avis/impression sur le contenu énoncé.	*Hein, tu sais, t'as vu, tu vois*
		Marqueur balisage	Marqueur permettant de mettre en évidence un lieu, une temporalité ou un élément du discours.	*Ah bah là, bah ça alors*
Marqueur spontanéité orale[101]	Marqueurs spécifiques de la compétence orale dans la conversation de la vie de tous les jours. Il s'agit notamment de phénomènes très fréquents à l'oral et influençant à la fois le niveau segmental (morphologie lexicale, emplois lexicaux stéréotypés et familiers), le niveau suprasegmental (intonation, débit de parole, disfluence) et le niveau interactionnel (chevauchement, digressions thématiques, tours de parole endémiques)	Contraction	Concerne les contractions morphologiques opérées par les locuteurs natifs pour raccourcir leurs phrases lors de l'usage du registre familier, par exemple *chais pas* pour *je ne sais pas*.	*Je ne sais => chais pas* *Il y a => y'a*
		Chevauchement	Concerne toutes les prises de paroles d'un locuteur alors qu'un autre est en train de parler. Ce phénomène est très fréquemment observé à l'oral quotidien chez les locuteurs natifs du français et témoigne souvent d'une certaine maitrise des spécificités du français oral ou du registre familier chez les locuteurs non natifs.	
		Idiomatique/familier	Concerne tous les emplois de lexique ou de séquences lexicales stéréotypées tels que les locutions verbales ou les expressions familières fréquemment retrouvées à l'oral.	*tout va au poil, c'est pas la mort, pas folle la guêpe*
		Élision	Concerne les élisions fréquemment retrouvées à l'oral lors de la conversation	*Je ne sais pas => j'sais pas*
		Négation simple	Concerne les omissions de la double négation caractéristiques de la langue orale spontanée.	*Je ne sais pas => je sais pas*
Marqueur anomalie interactionnelle	Marqueurs spécifiques révélant un manque de compétence interactionnelle de l'apprenant dans la langue cible. Ces marqueurs apparaissent notamment sous la forme d'une sous-représentation des emplois productifs chez les natifs, emplois compensés par des stratégies palliatives ou d'évitement qui peuvent parfois	Transfert lex	Renvoie aux alternances codiques, c'est-à-dire aux transferts de lexique de la langue maternelle vers la langue cible pour allier un manque en langue cible.	
		Transfert ex	Transfert des marqueurs/réactions expressifs de la langue maternelle vers la langue cible.	
		Transfert MD	Renvoie au transfert de marqueurs discursifs de la langue maternelle vers la langue cible pour pallier un manque en langue cible.	

[101] Il ne s'agit pas seulement ici de marqueurs lexicaux, mais également d'autres marqueurs paraverbaux.

	engendrer certaines anomalies dans le schéma communicatif habituel observé entre des locuteurs natifs.	Erreur prag	Renvoie à l'usage pragmatique inapproprié ou inadapté à la situation où une autre réponse pragmatique est habituellement attendue.	
		Erreur form	Erreur syntaxique ou grammaticale et usage de structures bancales.	
		Erreur lex	Usage inapproprié d'un terme.	

Tableau 4 : Deuxième version de la grille d'annotation incluant à la précédente version des marques spécifiques de la conversation quotidienne

Avant de procéder à l'annotation finale de nos données, nous avons validé cette seconde typologie à l'aide d'un accord interannotateur à trois codeurs, de manière à éliminer d'éventuels descripteurs ambigus ou peu pertinents. En effet, l'annotation de ce type de phénomène linguistique est souvent sujette à une grande part d'interprétation. Elle dépend également du cadre théorique employé lors de sa conception. Pour éviter ce biais interprétatif, la typologie d'annotation doit être la plus documentée et la plus complète possible afin d'éviter d'éventuelles ambigüités. Un schéma d'annotation clair permettra ainsi d'avoir des résultats fiables. Nous présentons cette procédure dans la partie qui suit.

3.3. L'accord interannotateur et l'amélioration de l'annotation

3.3.1. L'accord interannotateur

L'accord interannotateur (désormais AIA) permet de valider de manière statistique l'annotation des données à l'aide de métriques développées pour mesurer la fiabilité de la tâche d'annotation effectuée par plusieurs annotateurs sur un même segment de corpus. L'AIA teste également la reproductibilité du schéma d'annotation en évaluant le taux d'accord entre les différents annotateurs lors d'une tâche d'annotation similaire. Un score d'accord faible, c'est-à-dire un score de fiabilité de l'annotation peu satisfaisant, implique alors un remaniement du schéma d'annotation. C'est pour cela qu'il apparait également pertinent de procéder à la mesure de l'AIA au début de la procédure d'annotation, mais également tout au long de celle-ci pour assurer un contrôle de la fiabilité. Dans le cadre de notre AIA, nous avons choisi la mesure *gamma* (Mathet, Widlöcher & Métivier 2015, 2017). En effet, elle apparait particulièrement adaptée à l'utilisation sur corpus oral transcrit puisqu'elle permet de mesurer la fiabilité d'une tâche d'annotation impliquant le repérage, la segmentation et l'étiquetage d'un phénomène linguistique par plusieurs annotateurs (Mathet *et al.* 2015 : 474).

3.3.2. Résultats de l'AIA pratiquée sur le schéma d'annotation

L'interprétation de l'indice obtenu à partir de la métrique dépend grandement de la tâche d'annotation effectuée et de son caractère subjectif et complexe. Il est souvent admis qu'un accord supérieur à 0.8 est une garantie satisfaisante de la fiabilité des annotations. Toutefois, plus la tâche d'annotation est de haut niveau, plus elle est sujette à interprétation diverse et

donc plus l'AIA sera lui aussi moins concordant (Bregeon et al., 2019 : 51). La valeur seuil de 0,8 n'est donc pas absolue et doit être interprétée avec précaution. Le score d'AIA ne suffit pas à lui seul pour valider la pertinence et la fiabilité des annotations du corpus, un examen contextualisé doit être systématiquement pratiqué pour évaluer pleinement la qualité et la consistance du schéma d'annotation en fonction des spécificités du corpus.

La mesure *gamma* appliquée à notre étude diagnostique a été mise en œuvre sur 3 annotateurs, dont nous faisions partie. Les trois annotateurs sont des linguistes ayant une certaine expérience de l'annotation sur corpus. En outre, les trois annotateurs sont familiarisés avec les phénomènes linguistiques fréquemment retrouvés dans la conversation orale et ont des notions de base en pragmatique et en analyse de l'interaction. Afin de réaliser l'annotation qui servira de support à la mesure de l'AIA, il a été distribué à chaque annotateur un échantillon sélectionné au hasard dans le corpus Exolingue-Tsukuba. Cet échantillon de 1min15 a été transcrit et préparé pour être annotable directement. La préparation a inclus le paramétrage du fichier ELAN avec toutes les valeurs de l'annotation. Enfin, les annotateurs ont reçu un guide d'annotation contenant le schéma d'annotation détaillé un peu plus haut et un court guide d'utilisation d'ELAN explicitant les diverses étapes pour la délimitation et l'insertion d'une annotation dans la transcription. L'extrait étant riche en phénomènes à annoter, l'annotation a pris environ une heure par personne. Le but premier de cette démarche était de vérifier la stabilité et la fiabilité de notre schéma d'annotation. Les résultats obtenus sont les suivants :

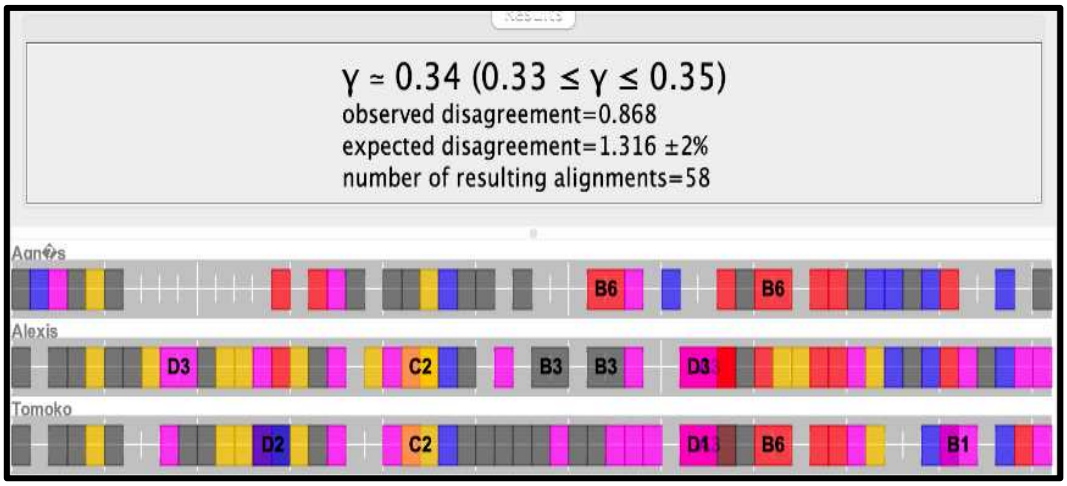


Figure 14 : Résultats de l'AIA mesuré avec la métrique gamma sur un échantillon d'annotation du corpus Exolingue-Tsukuba

Dans l'encadré contenant les données chiffrées, nous pouvons observer le score d'AIA qui correspond 0,34 avec une marge d'erreur comprise entre 0,33 et 0,35 (2%). Si l'on se réfère aux recommandations d'interprétation précédemment évoquées, on constate que l'AIA pour cette tâche d'annotation est faible. Cela est d'autant plus visible dans la partie basse de la figure où l'on peut observer une représentation graphique des annotations proposées par les trois annotateurs. Cette représentation graphique de la concordance des annotations et de leur alignement nous a permis de mettre en relief les éléments du schéma d'annotation étant à l'origine de la forte discordance constatée.

3.3.3. Amélioration du schéma d'annotation

Après avoir mesuré l'AIA de notre schéma d'annotation, nous avons entrepris un retour sur les données et une concertation avec les différents annotateurs afin d'isoler les valeurs du schéma d'annotation posant des difficultés ou étant non-pertinentes. Le premier problème rencontré est la catégorie des *articulateurs.* La catégorie manquait de délimitation par rapport à la catégorie des *marqueurs discursifs* ayant réellement une fonction logique et organisatrice dans le discours. Or, les objets linguistiques que nous souhaitions décrire dans la catégorie des *articulateurs* étaient plutôt des formes linguistiques ayant la particularité de connecter les divers éléments du discours pour fluidifier et lier l'interaction de manière naturelle, sans réellement impliquer de lien logique ou argumentatif. La deuxième difficulté rencontrée se trouvait au niveau de la différenciation entre la sous-catégorie des *régulateurs phatiques* et des *fillers*. En effet, nous nous sommes aperçu en consultant les annotateurs que les catégories étaient mal décrites, car on ne réussissait pas à faire une différence nette entre les formes employées dans les deux catégories. Nous avons donc entrepris de détailler plus finement ces deux catégories de manière à bien les différencier au niveau de leur fonction pragmatique et conversationnelle. Par ailleurs, la catégorie des marques de spontanéité orale semblait un peu confuse, car il ne s'agit pas de la caractérisation de la spontanéité orale, mais des spécificités du registre oral. Au sein de cette catégorie, il est apparu que l'annotation des doubles négations non réalisées et des élisions n'apportait pas grand-chose à notre démarche diagnostique. Nous avons donc choisi de les supprimer du schéma d'annotation. Enfin, nous avons approfondi les descriptions proposées pour chaque entrée afin de clairement exposer les spécificités de chaque objet. Nous avons dès que possible proposé un exemple précis et concret pour chaque phénomène linguistique annoté afin de faciliter l'identification et l'annotation de ces phénomènes.

Nous avons à la suite de cette procédure opéré un remaniement de notre schéma d'annotation en fonction des nouvelles données apportées par la consultation des annotateurs. De cette manière, nous avons pu écarter les valeurs d'annotation inadaptées, ainsi que les descriptions du schéma d'annotations jugées trop ambigües ou insuffisamment décrites. Ce remaniement de la grille d'annotation nous a permis d'élaborer la version finale de notre schéma d'annotation présentée ci-dessous :

	Sous-catégorie	Définition sous-catégorie	Exemple
Marques expressives	**PhPex**	Phraséologisme à fonction expressive permettant d'exprimer un état affectif du locuteur comme la surprise, la consternation, la colère ou l'incrédulité. Ces expressions sont généralement autonomes sur le plan syntaxique et sont utilisées de manière réactive à une situation particulière ou un dire de l'interlocuteur.	*Ex :* *-Tu as appris pour Paul ?* *- Oui, **ça craint !***
	Rire	Occurrence d'un rire qui peut avoir ici différentes fonctions comme le marquage de la connivence, l'empathie, l'amusement ou la gêne.	
	Réaction vocale	Ces marqueurs constituent des réactions vocales qui semblent habituelles en français et qui sont émises de manière à exprimer un affect. Dans le cas d'un locuteur non natif, le suremploi de ces réactions peut être un indice de stratégie palliative lorsque celui-ci ne possède pas la compétence lexicopragmatique suffisante pour exprimer son affect en langue cible. Dans ce dernier cas, si la forme expressive employée ne semble pas naturelle en français, annoter comme un « transfert expressif » (cf. Typologie anomalies plus bas).	Toutes manifestations vocales expressives similaires à celles employées en français
	Interjection	Marque l'usage d'une interjection dans le discours parlé. Ce sont généralement des réactions vocales, mais lexicalisées dans la langue française.	*Ex : **ah !** quelle bonne surprise de te voir ! ça fait un bail !*
Marques discursives et interactionnelles	**Filler**	Marqueur permettant de combler un silence entre deux tours de parole, une pause ou permettant de marquer les temps d'hésitation. Ces marqueurs sont une trace du travail de formulation (Candea 2000 ; Crible 2007, 2020) du locuteur. Le filler peut être employé comme structurant (pause volontaire rythmique ne marquant pas une recherche lexicale), comme appui (transition brève entre deux segments de parole) ou combiné avec un *connecteur* (ou *ligateur* chez Candea 2000) : *mais euh, donc euh, du coup euh*. Enfin, ce type de marqueur peut également avoir la fonction de continuateur en maintenant la prise de parole du locuteur (*idem* : 443).	*Ex : **euh…** je sais pas moi **hum**… vers 10h ça t'irait ?*
	Connecteurs	Marqueur permettant d'introduire ou d'articuler les différents éléments du discours du locuteur, sans pour autant marquer de relation logique (cf. *Marqueur de la structuration de la conversation*, Auchlin 1981). Marqueur permettant également au locuteur d'introduire une précision, un exemple ou un avis du locuteur vis-à-vis d'un élément de la conversation. Ils peuvent avoir plusieurs fonctions (Ferré 2011) : *iniative* (prendre la parole), *conclusive* (terminer le tour de parole), *résomptive* (reprendre un tour après une interruption brève ou une digression) et *ponctuante* (*turn construction unit*, unité de construction de la prise de parole).	*Ex : Je sais pas **quoi**… tu pourrais faire un effort quand même !* *Ex : **Du coup**, je vous propose de commencer la réunion sur l'étude des demandes de financement.*
	Marqueur phatique/régulateur	Marqueur permettant de capter l'attention ou d'ajuster l'implication / l'engagement du locuteur dans l'interaction en cours. Il peut aussi être utilisé comme un moyen de montrer à son interlocuteur une incompréhension liée à la rupture du canal de communication (Malinowsky 1923 ; Jakobson 1963 ; Kerbrat-Orecchioni 1990 ; Smith 2007 : 11). D'un point de vue distributionnel, ces marques sont dépendantes de la prise de parole de l'interlocuteur et accompagnent (ou se superposent) dans le discours de celui-ci. Ces marques sont généralement associées à des manifestations gestuelles stéréotypées telles que le hochement de tête.	*Ex :* *- Alors du coup je suis rentré sous la pluie, la gosse galère* ***- hm hm*** *- et c'est comme ça que j'ai choppé froid. Bref une sale journée*
	Marqueur épistémique	Marqueur permettant au locuteur d'exprimer un degré de certitude, de modaliser son dire ou de moduler son assertion (Tuchais 2015, 2020 ; Degoumois 2017). Le marqueur permet également au locuteur de se positionner face à ses connaissances et son savoir, d'apporter un éclairage personnel (Heritage 2012) et d'ajuster son domaine de connaissance (Kamio 1997) par rapport à son interlocuteur.	*Ex : **Je trouve** qu'il devrait être plus indulgent avec son frère.*
	Marqueur reformulation	Marqueur permettant d'exprimer une reformulation suite à une idée exprimée, mais jugée inappropriée ou non pertinente. Le marqueur peut aussi être employé afin de reformuler le dire du locuteur de manière à le rendre plus explicite ou à faciliter la compréhension de l'interlocuteur (Gülich 1983, Dufour 2007, Steuckardt 2018, Magri 2018).	*Ex : Mon frère abuse avec les jeux en ligne, '**fin je veux dire**, il passe sa vie dessus, c'est dingue !*

	Marqueur confirmation	Marqueur permettant au locuteur de confirmer ou valider ce que son interlocuteur dit. La réduplication peut éventuellement indiquer une intensification et un renforcement de la ratification (Lefeuvre et Legallois 2021 : 8). Cette sous-catégorie de marqueur discursif peut également permettre d'exprimer l'accord du locuteur sur ce que dit ou propose son interlocuteur (Kerbrat-Orecchioni 1990 : 18, Delahaie 2009)	*Ex :* *- On se voit demain ?* *- **Ouais,** ça roule ! à demain !*
	Marqueur intersubjectif	Dispositif d'ajustement permettant au locuteur d'impliquer/ d'orienter son interlocuteur (Anzai 2009 : 69) en sollicitant sa compréhension ou son engagement dans le contenu énoncé. L'usage de ce marqueur permet d'établir le rapport intersubjectif entre le locuteur et son interlocuteur (Quillard 2007, Noda 2011 : 244-245). Ces marqueurs sont très fréquemment utilisés en position parenthétique (Andersen 2007, Bolly 2010: 676) et permettent d'articuler le discours avec le maintien du rapport intersubjectif.	*Ex :* *- Comment tu fais pour rester calme avec tout ce qui t'arrive ?* *- ah **tu sais** j'essaie de faire la part de chose et je me dis que c'est pas la fin du monde hein*
	Marqueur balisage	Marqueur contextuel permettant au locuteur de mettre en évidence un lieu, une temporalité ou un élément du discours (Dostie 2007 : 36).	*Ex : **Ah bah là**, on est dans de beaux draps ! Je ne m'attendais pas à ça !*
Marque du registre oral	**Variations morphologiques orales**	Concerne les contractions morphologiques opérées par les locuteurs natifs pour raccourcir leurs phrases lors de l'usage du registre familier, par exemple *chais pas* pour *je ne sais pas* (Surcouf 2020 : 245).	*Je ne sais => **chais pas***
	Idiome familier	Concerne tous les emplois de lexique ou de lexique / séquences lexicales stéréotypées telles que les proverbes, les locutions verbales ou les expressions familières fréquemment retrouvées à l'oral.	*Ex :* *- Comment tu trouves le dernier film de Spielberg ?* *- Ah bah **je t'avoue** que je l'ai pas trouvé **folichon***
Marques anomalies lexicopragmatiques et formelles	**Transfert lexical**	Renvoie aux alternances codiques, c'est-à-dire aux transferts de lexique de la langue maternelle vers la langue cible pour pallier un manque en langue cible.	*Ex :* *- On utilise souvent **soya sauce** pour cuisiner.* *- la sauce soja ?* *- Oui*
	Transfert expressif	Transfert des marqueurs/réactions expressifs de la langue maternelle vers la langue cible.	Toutes manifestations expressives qui semblent inhabituelles en français. [102]
	Transfert MD	Renvoie au transfert de marqueurs discursifs de la langue maternelle vers la langue cible pour pallier un manque en langue cible.	Ex : -うーん (*un* allongé = *euh*) / なんだろう (*nan darô = comment dire…*) / *différent entre le Japon et la France* *- tu veux parler des différences entre France et Japon ?*
	Erreur pragmatique	Renvoie à l'usage pragmatique inapproprié ou inadapté à la situation où une autre réponse pragmatique est habituellement attendue lors de la communication en français.	*Ex :* *- je viens de perdre mon père hier ….* *- **oh que je te plains**, ça doit être dur*
	Erreur formelle	Erreur syntaxique ou grammaticale et usage de structures bancales lors de la communication en français.	*Ex :* *- **hier j'ai mangé trop bon repas ensemble*** *- oui c'était excellent !*

[102] Nous verrons quelques exemples de ce type de transfert dans la partie analyse.

	Erreur lexicale	Usage inapproprié d'un terme lors de la communication en français.	*Ex :* *- récemment **j'ai gagné** du français utile* *- super ! c'est cool si tu progresses bien !*

Tableau 5 : Version finale de la grille d'annotation suite aux améliorations suggérées par le processus d'accord interannotateur[103]

[103] Bien évidemment, l'objectif de ce travail de catégorisation n'a pas pour but de proposer une typologie fine et construite de la catégorie des marqueurs discursifs, mais vise avant tout à proposer un cadre d'annotation empirique pour la caractérisation de nos données en vue de leur analyse. Il ne sera donc pas rare que le locuteur observe que certaines des catégories soient poreuses et se recoupent entre-elles. Les choix qui ont été fait dans cette catégorisation sont donc les nôtres et ne constituent en aucune manière une typologie approfondie des marqueurs discursifs.

L'objectif principal de l'annotation ci-dessus est de détecter d'éventuels sous-emplois ou suremplois caractéristiques de la compétence lexicopragmatique de l'apprenant japonophone du français de niveau intermédiaire. Dans un second temps, il s'agit de quantifier ces emplois et de les mettre en contraste avec ceux du corpus contrôle afin d'observer les spécificités de la production des LNN par rapport aux emplois constatés lors d'une conversation quotidienne entre les LN. Les marqueurs identifiés dans cette annotation constituent des ancrages concrets et observables spécifiques à l'acte de communication et permettent de mettre en relief l'attitude et le positionnement du locuteur concernant différents aspects de la situation d'interaction en cours (Ostman 1982 : 152). Ces marques permettent de spécifier le dire du locuteur, d'exprimer une attitude ou une posture particulière ainsi que de renforcer l'engagement du locuteur dans l'interaction. En utilisant ces marques, le locuteur structure son discours, le rend plus naturel et cohérent, en s'appuyant sur des « façons de dire » conventionnellement établies dans la communauté linguistique à laquelle il appartient (Pellet 2005). C'est pour ces diverses raisons que nous avons principalement orienté le diagnostic de la compétence pragmatique sur les marques suivantes :

Marques expressives

Les marques expressives sont généralement très productives dans la conversation quotidienne et permettent au locuteur d'exprimer son état mental et émotionnel en réaction à un dire ou un évènement qu'il rencontre dans la situation d'énonciation (Searle 1975 : 356-357 ; Searle & Vanderveken 1985 : 12-20). Notre expérimentation diagnostique vise à observer les spécificités d'usage des PhPex chez les locuteurs natifs et les locuteurs non natifs du français. Nous avons ainsi tenté de proposer une typologie des différentes marques expressives susceptibles d'apparaitre dans notre corpus. Parmi ces marques, nous avons pris en compte les éventuelles occurrences de PhPex, les occurrences de rires, les réactions vocales et les interjections.

Marques discursives et interactionnelles (MDI)

Les MDI constituent des marqueurs pragmatiques fréquemment rencontrés dans la conversation quotidienne et statistiquement très représentés dans le discours du locuteur natif (Schiffrin 1987 ; Schourup 1999 ; Pellet 2005 ; Crible 2017, 2019). Ces marqueurs sont

généralement compacts dans leur forme, facultatifs, non essentiels à la construction du sens ou à la progression thématique (Schourup 1999 : 231-232 ; Dostie et Pusch 2007), mais occupent en revanche des fonctions discursives, interactionnelles et pragmatiques diverses (Pellet 2005 : 22). Ces marqueurs sont syntaxiquement indépendants (effaçables et autonomes) et peuvent se rencontrer en début, en incise ou à la fin d'un tour de parole (*idem* : 22). Elles ont une fonction *intersubjective* et *structurante* en ce sens qu'elles permettent à la fois d'organiser les rapports interpersonnels (Traugott 2010) et d'articuler les unités de communication (Bunt 2011).

Ces phénomènes linguistiques nous apparaissent tout à fait représentatifs des dynamiques lexicopragmatiques et discursives de co-construction qui caractérisent les interactions orales informelles. Nous avons tenté d'annoter des objets tels que les *marqueurs intersubjectifs* permettant d'impliquer l'interlocuteur dans son dire, d'attirer son attention sur des éléments importants, ou les *marqueurs épistémiques* qui permettent d'ajuster son degré de certitude ou de moduler son assertion. Il est à noter qu'une même forme linguistique peut occuper plusieurs fonctions et donc rentrer dans plusieurs catégories (par ex. *voilà* qui peut être tantôt *connecteur*, tantôt *marque de ratification*).

Marques du registre oral

Cette catégorie concerne principalement des marques spécifiques de la compétence orale et constitue tout un ensemble de phénomènes interactionnels fréquemment observés lors de la conversation quotidienne en français (Traverso 1996 : 12-13, 2003 : 4-5). Parmi les marques annotées, on retrouve les contractions morphologiques orales et le lexique relatif à la discussion quotidienne (Traverso 2003 : 7-8). Notre objectif ici était de caractériser quelques paramètres propres à l'oral et étant observable au niveau segmental (morphologie lexicale, emplois lexicaux, registres), car nous pensons que la maitrise de la compétence lexicopragmatique passe par la faculté à repérer les spécificités du registre oral et à adapter les emplois du lexique préfabriqué en fonction de ces spécificités. La caractérisation de ces paramètres nous permet ainsi d'observer la compétence lexicopragmatique des LNN en comparant les fréquences d'emploi de ces marques avec celle des LN.

Marques d'anomalies lexicopragmatiques et formelles

Les marques d'anomalies sont des marqueurs spécifiquement identifiés dans les productions orales des LNN étudiés dans notre corpus et permettant d'identifier les difficultés caractéristiques d'une compétence lexicopragmatique insuffisante dans la langue cible. Il est toutefois utile de mentionner que dans notre corpus, nous n'avons constaté aucun blocage interactionnel ni de malentendu manifeste entre les participants, la conversation se déroulant en général de manière fluide. En revanche, les anomalies annotées dans le corpus sont identifiées par l'observateur qui juge si l'élément linguistique annoté peut être considéré comme un emploi s'écartant des productions lexicopragmatiques observées dans le corpus contrôle. Ces marqueurs caractérisent notamment des sous-emplois d'objets linguistiques productifs chez les locuteurs natifs. Ces sous-emplois sont bien souvent des traces linguistiques des stratégies palliatives ou d'évitement utilisées par les locuteurs non natifs pour compenser les emplois linguistiques auxquels ils ne peuvent pas répondre avec le matériel linguistique dont ils disposent en langue cible. Parmi les marques annotées, nous observerons par exemple les différents types de transfert lexicaux (Masseron & Luste-Chaa 2008), expressifs ou discursifs (Corder 1973 ; Selinker 1969, 1972) qui peuvent se manifester sous la forme d'influences translinguistiques telles que l'alternance codique (Causa 1996 ; Trévisiol 2006 ; Trévisiol & Larruy 2017) ou des emplois inadéquats, par exemple des erreurs pragmatiques, formelles ou lexicales (Corder, Perdue & Porquier 1980 ; Dewaele & Wourm 2002). Une forte productivité de ces marques peut constituer une trace matérielle objective d'une compétence lexicopragmatique en cours d'acquisition.

3.3.4. Implémentation du schéma d'annotation et utilisation sur ELAN

L'implémentation du schéma d'annotation sur ELAN s'est faite en plusieurs étapes. Après avoir réalisé une première version de notre schéma d'annotation, nous avons procédé à la création de notre fichier d'annotation sur ELAN en fusionnant la vidéo et la bande-son. Nous avons ensuite attribué un canal de transcription (ou *acteur* selon la terminologie d'ELAN) pour chaque locuteur puis nous avons ajouté les transcriptions dans chaque canal. Nous avons ensuite établi une structure d'annotation arborescente en affectant plusieurs *acteurs enfants* à chaque *acteur parent* selon le modèle suivant :

Acteur parent 1 (Locuteur A)
-> Acteur enfant 1 : Marqueur expressif
-> Valeur A
-> Valeur B
-> Valeur C
-> Acteur enfant 2 : Marqueur discursif
-> Valeur A
-> Valeur B
-> Valeur C
-> …

Figure 15 : Schéma d'arborescence des annotations en partition sur ELAN

Ci-dessous, une impression d'écran du rendu de l'arborescence sous ELAN :

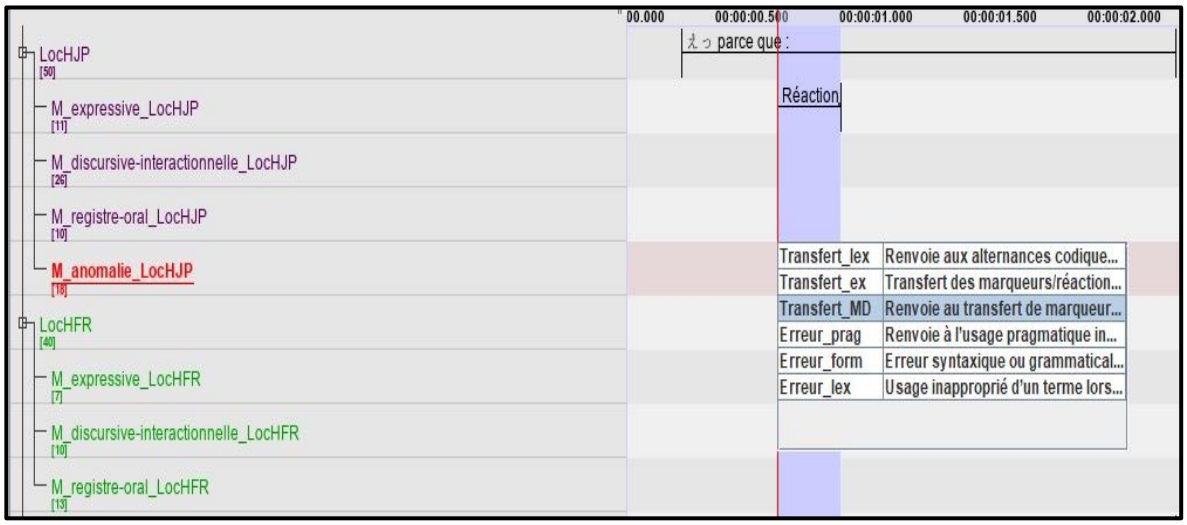


Figure 16 : Exemple d'arborescence annotation en partition ELAN

Nous avons par la suite intégré notre typologie d'annotation sur ELAN en paramétrant un *vocabulaire contrôlé,* outil permettant d'implémenter directement dans ELAN notre schéma d'annotation, ce qui permet d'affecter à un acteur enfant différentes valeurs correspondant aux différents phénomènes pragmatiques et interactionnels détaillés dans la typologie d'annotation. Enfin, nous avons paramétré le type de segmentation (*stéréotype* selon la terminologie d'ELAN) de nos annotations en optant pour une segmentation indépendante, mais étant strictement comprise entre le début et la fin du segment de transcription annoté (*included-in*). L'avantage de notre méthode est que nous pouvons obtenir ainsi des données quantitatives personnalisées pour chaque locuteur et pour chaque catégorie d'annotation de manière distincte et ainsi affiner la granularité de notre analyse et des mesures quantitatives.

Les méthodologies d'annotation de la compétence lexicopragmatique sur corpus oral étant encore peu développées sur le plan outillé, nous avons progressé à tâtons. Bien que cette typologie d'annotation constitue une tentative prometteuse de caractériser la compétence lexicopragmatique du LNN, il ne fait aucun doute qu'elle n'est pas exhaustive et encore perfectible étant donné la complexité du phénomène observé et le nombre important de paramètres linguistiques, socioculturels et extralinguistiques (voire psycholinguistiques) impliqués dans la production, la compréhension et l'acquisition. Il faut également mentionner la porosité de certaines des catégories annotées qui peuvent parfois se recouper avec d'autres catégories, augmentant ainsi la difficulté de la tâche d'annotation[104]. Dans la section qui suit, nous allons détailler notre analyse dans une perspective quantitative puis qualitative. Nous proposerons ensuite une synthèse diagnostique puis nous conclurons cette partie.

[104] Pausé et Tutin (2022) font la même constatation pour ce qui est de l'annotation des phrases préfabriquées de l'interaction. Elles ajoutent que c'est à la fois le caractère multidimensionnel et la polyfonctionnalité de certaines des expressions à fonction pragmatique qui complexifie la tâche d'annotation.

Chapitre 9 - Analyse des résultats du diagnostic

Nous allons dans ce qui suit synthétiser les résultats quantitatifs des phénomènes annotés dans les corpus exolingues puis nous contrasterons ces résultats avec ceux retrouvés dans le corpus contrôle. Dans un second temps, nous mènerons une analyse qualitative pour analyser plus concrètement les productions linguistiques des LNN et proposer quelques hypothèses concernant les difficultés observées. Enfin nous conclurons cette partie et ce chapitre sur une synthèse diagnostique qui nous permettra d'introduire les hypothèses qui seront observées dans le prochain chapitre centré sur une étude lexicopragmatique de la compétence des LNN japonophones du français de niveau avancé.

1. Synthèse des résultats quantitatifs par catégorie d'annotation

Comme évoqué précédemment dans la partie dédiée à la caractérisation du corpus employé pour notre analyse diagnostique, les extraits annotés ne sont pas strictement équivalents au niveau de la longueur (pour rappel : Exolingue-Grenoble = 11'45 / Exolingue-Tsukuba = 17'28 / Endolingue-Tsukuba = 9'10). De même, le nombre de locuteurs natifs et non natifs n'est pas équilibré pour tous les corpus et le temps de parole pour chaque locuteur est fortement inégal. De ce fait, nous nous proposons d'analyser dans cette partie non pas les fréquences brutes des annotations retrouvées dans notre corpus, mais la fréquence relative par locuteur (nombre d'occurrences par minute) qui sera pondérée sur un temps de parole de 10 minutes par locuteurs (multiplication de la fréquence relative par 10). Les résultats proposés dans les graphiques ci-dessous sont donc des projections de la fréquence d'emploi des marques annotées sur dix minutes de temps de parole pour chaque locuteur. Nous nous servons ensuite de ces fréquences relatives pour calculer la fréquence moyenne pour chaque population.

1.1. Marques expressives

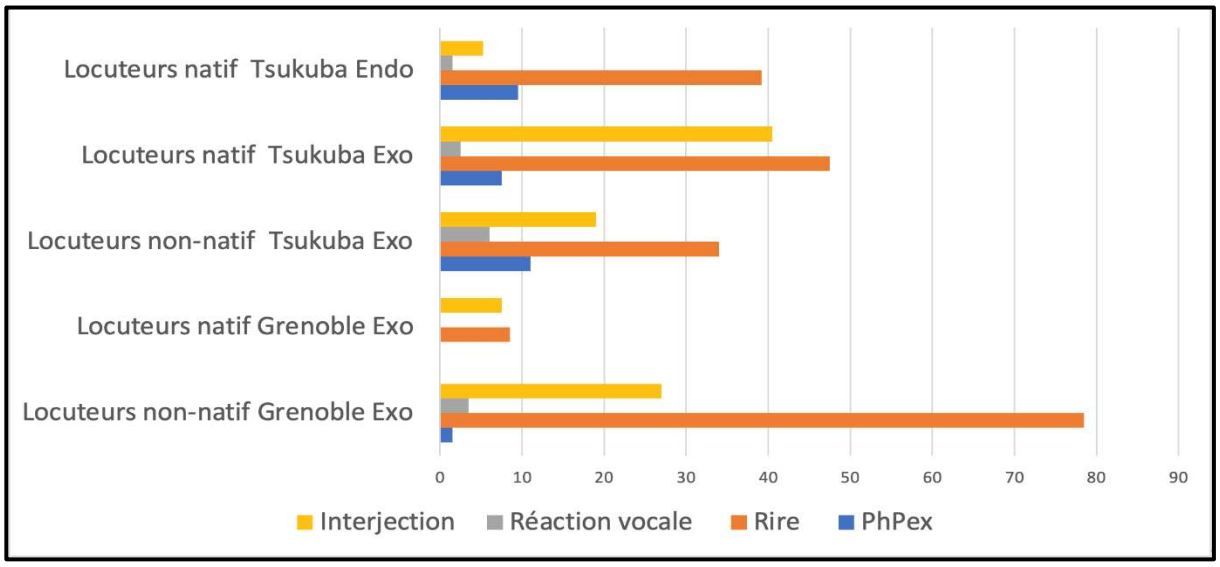


Figure 17 : Moyenne des fréquences relatives d'emploi des marques expressives par groupe de locuteurs

Dans le tableau ci-dessus, le nombre de marques expressives repérées dans le corpus Exolingue-Grenoble compte un total de 16 marques pour les locuteurs natifs contre un total de 110 marques pour les locuteurs non natifs pour 5 extraits totalisant 11'45 de conversation. Les sous-types de marques expressives les plus représentées sont les rires (87) et les interjections (34). Les rires et les interjections sont très représentés chez les locuteurs non natifs, mais moins chez les locuteurs natifs. Les occurrences de phraséologismes expressifs ou de réactions expressives sont en revanche marginales. Les rires observés chez les locuteurs non natifs sont ici intéressants puisqu'ils interviennent à des moments où l'interlocuteur ne semble pas évoquer de faits amusants ou ne semble pas faire de trait d'humour. Nous pensons que ce rire joue une fonction pragmatique spécifique au LNN interagissant avec des LN dans un cadre exolingue. Nous le verrons dans la section suivante, le rire peut être une trace intéressante sur laquelle peut reposer une partie du diagnostic de la compétence lexicopragmatique du LNN.

Le nombre de marques expressives repérées dans le corpus Exolingue-Tsukuba compte un total de 98 marques pour les locuteurs natifs contre un total de 70 marques pour le locuteur non natif. Les sous-types de marques expressives les plus représentées sont comme précédemment les rires (81) et les interjections (59). À l'inverse du corpus Exolingue-Grenoble, les rires et les interjections sont plus représentés chez les natifs que chez les locuteurs non natifs. Ceci pourrait s'expliquer par deux raisons : a) le locuteur non natif était

seul, ce qui diminue proportionnellement les valeurs des résultats par rapport au corpus Exolingue-Grenoble, b) le contexte de discussion étant un club de conversation libre, le climat de convivialité s'est installé beaucoup plus vite, alors que les locuteurs du corpus Exolingue-Grenoble ont éprouvé certaines difficultés à se détacher du cadre institutionnel et à se « lâcher ». Toutefois, à l'instar du corpus Exolingue-Grenoble, les rires des locuteurs natifs et des locuteurs non natifs ne semblent pas intervenir aux mêmes moments ni pour les mêmes raisons. Les PhPex sont également plus représentés dans le groupe Exolingue-Tsukuba, là aussi en raison du cadre propice à l'emploi de ces expressions et aux réactions expressives. On remarque que le locuteur non natif cumule à lui seul 6 occurrences de PhPex.

Le corpus d'interaction endolingue a donné des résultats surprenants. En effet, seulement 55 marques expressives ont été relevées. Étant donné le cadre de discussion libre et le fait que les interlocuteurs se connaissaient, nous avions envisagé un plus grand nombre de marques expressives. Nous pensons que cette faible proportion est probablement due au fait que l'extrait étudié est plus petit en comparaison des extraits exolingues. Nous remarquons toutefois que les marques les plus représentées sont les PhPex (6) et les rires (26), ce qui ne semble pas étonnant étant donné le contexte d'interaction libre propice à l'usage des marques expressives familières.

1.2. Marques discursive

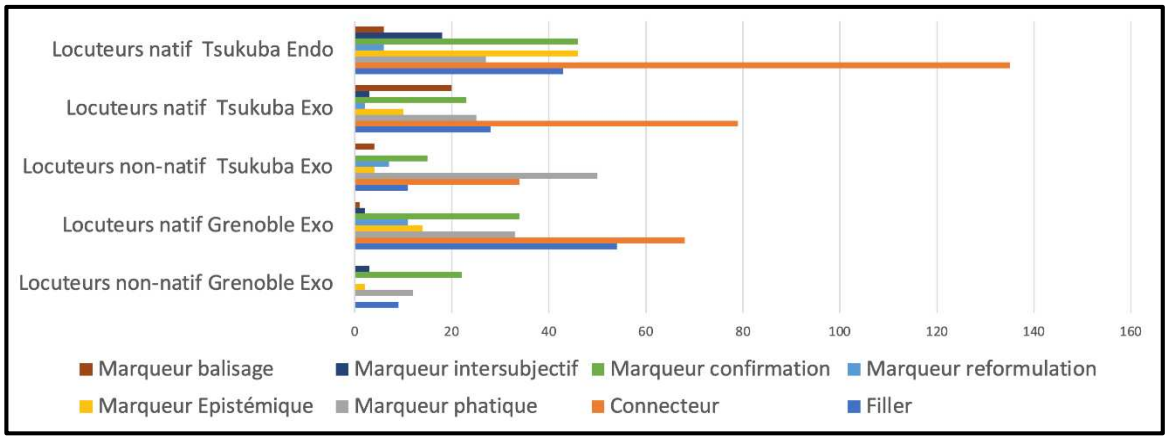


Figure 18 : Moyenne des fréquences relatives d'emploi des marques discursives et interactionnelles par groupe de locuteur

On observe un total de 46 marques discursives et interactionnelles (désormais MDI) pour les LNN contre 216 MDI pour les LN dans le corpus Grenoble-Exolingue. Chez les LNN, les

marques les plus représentées sont les marques de confirmation (21) et les marques phatiques[105] (12) en français. Pour ce qui est des locuteurs natifs, ce sont les *fillers* (54), les connecteurs (68) et les marques phatiques (32) qui sont les plus représentés. On peut déjà remarquer à ce stade de l'analyse une plus grande productivité des MDI chez les locuteurs natifs que chez les locuteurs non natifs, ce qui semble être un indice d'une différence de compétence lexicopragmatique entre les deux groupes de locuteurs. Cette tendance se confirme encore plus lorsque l'on contraste ces résultats avec ceux du corpus endolingue. Les résultats montrent en effet une très grande productivité et une grande variété de l'usage des MDI avec notamment une très forte prépondérance des connecteurs.

On retrouve la même tendance dans le corpus Exolingue-Tsukuba avec d'un côté 188 MDI pour les LN contre 125 marques pour le LNN. Les marques les plus représentées chez les LN sont les connecteurs (79), les *fillers* (28) et les marques phatiques (25). En ce qui concerne les marques comptabilisées chez le LNN, les marques les plus représentées sont les marques phatiques (50), les connecteurs (34) et les confirmations (15). On peut observer dans les résultats du tableau que globalement, le LNN du corpus Exolingue-Tsukuba semble cumuler presque trois fois plus de MDI que les deux LNN du corpus de Grenoble. Cela pourrait s'expliquer par la différence de la qualité, de la longueur et de la diversité d'exposition à la langue française entre les deux groupes de LNN. D'un côté, on peut observer des primo-arrivantes en début de séjour en France dans le cadre scolaire, de l'autre côté nous avons un locuteur ayant discutant régulièrement de manière informelle avec des locuteurs natifs du français dans la vie privée et lors du club de discussion Francophonie. Cette différence d'exposition au français semble se manifester par une diversité des formes linguistiques employées plus importante et une plus grande occupation de la parole dans l'interaction chez le locuteur du corpus Exolingue-Tsukuba que chez les locutrices du corpus Exolingue-Grenoble.

Les résultats retrouvés dans le corpus endolingue affichent un total de 326 marques discursives et interactionnelles soit environ 20 % de plus que dans le corpus Exolingue-Tsukuba et 40% de plus que le corpus Exolingue-Grenoble. Les marques les plus représentées sont les connecteurs (135), les *fillers* (43) et les marqueurs de confirmation (45). On remarque la représentation importante des marqueurs épistémiques (45) et intersubjectifs

---

[105] Cette valeur ne prend pas en compte les marques phatiques transférées de la LM que nous traiterons dans la section suivante.

(18), marqueurs faiblement représentés chez les LN dans les corpus exolingues. Une hypothèse que l'on peut émettre concernant cette faible représentation est celle d'un changement de posture interactionnelle chez les LN lors de la conversation exolingue. En effet, ceux-ci jugeant que la compétence conversationnelle de l'interlocuteur non natif n'est pas suffisante vont avoir tendance à adapter leur production langagière de manière à la rendre plus saillante et plus simple à comprendre pour le locuteur non natif. Cette simplification va impliquer une diminution des marqueurs complexes et du lexique stéréotypé au profit de tournures courtes et de nombreuses marques de reformulation.

1.3. Marques de l'oralité

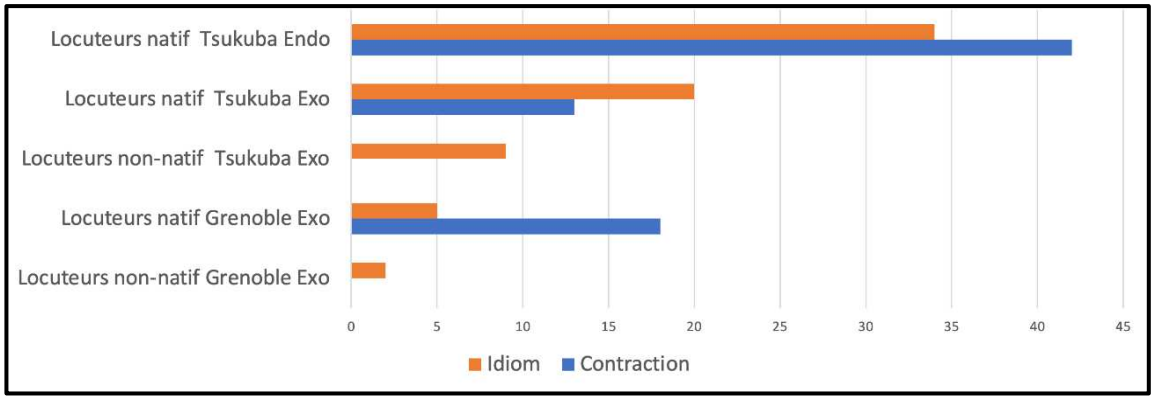


Figure 19 : Moyenne des fréquences relatives d'emploi des marques du registre oral par groupe de locuteurs[106]

Le corpus Exolingue-Grenoble présente ici une très grande disparité entre le nombre total de marques du registre oral pour les LN (51) et pour les LNN (15). Sans grande surprise, les deux locutrices ayant participé à notre étude n'ont produit aucune contraction morphologique contrairement aux locuteurs natifs qui ont recours de manière assez fréquente à ces contractions lexicales spécifiques de l'oral. On observe cependant que contrairement aux deux autres corpus, la fréquence de ces marques reste relativement faible chez les locuteurs natifs également. Une des explications plausibles est qu'un seul des deux locuteurs natifs a pris de nombreuses fois la parole et semblait adopter une posture d'adaptation envers les deux locutrices japonaises. Ce faisant, il est apparu dans le corpus que le locuteur prenait soin de parler clairement et de ne pas employer de structures trop complexes pour faciliter

[106] *Idiom* = emplois de constructions idiomatiques / *contraction* = emploi de contractions morphologiques caractéristique du registre oral spontané.

l'interaction avec les locuteurs non natifs. C'est selon toute logique ce qui pourrait expliquer l'usage peu fréquent des contractions ou des chevauchements du côté des natifs. On remarque également un emploi faible des structures lexicales spécifiques à la conversation (idiom).

Le corpus Exolingue-Tsukuba affiche des valeurs plus importantes avec un total de 67 marques du registre oral pour le locuteur non natif et 101 marques pour ses interlocuteurs natifs. Nous remarquons plusieurs choses intéressantes. Premièrement, à l'instar des locutrices japonaises du corpus Exolingue-Grenoble, le locuteur du corpus Exolingue-Tsukuba semble ne pas avoir recours aux formes contractées, mais semble en revanche utiliser de manière assez fréquente quelques expressions spécifiques à la conversation orale (quelques exemples seront donnés dans l'analyse qualitative). Il semblerait que son expérience plus importante des interactions informelles avec des natifs ait permis à ce locuteur d'avoir recours plus aisément à ce marqueur du registre oral. On remarque également que l'usage des structures idiomatiques et familières propres à l'oral est plus important chez les locuteurs natifs (20) que chez le locuteur non natif (9).

Le corpus endolingue affiche des valeurs qui semblent représentatives de la dynamique de la conversation classique entre locuteurs natifs. De plus, la constatation d'une posture d'adaptation du locuteur natif en réponse à la perception de la compétence du locuteur non natif qui apparait comme insuffisante semble ici s'étayer à nouveau. En effet, on constate que le nombre de contractions employées est bien plus important (42) que pour les deux corpus exolingues, car les locuteurs natifs ne se contiennent pas sur le plan lexical et utilisent leur compétence lexicopragmatique de manière spontanée. De même on observe un usage important du lexique familier spécifique à la conversation.

1.4. Marques d'anomalies lexicopragmatiques et formelles

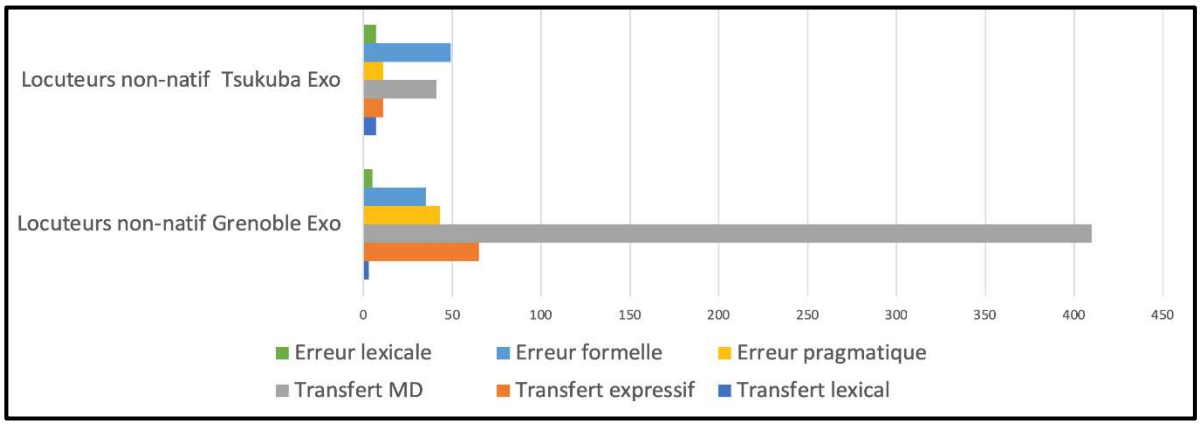


Figure 20 : Moyenne des fréquences relatives des anomalies interactionnelles entre les locuteurs non natifs des corpus de Grenoble et de Tsukuba

Le corpus Exolingue-Grenoble totalise un nombre relativement important d'emplois inhabituels. Ainsi, nous avons pu observer au cours des 11'45 du corpus majoritairement des transferts [107] de marques discursives (410), des transferts expressifs (65), des erreurs pragmatiques (43) et des erreurs formelles (35). La prépondérance des transferts de marques discursives attire notre attention, car ceux-ci sont principalement de nature phatique ou marquant une recherche lexicale. Le lien entre cette forte prépondérance des marques phatiques et un transfert de la langue maternelle du LNN vers la langue cible apparait comme l'hypothèse la plus probable. En effet, la manière de communiquer des locuteurs japonais implique une forte propension à l'interaction convergente et harmonieuse (Higashi 1992 ; Kita & Ide 2007). De ce fait, le recours fréquent à des marques phatiques semble être une des spécificités interactionnelles propres à l'éthos communicationnel japonais, spécificité qui se manifeste ici sous forme d'emprunts de marques discursives. Par ailleurs on remarque que les erreurs pragmatiques constituent pour la plupart des erreurs de MDI. Autrement dit, l'emploi du matériel linguistique en français ne correspond pas à l'emploi habituellement observé en communication endolingue ou est employé d'une manière détournée de l'emploi habituel. Les erreurs formelles désignent des erreurs syntaxiques, lexicales et de prononciation.

---

[107] Les alternances codiques sont considérées ici comme erreurs étant donné la nature diagnostique de cette phase. Toutefois, ce type d'interaction est caractéristiques des situations exolingues et n'est pas gênant pour le déroulement de l'interaction. Tomoko Higashi (LIDILEM) a mentionné lors de la relecture de ce manuscrit que ces alternances sont naturelles et ont un rôle ludique dans une conversation exolingue où les locuteurs partagent la connaissance des deux langues.

Le corpus Exolingue-Tsukuba comporte également un nombre significatif d'anomalies lexicopragmatiques et formelles, si l'on prend en compte que le locuteur était seul contrairement au corpus Exolingue-Grenoble. Les anomalies majoritairement repérées concernent des transferts de marques discursives (41) et des erreurs formelles (49). Il est intéressant de constater ici que le LNN commet moins d'erreurs pragmatiques (11) que ses homologues du corpus Exolingue-Grenoble. Cela vient probablement de sa plus grande expérience d'interaction avec les Français. En revanche, celui-ci semble avoir commis plus d'erreurs de type formel (49).

1.5. Synthèse

Dans la figure qui suit, nous pouvons observer un graphique reprenant la moyenne des fréquences relatives par type de marque et par groupe de locuteurs (natif ou non natifs de Grenoble ou de Tsukuba) :

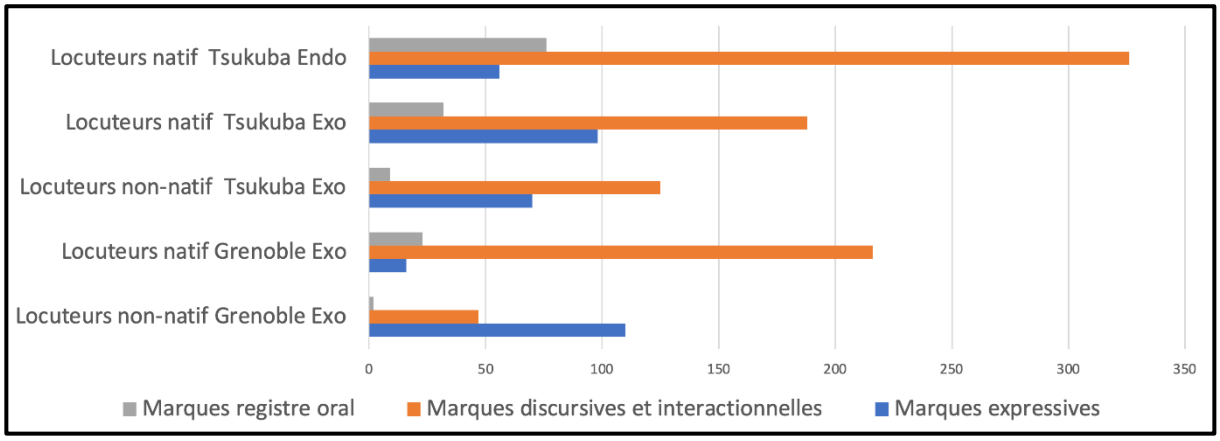


Figure 21 : Moyenne des fréquences relatives d'emploi par type de marques et par groupe de locuteurs

Le graphique laisse apparaitre de manière générale une grande productivité des marques discursives et interactionnelles chez les LN et une productivité très modérée chez les LNN. On remarque toutefois une exception chez le LNN de Tsukuba qui semble avoir de fortes propensions à employer des objets phraséopragmatiques plus fréquemment. Après observation minutieuse des données, une des hypothèses plausibles pouvant expliquer cette

propension est que le locuteur HJP semble très motivé dans son apprentissage[108]. En effet, il pose de nombreuses questions et n'hésite pas à tenter d'employer des tournures familières. De ce fait, il apparait que ce locuteur semble apporter un soin tout particulier à varier son attitude et le lexique employé pour se rapprocher de ce que l'on observe dans une conversation en contexte endolingue. De même, nous remarquons que les marques expressives semblent globalement nombreuses chez les LN et chez les LNN, à l'exception du groupe de LN de Grenoble qui affiche un score assez faible. Le contexte d'interaction semble être ici un facteur probable de cette différence. En effet, les marques expressives principalement observées sont des rires ou des marques interjectives. Ces rires sont très présents à la fois chez les LN et le LNN de Tsukuba alors que ces rires se retrouvent en majorité chez les LNN du corpus de Grenoble uniquement. Nous pensons donc qu'en fonction du contexte et des objectifs de communication, la nature et la productivité des marques expressives tendent à varier.

Dans le corpus de Grenoble, la discussion est libre, mais cette discussion prend un certain temps à démarrer et les locutrices natives semblent un peu plus en retrait que les locuteurs natifs. De même, les locuteurs présents savaient au préalable qu'ils allaient être réunis dans le cadre d'une étude linguistique, ce qui peut avoir eu une influence sur la prise de parole des uns et des autres. De même, l'enchainement des tours de parole, les thématiques de discussions et les réactions expressives semblaient plutôt similaires à une interaction exolingue avec un enchainement où le locuteur expert rythme la conversation et les locuteurs non natifs réagissent par moment en attendant le moment propice pour prendre la parole. Les rires annotés sont en majorité des rires de convivialité et/ou de temporisation face à une difficulté linguistique (nous le verrons dans la section suivante). De même, on observe qu'il y a moins d'opportunité dans le corpus pour employer les tournures expressives, car les thématiques discutées sont bien souvent explicatives et/ou comparatives et ne semblent pas particulièrement propices à la subjectivité ou à l'expression de l'affect.

---

[108] L'enregistrement vidéo montre par ailleurs que le locuteur japonophone du corpus exolingue de Tsukuba utilise son téléphone portable tout au long de la discussion pour rechercher du lexique et vérifier le sens des termes qu'il emploie. Il adopte une posture d'apprenant, contrairement aux étudiantes japonaises de Grenoble. Celles-ci ne font aucune recherche lexicale et font peu de demande d'élicitation. En outre, elles ont l'habitude d'interagir en japonais avec le locuteur masculin du corpus exolingue de Grenoble, leur posture est donc plutôt bilingue.

Dans le corpus de Tsukuba, la discussion est également libre et se déroule dans le cadre d'un club de discussion. L'objectif communicatif est différent, car les locuteurs présents viennent en général pour discuter dans le Club Francophonie sans réellement qu'il y ait d'enjeu à la conversation. Les locuteurs ont été sélectionnés sur place et ne savaient pas au préalable qu'ils allaient être sollicités dans le cadre de cette étude. De même, les locuteurs natifs étaient bien plus nombreux que le locuteur non natif (4 natifs contre 1 locuteur non- natif) ce qui là aussi peut avoir influencé le cadre de communication. La conversation apparaissait plus spontanée et à « bâtons rompus », c'est-à-dire de manière dynamique et plus proche de ce que l'on pourrait habituellement observer avec des LN en français. Les thématiques qui étaient identiques à celles abordées dans le corpus de Grenoble ont très souvent débordé sur des récits d'expérience et des thématiques personnelles comme la relation avec le petit-ami de l'une des locutrices natives ou les relations amoureuses du LNN. Les enjeux personnels et subjectifs étant nombreux, les réactions expressives ont été elles aussi plus fréquentes. Le rire notamment était très présent et fréquemment associé à des situations humoristiques. Cette différence entre les contextes et les objectifs de communication pourrait expliquer ces fréquences anormalement basses chez les locuteurs natifs de Grenoble par rapport aux natifs des deux autres corpus.

Dans cette section, nous avons analysé les résultats quantitatifs obtenus à partir de l'annotation des extraits d'interaction exolingue et endolingue sélectionnés pour cette étude diagnostique. Cette analyse des données nous a permis de relever certaines anomalies lexicopragmatique et communicationnelles intéressantes que nous nous proposons de développer point par point dans la partie suivante.

2. Analyse qualitative des anomalies pragmatiques et interactionnelles

2.1. Un usage du rire propre à la variété non-native

Dans cette section, nous allons tenter de montrer que l'usage du rire peut constituer une trace empirique du niveau de compétence lexicopragmatique dans la langue cible. Nous verrons que les lacunes lexicopragmatiques ou expressives sont compensées par l'utilisation abondante des rires. Ces rires sont employés par les LNN comme des procédés de facilitation, de palliation ou de connivence interactionnelle.

L'usage du rire étudié dans ce corpus n'est pas nécessairement déclenché par un évènement humoristique chez les LNN. Il relève bien souvent des stratégies d'adaptation communicationnelle en réponse à une situation de conflit ou d'incompréhension que le locuteur rencontre. Le rire apparait ainsi comme une stratégie permettant la gestion de moment de conflit ou de posture interactionnelle délicate (Matsumoto 2018 ; Mazzocconi 2019). En outre, le rire d'un LNN peut constituer la trace d'une stratégie palliative visant à compenser une compétence insuffisante ne permettant pas d'atteindre l'objectif communicationnel voulu (Matsumoto 2018 : 241). Le rire est aussi fréquemment associé à l'incertitude ou la lacune lexicale (Mazzocconi 2019 : 31). Chez les locuteurs japonais, le rire constitue une convention sociale bien définie (Tanaka et Campbell 2014 : 316) utilisée pour réguler la relation interpersonnelle. Le rire possède une fonction d'ajustement social (Maemura 2013 : 104) et peut être utilisé comme une marque d'empathie ou de convivialité (Miyachi 2009 ; Maemura 2013 ; Mazzocconi 2019). Cette pratique est culturellement encouragée au Japon pour faciliter les relations interpersonnelles et cacher d'éventuels sentiments négatifs ou positions délicates dans l'interaction. Miyachi (2009), partant du constat que les Japonais produisent beaucoup plus de rires que les natifs de la langue cible lors de l'interaction exolingue, étaye cette hypothèse. Il montre que les locuteurs japonais emploient les rires pour maintenir une bonne relation au sein de l'interaction ou pour pallier une lacune langagière.

En nous basant sur les observations de Miyachi (2009) et sur les données retrouvées dans le corpus, nous avons distingué les fonctions du rire suivantes : la *fonction de convivialité* (attitude communicative), la *fonction de temporisation* (gagner du temps) et la *fonction de palliation* (remplacer une lacune lexicale)[109]. La fonction conviviale correspond à une attitude communicative souvent héritée de la langue maternelle vers la langue cible et vise à renforcer la convivialité et l'harmonie de l'interaction. La fonction de temporisation correspond à une stratégie visant à gagner du temps en allongeant le temps de réponse pour trouver le lexique approprié à la réponse ou pour avoir accès au sens de ce qui est dit en laissant l'interlocuteur parler sans l'interrompre. La fonction de palliation renvoie à l'usage du rire pour pallier une lacune lexicale en langue cible.

---

[109] Le repérage et l'annotation de ces fonctions ont été faits à partir de critères prosodiques, intonatifs et interactionnels. Nous avons observé chaque rire au cas par cas puis nous les avons catégorisés. Nous avons ensuite soumis notre catégorisation à des locuteurs japonophones afin de confirmer la catégorisation.

2.1.1. La fonction de convivialité

Le premier cas étudié est celui du marquage de la connivence sous la forme d'un emploi de rire convivial. Ce rire court est exprimé en réaction à un dire de l'interlocuteur natif qui n'est pas de nature humoristique. Ce premier type de rire se remarque aisément, car au moment de l'emploi, seules les locutrices non natives rient alors que les locuteurs natifs eux n'expriment pas spécifiquement d'amusement en réaction à ce qui est dit. Le rire a donc ici une dimension sociale et empathique. Toutefois, la fréquence et la durée de ces rires nous amènent à nous demander s'il ne s'agit pas ici d'une influence de la langue maternelle, car d'une part on retrouve très peu ce type de rire chez les locuteurs natifs de Grenoble et de Tsukuba, d'autre part, cet emploi du rire semble correspondre à ce que l'on observe habituellement chez les locuteurs japonais en milieu endolingue :

Extrait 1 : Corpus Exolingue-Grenoble [00:01:14] :

Les locuteurs discutent à propos des différences entre la France et le Japon. L'une des locutrices japonaises prend la parole à propos des différences de comportement lorsque l'on entre dans un magasin français ou dans un magasin japonais.

LocFJP2 : et euh:
LocFJP1 : なんだろう　うーん (*nan darô uun = comment dire euh*)
LocFJP2 : quand-quand on entre euh: dans un magasin
LocHFR : ouais / magasins
LocFJP1 : あ　うん (*a un = ah oui*)
LocFJP2 : on dit bonjour
LocFJP1 : bonjour
LocHFR : bonjour ouais
LocFJP1 : うん　(*un* = marque phatique)
LocHFR : alors qu'au japon on ne dit pas bonjour
LocFJP1 : うん dit pas (*un* = marque de validation)
LocHFR : on nous dit bonjour mais nous on ne dit pas bonjour
LocFJP1 : (rire)
LocFJP2 : non mais merci
LocHFR : non merci
LocFJP2 : (rire)

Dans l'exemple ci-dessus, nous remarquons que le rire des deux locutrices japonaises se déclenche à un moment de la conversation qui ne semble pas être de nature humoristique. Le locuteur HFR explique une différence de comportement entre les clients qui rentrent dans un magasin en France par rapport aux clients du Japon. Ce contenu informationnel déclenche le rire de la locutrice japonais FJP1 puis FJP2, mais ne déclenche pas le rire des locuteurs natifs. Nous avons donc considéré qu'il s'agissait ici d'un rire de convivialité visant à assurer le bon déroulement de la conversation. Les locutrices ne semblent pas présenter de difficulté à la compréhension, ce qui écarte l'hypothèse d'un rire à fonction de temporisation ou de palliation face à une lacune lexicale. Nous observons également que les plaisanteries proposées par les locuteurs natifs sont parfois incomprises par les locutrices japonaises, qui réagissent souvent avec un temps de latence visible, puis rient en réaction aux rires des interlocuteurs natifs. Ces rires sont également des traces intéressantes permettant de caractériser le niveau de compétence lexicopragmatique du LNN :

Extrait 2 : Corpus Exolingue-Grenoble [00:03:49] :

Les locuteurs discutent à propos des différences entre la France et le Japon. Le locuteur français propose aux autres locutrices de commencer à déguster la collation offerte dans le cadre de l'étude. Ce faisant, il sert une à une les autres personnes.

LocHFR :	euh: du coup apparemment on a du jus de pomme et/ pomme groseille donc euh:
LocFJP2 :	pomme / pomme et (intonation montante)
LocHFR :	euh: pomme groseille
LocFJP1 :	groseille
LocFJP2 :	groseille
LocFJP1 :	Qu'est-ce que c'est groseille ?
LocHFR :	hum クランベリかな (*kuranberi ka na = je me demande si ce n'est pas cranberry*)
LocFJP1 :	うーん (intonation descendante) (*un = ah oui*)
LocFJP2 :	ah:
LocHFR :	enfin ça dépend si c'est de la groseille ou non
LocFJP2 :	hm hm
LocHFR :	euh: / ok alors (nom LocFJP2) qu'est-ce que tu veux / pomme (intonation montante)
LocFJP2 :	groseille (rire)
LocFJP1 :	(rire)
LocHFR :	groseille / très bien
LocFFR :	(rire)
LocFJP2 :	(rire) / je vais essayer
LocHFR :	(sert un verre de jus) je pense que c'est pas mal

LocFJP2 : うーん (*un* allongé = régulateur phatique)
LocHFR : comme à la maison
LocFFR : (rire)
LocHFR: on prend le petit déjeuner comme ça
LocFJP2: (rire)
LocFJP1: (rire)

Dans cet extrait, on peut remarquer l'occurrence de deux types de rires chez les locutrices non natives. Un premier rire de convivialité qui est employé lorsque la locutrice FJP2 choisit le jus de groseille et un deuxième rire qui sera déclenché à la fin de l'extrait. C'est ce dernier rire qui va nous intéresser. En effet, il s'agit d'une part d'un rire qui n'est pas réactif au trait d'humour du locuteur HFR, mais qui intervient après. Il apparait ici que les locutrices, n'ayant pas compris ce trait d'humour, ont réagi en différé en riant en réaction au rire de la locutrice FFR qui a été déclenché par le trait d'humour de HFR. D'autre part, nous avons observé qu'il s'agissait d'un rire ayant une intonation et une expressivité inhabituelle par rapport au rire de la locutrice française ou les autres rires employés par les LNN auparavant. Les locutrices japonaises n'ont pas compris ce trait d'humour, mais réagissent tout de même par un rire en réaction à celui de la locutrice HFR, et ce de manière à conserver la dynamique du groupe. Cette incompréhension est un peu moins fréquente dans le corpus Exolingue-Tsukuba, car le locuteur non natif, qui est un membre régulier du club de discussion, est plus habitué à l'interaction avec des locuteurs natifs. En outre, il est plus réceptif à l'humour des locuteurs natifs et aux marques qui définissent ce type de discours, ce qui se caractérise par des rires réactifs plus ou moins synchronisés avec le moment de la plaisanterie.

2.1.2. La fonction de palliation

Le deuxième cas d'emploi du rire est celui d'une réaction palliative face à une lacune linguistique généralement lexicopragmatique, qui empêche le locuteur de communiquer l'intégralité ou la complexité de ce qu'il souhaite exprimer. Cette incapacité à communiquer pleinement engendre une certaine tension interactionnelle qui se manifestera sous la forme d'un rire qui viendra compléter la phrase lacunaire ou terminer le tour de parole du LNN. Cette forme d'emploi du rire s'est retrouvée de manière assez fréquente à la fois dans le corpus Exolingue-Grenoble et dans le corpus Exolingue-Tsukuba :

Extrait 3 : Corpus Exolingue-Grenoble [00:04:23] :

Les locuteurs discutent à propos des différences entre la France et le Japon. Le locuteur français propose aux autres locutrices de discuter sur les différences culinaires entre les deux pays.

LocHFR : après on pourra discuter si vous voulez des différences entre la cuisine française et japonaise
LocFJP2 : ah oui oui / oui
LocFJP1: ah:
LocHFR : sur euh par exemple euh en france euh on utilise par exemple beaucoup de fromage ou de crème
LocFJP1 : うんうん (*un un* phatique)
LocFJP2 : うん(*un* phatique)
LocHFR : alors qu'au japon peut-être un peu moins
LocFJP1 : うんうんうん / うん うん (*un* réduplqué phatique)
LocFJP1 : on utilise ブイヨン (*buiyon* = bouillon)
LocHFR: bouillon / oui beaucoup ouais
LocFJP1 : うんうんうん うん (*un* phatique)
LocFJP2 : bouillon et
LocFJP1 : et: on mange souvent うーんなんだろう(*un* allongé + *nan darô* = recherche lexicale + *comment dire. . .*) produit / うーんなんだっけ (*un* allongé + *nan dakke* = recherche lexicale + *comment on dit. . .*) うん / haricot / haricot
LocHFR : haricot / les légumes (intonation montante)
LocFJP1 : oui (rire) oui oui (rire)
LocHFR : les légumes / d'accord / salades etc. oui c'est vrai
LocFJP2 : うーん (*un* allongé = validation)

Le rire employé à la fin par la locutrice FJP1 permet ici de maintenir la fluidité de la conversation malgré le manque de matériel linguistique pour finir le tour de parole. De même, nous avons pu remarquer qu'une gestuelle caractéristique de la locutrice FJP1 était visible à plusieurs reprises. Cette gestuelle qui consiste à se toucher le sommet de la tête avec la main ou l'index est employée lors d'un processus de recherche lexicale intensive par la locutrice. Cet indice supplémentaire nous permet d'avancer l'hypothèse selon laquelle la locutrice était effectivement en situation de lacune linguistique qu'elle a tenté de contourner avec le rire. Enfin, nous avons pu remarquer qu'il y avait un léger dysfonctionnement de la communication entre le locuteur FJP1 et HFR. En effet, celle-ci en employant *haricot* faisait en réalité référence aux produits de base de la cuisine japonaise qui utilisent les haricots comme ingrédient principal (*sauce soja, tofu, nattô, misô*, etc.). Toutefois, la locutrice japonaise ne disposant pas des compétences lexicales suffisantes pour étayer son propos, le locuteur HFR a pensé que la locutrice faisait en réalité référence aux légumes.

Cette deuxième forme de rire se distingue également par sa capacité à constituer un tour de parole à elle seule. Ainsi, nous avons fréquemment constaté l'usage d'un rire chez les locutrices japonaises employé afin de répondre à l'interlocuteur alors qu'il serait habituellement attendu une expression réactive du type *sérieux ?, ah bon ?, ah ouais !, t'as raison*, entre autres. L'emploi du rire de la deuxième catégorie nous apparait d'autant plus inhabituel que la plupart des occurrences s'observent à des moments de la conversation où le LN ne semble pas faire preuve d'humour. Ce type de rire est également très représenté dans le corpus Exolingue-Tsukuba. En effet, le locuteur non natif de ce corpus prend plus souvent la parole (318 sec) que les locutrices du corpus Exolingue-Grenoble (267 sec) ce qui se traduit par plus d'occasions de rencontrer des lacunes lexicales. Dans le cas du corpus exolingue de Tsukuba, nous avons pu observer plusieurs fois que le LNN commence son tour de parole, puis, rencontrant une lacune lexicale qu'il n'est pas à même de surmonter, va céder son tour de parole en le concluant par un rire.

2.1.3. La fonction de temporisation

Le troisième cas renvoie à la fonction de temporisation du rire pour dissimuler la contrariété ou la confusion du LNN liée à l'incompréhension du lexique utilisé par le LN. Il s'agit plutôt d'un rire relativement bref, mais accompagné d'une contraction faciale traduisant une certaine difficulté ressentie par le locuteur non natif. À l'instar du deuxième type, l'incapacité à comprendre la communication en cours ou à faire le lien avec le lexique semble être source de confusion et de stress interactionnel. Le locuteur non natif va employer ce rire pour en quelque sorte dissimuler son état d'incompréhension, de manière à ne pas rompre la fluidité de l'interaction et à gagner du temps pour comprendre ce que l'interlocuteur dit :

Extrait 4 : Corpus Exolingue-Grenoble [00:20:05] :

Les locuteurs discutent à propos de l'empereur actuel du Japon qui va abdiquer pour laisser la place à son fils lors du printemps 2020. Le locuteur HFR demande aux étudiantes japonaises ce qu'elles en pensent.

LocHFR : d'accord / euh: / un autre sujet / euh par exemple euh Alexis 'fin 先輩 (*senpaï* = ainé / étudiant ou collègue sénior) disait tout à l'heure que en: avril mai l'empereur du Japon va changer.

LocFJP2 : うーん (*un* allongé = phatique)

LocHFR: c'est très rare / très très rare c'est la première fois / l-l'une des premières fois que l'empereur change qui va-qu'il arrête
LocFJP2 : うん(*un* = phatique) qui arrête
LocHFR : il va: / euh: l'empereur hm actuel
LocFJP2 : うん(*un* phatique)
LocHFR : il va arrêter d'être l'empereur
LocFJP2 : うん (*un* validation)
LocHFR : et c'est un nouvel-c'est son fils qui va devenir empereur c'est ça (intonation montante) / en: avril peut-être (regarde FJP1 pour confirmation)
LocFJP1 : (rire)
LocFJP2 : (rire)
LocHFR: et le le prince euh comment je peux expliquer hum: le fils de l'empereur / actuel
LocFJP2 : fille
LocHFR : le fils
LocFJP1: le fils
LocFJP2 : le fils oui
LocHFR : le fils de l'empereur
LocFJP1 : -preur
LocHFR : de l'empereur
LocFJP1 : de l'empereur
LocHFR : c'est euh celui qui dirige

Dans cet exemple, nous pouvons observer que les deux locutrices japonaises rencontrent une lacune lexicale importante. En effet, elles ne vont pas trop comprendre ce dont il est question, car elles ignorent le sens du mot *empereur*. Vers le milieu de l'extrait, le locuteur HFR va demander une confirmation aux deux locutrices que l'abdication de l'empereur est bien prévue pour avril. Celles-ci ne comprenant pas la demande du locuteur HFR vont employer un rire de temporisation afin de retarder leur prise de parole et laisser continuer le locuteur HFR afin de mieux comprendre la thématique de la conversation. Le locuteur va ensuite tenter plusieurs étayages afin d'aider les deux locutrices à comprendre la conversation. Ci-dessous, le tableau 6 résume ces trois cas observés dans notre corpus :

	Description
Fonction de convivialité	Marquage de la connivence sous la forme d'un emploi de rire convivial court et expressif.
Fonction de palliation	Réaction face à une lacune linguistique généralement lexicale ou lexicopragmatique, qui empêche le locuteur de communiquer l'intégralité ou la complexité de ce qu'il souhaite exprimer.
Fonction de temporisation	Dissimuler la contrariété ou la confusion du locuteur non natif liée à l'incompréhension du lexique utilisé par le locuteur natif.

Tableau 6 : Les trois fonctions du rire chez les LNN observés

On remarque que le rire, au-delà d'être une marque d'empathie et de connivence, est essentiellement utilisé comme stratégie pour remplir les blancs dans la conversation, pallier les lacunes lexicales ou éventuellement dissimuler un moment d'incompréhension. Ce type de rire social est donc mis en place ici comme une ressource permettant de compenser une lacune dans la compétence lexicopragmatique. Notons également que les résultats nous montrent que les locuteurs du corpus Exolingue-Grenoble sont moins à l'aise dans la conversation orale et ont tendance à utiliser beaucoup les rires pour pallier ce manque de compétence. En revanche, le locuteur du corpus Exolingue-Tsukuba a quant à lui plus l'habitude de la conversation, les lacunes lexicopragmatiques et communicationnelles sont moins fréquentes, ce qui explique qu'il a moins recours au rire et que celui-ci se retrouve souvent lors des phases de connivence ou lors des moments cocasses. Les éléments évoqués précédemment semblent donc étayer notre hypothèse que ces emplois inhabituels du rire sont tout à fait utiles dans le diagnostic de la compétence lexicopragmatique des LNN observés. En effet, la surabondance du rire peut constituer selon nous une trace concrète d'une recherche ou d'un manque lexicopragmatique.

2.2. Un suremploi des formes phatiques

Dans cette section, nous allons tenter de montrer que le suremploi des formes phatiques remarqué dans notre étude peut constituer une autre trace empirique utile pour le diagnostic de la compétence lexicopragmatique des locuteurs non natifs dans la langue cible. Comme pour le rire, notre hypothèse ici est que le manque de compétence et les lacunes lexicopragmatiques ou expressives sont compensés par l'utilisation abondante des marques qui sont utilisées par les locuteurs non natifs comme procédé d'évitement ou de palliation. Nous verrons également qu'un transfert s'effectue au niveau lexical et communicatif et que ce transfert donnera lieu à des emplois inhabituels pour le français.

L'analyse quantitative des résultats a montré une très forte fréquence de l'emploi des marques phatiques, et notamment de marques empruntées à la langue japonaise. De nombreuses recherches sur la manière de communiquer des locuteurs japonais montrent

qu'ils sont enclins[110] à adopter une attitude taciturne et en retrait plutôt qu'à adopter des réactions expressives exubérantes (Maynard 1986). Ils préfèreront ainsi une communication plus harmonieuse et emphatique basée sur l'écoute et l'accompagnement de la prise de parole de l'interlocuteur. Un des instruments privilégiés de la mise en place de cette communication harmonieuse constitue les marques phatiques à valeur sociale nommée *aizuchi*[111]. La notion renvoie à l'origine à une technique de forge durant laquelle le maitre forgeron et son assistant vont battre le fer à tour de rôle, le maitre indiquant précisément avec son marteau l'endroit où doit frapper l'apprenti. En référence à cette image d'alternance parfaite et harmonieuse, l'utilisation des *aizuchi* symbolise cette volonté de convergence et d'adaptation avec l'interlocuteur. De même, les *aizuchi* constituent des marques phatiques généralement accompagnées des hochements de têtes fréquents ou *unazuki*[112] (Maynard 1987 ; Kita 1996 ; Kita et Ide 2007). Notons que ces marques semblent avoir une importance telle que leur omission peut être perçue comme une impolitesse par l'interlocuteur de langue japonaise (Locastro 1987 : 103). L'harmonie au sein de l'interaction et des interactants semble donc être des notions plus centrales pour le locuteur japonais que la clarté et la franchise, quitte à s'adapter au rythme et à la distribution de la parole imposée par l'interlocuteur (Matsumoto & Boye Lafayette 2000 : 193).

Maynard (1986, 1990) et Locastro (1987, 1999) remarquent que les locuteurs japonais ont de fortes tendances à reproduire les habitus interactionnels développés en langue maternelle lors de l'interaction exolingue avec des LN de la langue cible. Ceux-ci utilisent de manière abondante les marques phatiques accompagnées de hochement de tête reproduisant ainsi la manière de communiquer en langue japonaise[113]. Maynard remarque également que ces marques prennent généralement la forme de courtes productions présentant peu de variété au niveau de la forme linguistique contrairement aux locuteurs anglophones qui ont tendance à utiliser peu de marques phatiques, mais avec une plus grande variété des formes. De plus, elle observe qu'une grande partie de ces marques phatiques sont employées par les LNN de manière simultanée et coordonnée (à la manière des *aizuchi*) à la prise de parole de

---

[110] Attention toutefois, il s'agit d'une constatation générale qui peut varier en fonction des individus, des contextes de communication, voire en fonction de variantes régionales.
[111] Littéralement « mutuel » et « marteau »
[112] Donner un signe de la tête ou faire oui en inclinant la tête.
[113] Approximativement 3x plus de marques que les locuteurs natifs, ce que nous avons pu confirmer dans notre étude quantitative.

l'interlocuteur, ce qui semble n'être jamais le cas chez les LN (Hayashi 1988 ; Maynard 1997 ; Cutrone 2005).

Nos résultats rejoignent les résultats obtenus par Maynard et Cutrone dans leurs études respectives en montrant que les locuteurs non natifs utilisent en moyenne 5 fois plus de marqueurs phatiques que les locuteurs du corpus endolingue pour le corpus Exolingue-Grenoble et deux fois plus pour le corpus Exolingue-Tsukuba. Ces marqueurs sont le plus souvent employés durant la prise de parole du LN ou dans les moments de pause que prend le LN pour articuler son propos. Par ailleurs, suite à des remarques[114] proposées lors d'une réunion d'observation de ces données, une hypothèse concernant la forte prépondérance de ces transferts s'est dégagée. En effet, on peut suspecter le fait ici que les LNN ont probablement identifié la capacité des locuteurs natifs à parler et comprendre le japonais, ce qui a vraisemblablement accentué le recours à la LM comme bouée de secours, car ils savaient que les locuteurs natifs étaient à même d'être réceptifs à ces marques phatiques. Il y a donc ici un effet d'accentuation du recours au transfert de la langue maternelle en fonction de la représentation qu'ont les locuteurs non natifs du répertoire linguistique de leur interlocuteur natif. En outre, les formes phatiques employées par les locuteurs natifs étaient d'une part brève et peu diversifiée (uniquement des formes vocales monosyllabiques, parfois rédupliquées), d'autre part, elles constituaient majoritairement des formes empruntées à la langue maternelle. Ainsi les formes principalement employées étaient les suivantes :

- あ (*ah*)
- うん (*un*)
- えー (*hé* allongé)
- ふーん (*hun* allongé)
- おー (*oh* allongé)

Ces formes phatiques correspondent aux formes observées habituellement dans l'interaction entre locuteurs japonais, corroborant les observations faites par Maynard (1997), Cutrone (2005) et Ike (2010). En outre, lorsque les LNN de nos données emploient des formes phatiques en français, on remarque que cet emploi semble quelque peu différer de l'emploi habituel par les LN et semble parfois se rapprocher de la pratique conversationnelle en

[114] Nous remercions ici Tomoko Higashi (LIDILEM) pour ces remarques.

langue japonaise. Par exemple, nous avons pu observer plusieurs fois dans le corpus Exolingue-Grenoble que l'une des locutrices avait tendance à utiliser *oui* et *ouais* en tant que marque phatique, mais l'emploi semble inhabituel, car proche d'une marque d'accord en français :

Extrait 5 : Corpus Exolingue-Grenoble [00:11:57]

Les locuteurs discutent des raisons pour lesquelles une des locutrices a débuté le cursus universitaire en langue japonaise

LocFJP2 :	Pourquoi tu as choisi le cours de japonais ?
LocFFR :	Pourquoi le japonais ? Euh parce que euh l'année dernière euh j'étais en médecine
LocFJP2 :	ふーん (*hun* allongé)
LocFJP1 :	えー (*hé* allongé)
LocFFR :	Oui, et c'était difficile
LocFJP2 :	**Oui oui**
LocFFR :	Et du coup j'ai voulu changer
LocFJP2 :	うん (*un*)
LocFFR :	Et je suis venue en japonais parce que ça m'intéressait beaucoup

Nous pouvons observer dans la séquence ci-dessus que *oui oui* est utilisé comme marque phatique. Cela se remarque notamment dans le caractère bref et l'alternance avec les tours de parole du LN. De même, l'usage de *oui oui* s'intègre à une succession de marques phatiques transférées de la langue maternelle à la langue cible. Toutefois, le marqueur *oui oui* en français peut constituer une marque de confirmation ou un *évaluateur résomptif* (Kerbrat Orecchioni 1980 ; Lefeuvre 2012) associé à une manifestation du degré d'adhésion du locuteur au dire de son interlocuteur plutôt qu'une marque phatique (Lefeuvre & Legallois 2020). Lorsque *oui* est utilisé une seule fois, son emploi peut avoir tantôt une valeur de validation, tantôt de régulateur phatique. Sa réduplication lui confère une valeur de confirmation. Il nous semble donc que c'est cette distribution de la marque discursive qui se trouve dans une succession de marques phatiques et l'usage d'un marqueur habituellement associé à la confirmation qui crée cet emploi considéré ici comme inhabituel. En effet, il apparait curieux que la LNN confirme ce que dit la LN, alors que cette même LN vient d'apporter une nouvelle information inconnue de tous. En adoptant une approche translinguistique, on s'aperçoit que l'usage de *oui oui* se rapproche étrangement de la marque

de confirmation はい (*haï* avec le *h* aspiré) ou うん[115] (*un*) utilisées en japonais pour confirmer ou ratifier un propos. Toutefois lorsque cette marque est rédupliquée (*haï haï* ou *un un*) et employée simultanément durant le tour de parole de l'interlocuteur, elle remplit la fonction de marqueur phatique type *aizuchi*. Cette réduplication permet également au locuteur japonais d'exprimer son intérêt. Cette comparaison interlangue nous amène donc à penser qu'il y a bien ici un transfert de la langue maternelle de la LNN afin de pallier une lacune lexicopragmatique. En effet, on s'attendrait ici à des expressions du type *ah bon ?*, *ah ouais ?*, *C'est vrai ?* ou *Tu m'étonnes* qui en plus d'avoir une fonction phatique, permettent d'exprimer l'état d'esprit du locuteur et sa surprise en réaction au dire de l'interlocuteur.

Dans le corpus Exolingue-Grenoble, le temps de parole des deux LNN est court (4'45 de parole sur un total de 11'45) et comporte une grande proportion d'emploi de marques phatiques ce qui là aussi semble corroborer les travaux de Maynard (1997), Cutrone (2005) et Ike (2010). Les emplois de ces marques phatiques font souvent l'objet d'une utilisation synchronisée ou quasi simultanée entre les deux locutrices, ce qui est assez inhabituel dans une conversation en français. Le corpus Exolingue-Tsukuba montre en revanche une plus grande participation du LNN qui cumule à lui seul 5'30 de temps de parole sur un total de 17'28 de corpus. Cette participation plus active se manifeste par des prises de parole plus récurrentes et donc moins de postures d'écoute propice à l'usage des marques phatiques. Néanmoins, les résultats montrent tout de même une grande productivité de l'emploi de phatiques directement empruntés à la langue japonaise. Par ailleurs, on note à plusieurs moments l'occurrence de phatiques employés par le LNN lorsqu'il ne comprend pas le propos de son interlocuteur, mais ne souhaite pas l'interrompre. L'emploi des phatiques de cette catégorie se traduiront notamment par une posture d'écoute passive et généralement peu de réactivité au propos de l'interlocuteur :

Extrait 6 : Corpus Exolingue-Tsukuba [00:38:47]

Les locuteurs discutent des lois anciennes et absurdes qui perdurent dans différents pays du monde, mais qui n'ont plus lieu d'être, car devenues désuètes.

LocFFR1 : mais en france, il y a plusieurs lois comme ça où euh: on / elles servent plus à rien et-mais on les a oubliées du coup elles sont encore là

---

[115] *un* employé seul peut également constituer une marque phatique.

LocHJP : あー (***ah* allongé)**
LocHFR : C'est parce que en fait il est plus difficile de supprimer une loi
LocFFR1 : et ça prend du temps
LocHFR : que d'en inventer une nouvelle
LocHJP : うーん (***un* très allongé)**
LocFFR1: ça ça prend du temps: et du coup euh:

Dans cet extrait, on peut observer que le locuteur non natif va employer à deux reprises une marque phatique en réaction à ce que lui expliquent deux locuteurs natifs. La thématique semble un peu complexe puisqu'elle touche à une thématique juridique et de culture générale. Le locuteur japonais semble peu réactif et va demander des étayages lexicaux à plusieurs reprises avant et après cet extrait, car il ne comprend pas bien le contenu de la conversation. Nous pensons donc ici que le locuteur HJP emploie ces marqueurs phatiques comme une marque d'un manque de compétence conversationnelle. Toutefois, en utilisant ces marques, il va donner le signal à son interlocuteur que la conversation se déroule normalement, et ce de manière à ne pas l'interrompre. On peut y voir ici une fonction de temporisation similaire à ce que nous avons observé avec le rire. L'emploi des marqueurs phatiques permet au LNN de gagner du temps jusqu'à comprendre le contenu de la conversation, sans pour autant interrompre sans cesse son interlocuteur. Enfin, on observe que dans le cadre du corpus Exolingue-Grenoble, les marques phatiques semblent être employées de manière à « meubler » la conversation, permettant aux locutrices de se désengager de toute prise de parole en indiquant à l'interlocuteur qu'elles sont à l'écoute de leur prochaine prise de parole.

2.3. Un sous-emploi des marques discursives et interactionnelles

Dans cette section, nous explorons de quelle manière le sous-emploi des marques discursives et interactionnelles constaté dans notre étude quantitative peut constituer une trace pertinente pour le diagnostic de la compétence lexicopragmatique des locuteurs non natifs observés.

Polat (2011) montre la complexité et la grande utilité des *marques du discours et de l'interaction* (MDI), mais constate aussi que les LNN tendent à ne pas les utiliser de la même manière que les LN ni à la même fréquence. Romero Trillo (2002), Buysse (2017) puis Wolk *et al.* (2021) observent que les LNN montrent un usage limité des MDI avec notamment un sous-emploi des MDI fréquemment employés par les natifs. Rejoignant Polat (2011), Sahin Kizil (2021) démontre que les MDI ont une fonction cruciale dans la communication et

ajoute qu'ils jouent un rôle prédéterminant dans la pratique et le développement de la compétence lexicopragmatique des LNN en facilitant les opportunités d'interagir avec les LN. Le degré de maitrise des MDI semble donc fortement corrélé au degré de compétence lexicopragmatique du LNN et vice versa. Fung et Carter (2007), Gilquin (2008), Aijmer (2011) et Huang (2019) montrent que les MDI interpersonnels du type *so, you know, like, well* sont majoritairement sous-employés par les LNN alors que l'inverse se produit chez le natif. D'autres études se sont également penchées sur les moyens de substitution à l'emploi des MDI. Ainsi, De Cock (1999), Müller (2005), Gilquin (2008) et Dumont (2018) relèvent dans leurs études respectives que les LNN sous emploient les MDI, mais préfèrent en revanche employer des stratégies d'évitement ou de palliation en employant le silence, les marques phatiques ou les rires. D'autre part, Gilquin (2008, 2016) et De Cock (2019) montrent que les apprenants n'ayant pas complètement systématisé l'emploi des MDI dans la langue cible ont tendance à utiliser des marqueurs transférés de la LM.

Shirato and Stapleton (2007) montrent que les locuteurs japonais apprenant de l'anglais sous-emploient les MDI tels que *well*, *mind you, you see*, *as you know*, *I mean*, entre autres. Shimada (2014) observe quant à lui que la fréquence d'emploi des MDI est fortement corrélée à la progression de la compétence générale des locuteurs japonais en langue anglaise. Il constate également que plus la compétence générale des locuteurs est élevée, plus les emprunts à la langue maternelle se font moindres. Il remarque ensuite que les locuteurs japonais acquièrent plus vite l'usage des marqueurs ayant une fonction interactionnelle simple comme la confirmation ou le désaccord, mais utilisent beaucoup moins les marques interpersonnelles et/ou mettant en avant un processus subjectif complexe (2014 : 63). Miura (2011 dans Shimada 2014) comparant la fréquence des MDI utilisés par les japonophones apprenant l'anglais avec celle des LN anglophones, constate que les marqueurs les plus courants chez les natifs sont peu utilisés par les Japonais.

Lors de l'étude quantitative des données, nous avions remarqué une très forte disparité entre les emplois des MDI des LN et ceux des LNN en contexte de communication exolingue. Cette disparité s'est avérée encore plus forte avec le corpus endolingue :

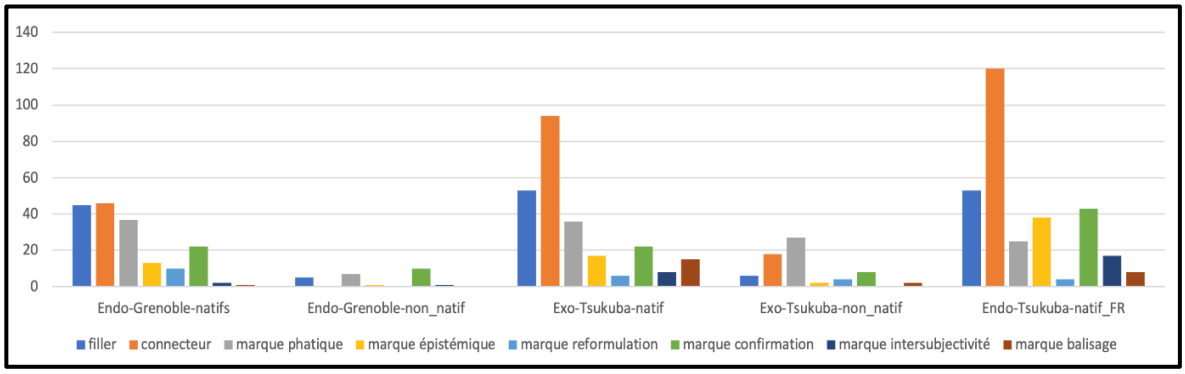


Figure 22 : Moyenne des fréquences brutes d'emploi des marques discursives et interactionnelles par groupe de locuteurs

Le graphique des fréquences brutes d'emploi des MDI par groupe de locuteurs ci-dessus montre de manière plus visible cette disparité et corrobore les résultats retrouvés dans les études de Shirato & Stapleton (2007) et Shimada (2011, 2014) en montrant qu'il y a effectivement une très faible représentation de l'emploi des MDI chez les LNN. Comme évoqué précédemment ce sont les marqueurs phatiques et les confirmations qui sont les formes les moins complexes à acquérir et donc privilégiées par les LNN. En revanche, les marques plus complexes (intersubjectives, épistémique, connectives) sont plus utilisées du côté des LN. On remarque également une très faible diversité des MDI employés par les LNN comparés aux locuteurs natifs qui emploient un plus large spectre d'expressions permettant d'exprimer des fonctions pragmatiques et interactionnelles plus complexes, exception faite du locuteur du corpus Exolingue-Tsukuba qui, pour les raisons évoquées dans les sections précédentes (exposition plus fréquente au français quotidien), semble varier un peu plus son registre de MDI que les locutrices du corpus Exolingue-Grenoble. On observe ainsi sur le graphique proposé que le locuteur non natif de Tsukuba emploie quasiment toutes les MDI annotées dans le corpus à l'exception des marques d'intersubjectivité. En effet, ces dernières constituent des structures dont l'emploi répond à une connaissance approfondie de l'interaction en français. On peut également émettre l'hypothèse que le contexte d'interaction peut avoir eu un effet sur la fréquence et la diversité des MDI employés. En effet, le contexte d'interaction du corpus prélevé à Grenoble semble se dérouler dans un cadre un peu moins enthousiasmant que le club de discussion de Tsukuba où les membres ont l'habitude de jouer à des jeux de plateau et de discuter de tout et de rien. Ce faisant, il nous semble que dans le premier cas, la modalité d'interaction était plus propice à l'écoute et la discussion consensuelle tandis que dans le second cas, la discussion nous est apparue plus dense et dynamique. Cette différence de cadre semble avoir eu une influence

sur la densité et la diversité des MDI, ce que nous pouvons observe dans le graphique plus haut.

Si l'on regarde l'emploi des *fillers* d'un peu plus près, on remarque que ces marques restent tout de même assez peu utilisées, car leur emploi nécessite à la fois une compréhension des stratégies inhérentes à la communication et une conscience approfondie de l'articulation interne de la prise de parole en français. La catégorie des connecteurs semble sous-employée par les locuteurs japonais alors qu'elle est utilisée fréquemment par les LN dans les corpus exolingues et deux fois plus employée dans le corpus endolingue. Rappelons que ces marqueurs permettent d'introduire ou d'articuler différents contenus dans le discours du locuteur sans pour autant entrainer de relation logique. Ce sont de petits mots dont la fonction n'est pas forcément indispensable sur le plan référentiel, mais qui permettent néanmoins de lier les éléments du discours du locuteur de manière à les rendre plus fluides et naturels tout en renforçant l'interactivité dans la conversation. On peut par exemple citer *du coup, voilà, quoi, déjà, au fait, genre* qui constitue autant d'expressions que les LNN ne semblent pas utiliser fréquemment dans les données que nous avons observées. L'emploi de ces connecteurs nécessite à la fois une certaine connaissance des dynamiques de la conversation en langue cible et une sensibilité à la gestion de son tour de parole qui ne peuvent s'acquérir de manière pérenne que par l'expérience de l'interaction en milieu écologique (Kasper & Rose 2002 : 195-196). De même, ce type de phénomène est très peu enseigné et ne fait pas l'objet d'un traitement adéquat dans les manuels d'apprentissage (Shimizu *et al.* 2007, Meléndez Quero 2010, Lorrillard 2018).

Les marques épistémiques sont également en sous-représentation chez les LNN comparés aux LN en interaction exolingue et aux LN en interaction endolingue. Le constat est le même pour les marques d'intersubjectivité ce qui rejoint Shimada (2011, 2014) qui montre dans ces deux études que ces MDI complexes sont faiblement représentés, car leur emploi nécessite une compétence pragmatique et conversationnelle avancée en langue cible. Un autre point intéressant est que dans le corpus Exolingue-Grenoble, les marques épistémiques et d'intersubjectivité employée par les LNN sont en nette diminution par rapport aux deux autres groupes de locuteurs natifs. Cela peut être notamment dû au fait que la communication était asymétrique et nettement moins interactive dans le corpus Exolingue-Grenoble. En effet, l'emploi de ces marques construit le lien dialogal et structure le rapport des interlocuteurs par rapport à l'énonciation. Or, l'interaction comportait en l'occurrence peu

d'échange et était très convergente. Les locutrices japonaises avaient adopté une posture d'écoute, avec une forte prépondérance des marques phatiques empruntées à la langue maternelle et peu de négociation du terrain de la parole. Cela s'est manifesté avec un très faible taux de chevauchement et un faible temps de parole. De ce fait, l'emploi de ces marques a fortement chuté, car le locuteur natif, qui a pris essentiellement la parole, a aussi eu tendance à adapter sa posture de locuteur en fonction du niveau de ses interlocutrices japonaises. On observe aussi une certaine productivité des marques de reformulation utilisées par les LN dans ce corpus par rapport aux autres corpus. En regardant de plus près le corpus, on s'aperçoit que ces marques de reformulation se retrouvent le plus souvent chez le locuteur HFR qui bénéficie de tours de parole bien plus longs que ses partenaires, ce qui favorise le recours à des reformulations diverses pour étayer son propos.

Enfin, tout comme Gilquin (2008, 2016) et De Cock (2019) nous avons pu observer que les locuteurs de notre corpus ont parfois tendance à utiliser des marqueurs empruntés de la langue maternelle lorsque la compétence en langue cible ne suffit plus à exprimer ce que veut dire le locuteur non natif :

Extrait 7 : Corpus Exolingue-Grenoble [00:04:23]

Les deux locuteurs natifs et les deux locutrices non natives entament une conversation sur les différences entre la cuisine française et la cuisine japonaise.

LocHFR: après on pourra discuter si vous voulez des différences entre la cuisine française et japonaise
LocFJP2 : ah oui oui / oui
LocFJP1: ah:
LocHFR : sur euh par exemple euh en france euh on utilise par exemple beaucoup de fromage ou de crême
LocFJP1 : うんうん **(*un un* phatique)**
LocFJP2 : うん**(*un* phatique)**
LocHFR : alors qu'au japon peut-être un peu moins
LocFJP1 : うんうんうん / うん うん **(*un* phatique)**
LocFJP1 : on utilise ブイヨン (*buiyon* = *bouillon*)
LocHFR: bouillon / oui beaucoup ouais
LocFJP1 : うんうんうん うん **(*un* phatique)**
LocFJP2 : bouillon et
LocFJP1 : et: on mange souvent うーんなんだろう (*un* allongé + *nan darô* = recherche lexicale + *comment dire. . .*) produit / うーんなんだっけ (*un* allongé + *nan dakke* = recherche lexicale + *comment on dit. . .*) うん / haricot / haricot

LocHFR : haricot / les légumes (intonation montante)

Dans cet extrait, nous remarquons des occurrences de transferts de marqueurs phatiques type *aizuchi* (*un un*), de marqueur d'hésitation (*un* allongé) ou de marqueur de recherche lexicale (*nan darô, nan dakke*) chez les deux LNN. Deux hypothèses nous semblent ici plausibles. La première, qui semble la plus évidente, est que ces emprunts à la langue maternelle sont des traces d'une compétence lexicopragmatique encore insuffisante pour exprimer la reformulation ou la régulation phatique en français. La deuxième hypothèse nous amène à penser que l'emploi de ces emprunts peut également être motivé par le fait que les LNN sont conscientes du niveau de compétence en langue japonaise de leurs interlocuteurs et semblent être plus flexibles sur le recours à la langue maternelle, car elles savent que cela ne constituera pas un problème de communication. De même, la présence d'une autre locutrice native du japonais semble avoir un effet sur la fréquence d'activation de la compétence en langue maternelle lors de la conversation.

2.4. Particularités structurelles, lexicales et pragmatiques de la variété non-native observée

Dans cette partie, nous allons examiner quelques anomalies concernant à la fois l'emploi du lexique, la structure de l'énoncé et la visée pragmatique de l'énoncé. Nous montrerons ici que le manque de compétence pragmatique est corrélé à un manque de compétence dans la structure formelle du langage et que ce faisant, un fort taux d'erreur d'emploi lexical ou de structures mal construites peut être un critère annonciateur de lacunes sur le plan pragmatique. Tout d'abord on peut remarquer que les lacunes dans la compétence lexicopragmatique se manifestent par des transferts d'expressions à fonction pragmatique depuis la langue maternelle :

Extrait 8 : Corpus Exolingue-Grenoble [00:00:10]

Les deux locuteurs natifs et les deux locutrices non natives entament une conversation ensemble dans le cadre d'un atelier de discussion libre.

LocHFR: C'est bon ça filme
LocFJP1 : うん　うん **(*un un* de validation)**
LocFJP2 : c'est bon
LocHFR : okay / bah merci à tous d'être venus

((Tout le monde rit))

LocHFR : merci Agathe / hum : on va parler un petit peu en français pendant une demi heure à peu près ça va aller pour tout le monde
LocFJP2 : ouais
LocHFR : euh : le but c'est qu'on puisse parler un peu de-de thèmes différents comme si on était à la maison sans trop de-de pression / euh : est-ce que vous avez des thèmes dont vous voulez parler / par exemple
LocFJP1 : うーん **(*un* allongé = recherche lexicale)** / なんだろう(*un* allongé + *nan darô* = recherche lexicale + *comment dire. . .*) / différent entre le japon et la france

On peut observer dans cet extrait que la locutrice non native emploie à deux reprises un transfert de marque discursive depuis sa langue maternelle et produit une erreur syntaxique. Le premier transfert constitue un emploi d'une marque phatique redoublée (cf. Section (2) suremploi phatique) qui pourrait être remplacé par une expression de validation telle que *ah oui* ou *ouais*. Le deuxième transfert est la formule *nan darô* utilisée pour marquer le processus cognitif de recherche lexicale ou conceptuelle et précédée de l'emploi de *un* allongé qui indique également un processus de recherche, expressions qui pourraient être remplacées en français par *euh comment dire*. Dans le dernier tour de parole, on remarque que la locutrice FJP1 produit une erreur lexicale en employant le terme *différent* au lieu du terme *différence*. Voici un autre exemple :

Extrait 9 : Corpus Exolingue-Grenoble [00:08:15]

Les deux locuteurs natifs et les deux locutrices non natives discutent à propos des traits culturels un peu bizarres en France ou au Japon.

LocFJP1 : par contre / うーん **(*un* allongé de recherche lexicale)** pour les Français
LocHFR: ouais
LocFJP1 : qu'est-ce que うーん **(*un* allongé de recherche lexicale)** qu'est-ce que vous pensez / ah non non non non non non il y a / il y a les : choses bizarres /
LocFJP2 : うん **(*un* phatique)**
LocHFR : au japon ((intonation montante))
LocFJP2 : au japon

Dans ce nouvel extrait, nous pouvons observer, en plus des transferts de marques de recherche lexicale ou de marques phatiques que l'on pourrait remplacer en français respectivement par *euh:* et *hum* et d'une erreur de déterminant, un emploi pragmatique assez curieux chez la locutrice FJP1. En effet, on observe que celle-ci va en plein milieu de son

tour de parole employer une interjection suivie de six répétitions de l'adverbe de négation *non*. Cet emploi qui nous semble ici erratique sur le plan pragmatique prend place à un moment où la locutrice tente de formuler une phrase puis essayer de reformuler celle-ci. Toutefois, voyant que la formulation ne correspond pas à ce qu'elle veut exprimer, elle va utiliser cette succession d'adverbes de négation pour marquer le fait que sa production ne lui convient pas. Cette succession ne semble pas tout à fait naturelle en français, ce qui nous amène ici à penser que nous avons un emprunt de la langue maternelle. Une expression plus appropriée en français serait l'emploi par exemple de *ah mince c'est pas ça / c'est pas ce que je voulais dire* ou *ah comment dire* qui permet d'exprimer la difficulté du locuteur à formuler un propos. Les lacunes repérées dans notre corpus peuvent aussi prendre la forme d'erreurs lexicales chez les locuteurs non natifs qui vont employer des formes empruntées à la langue maternelle ou une langue tierce, ou employer une variante erronée d'un terme français :

Extrait 10 : Corpus Exolingue-Grenoble [00:14:50]

Les deux locuteurs natifs et les deux locutrices non natives discutent à propos de la façon dont les Japonais perçoivent la France et notamment du choc fréquent entre les stéréotypes véhiculés sur la France et la désillusion quand les Japonais font l'expérience de la vie en France pour la première fois.

LocFJP1 : après / en particulier / j'aime **la culture cuisine**
LocHFR: la cuisine d'accord
LocFJP1 : cuisine
LocHFR : la gastronomie d'accord
LocFJP1: gastronomie
LocHFR : c'est bien (rire)
LocFJP1 : (rire) et / les japonais / ont / なんだろう(*nan darô* = recherche lexicale + *comment dire. . .*) / パリ　シンドロム (*pari shindoromu = syndrome de Paris*[116])
LocFJP2 : ouais (rire)
LocFJP1 : (rire) パリ　シンドロム
LocFJP2 : syndrome
LocHFR : ah syndrome de paris

[116] Syndrome psychologique transitoire constaté chez les personnes japonaises en réaction au choc culturel ressenti lors de l'expatriation en France, et plus particulièrement face au fossé entre représentations idéalisées de la France et réalité.

Dans cet extrait, on remarque d'une part une structure inhabituelle *la culture cuisine* où cuisine est vraisemblablement employé pour exprimer l'adjectif *culinaire*. On aurait pu également utiliser le syntagme prépositionnel *de la cuisine* pour exprimer l'idée d'une culture relative à la cuisine. On retrouve une nouvelle fois chez la locutrice FJP1 l'emploi de *nan darô* qui constitue un marqueur de recherche lexicale emprunté à la langue maternelle utilisée ici par la locutrice, car elle ne trouve pas l'équivalent français du terme japonais *pari shindoromu* (syndrome de Paris) qu'elle exprime en ayant recours au japonais. Elle va ensuite effectuer une relance sur la deuxième locutrice japonaise dans l'espoir que celle-ci puisse l'aider à pallier sa lacune lexicale. Toutefois, les locutrices semblent un peu incommodées par leur incapacité à trouver l'équivalent français, ce qui va se traduire par deux rires de temporisation alors que les locuteurs natifs, qui semblent éprouver quelques difficultés pour comprendre le thème de la conversation, ne vont esquisser aucun rire. La locutrice FJP2 va ensuite proposer le terme *syndrome* dont elle semble s'être souvenue, ce qui va tout de suite activer la compréhension du thème chez le locuteur HFR (*ah syndrome de paris*). On peut également trouver dans le corpus des erreurs de grammaire et de syntaxe qui là aussi peuvent être un indice intéressant pour cerner la compétence lexicopragmatique en langue cible :

Extrait 11 : Corpus Exolingue-Grenoble [00:12:22]

Les deux locuteurs natifs et les deux locutrices non natives discutent à propos des choses qui intéressent les locuteurs natifs dans la culture japonaise.

LocFJP2 : **qu'est-ce que tu : t'intéresse / t'intéresse le cul-la culture en japon culture japonais** ((intonation montante))
LocFFR: hum :
LocFJP2 : par exemple animé / manga
LocFFR : ha
LocFJP2 : **la culture**
LocFFR : euh : vraiment tout 'fin pas quelque chose en particulier de tout <u>la musique</u> les animés la culture
LocFJP1 : **うーん (*un* allongé = surprise modérée)**
LocFJP2 : **la musique**
LocFFR : **les musées** aussi
LocFFJP1 : **うーん (*un* allongé = surprise modérée) ((intonation expressive))**
LocFJP2 : **ah musées** ((intonation montante)) / へー (*he* allongé = *ah bon ?* = marque la surprise)

On observe dès le premier tour de parole que la locutrice FJP2 va produire un énoncé présentant des erreurs au niveau de l'emploi du pronom relatif, de l'emploi des prépositions et de l'emploi des déterminants. On constate également que les locutrices vont employer à nouveau des marques phatiques directement empruntées de leur langue maternelle (*un* allongé évoquant une surprise modérée). Un phénomène intéressant est ici les deux hétéro-répétitions (ou répétition diaphonique) *la musique* et *les musées* par la locutrice FJP2. Ce phénomène permet de marquer l'engagement de l'interlocuteur dans l'interaction en cours en reprenant un mot ou partie de l'énoncé précédemment produit par le précédent locuteur (Bigi *et al.* 2010). Ce phénomène est assez typique de l'interaction en langue japonaise (Nishizaka 1995 : 309) et permet d'une part de valider l'information entendue par le locuteur (Yokomori *et al.* 2018), mais permet aussi de faciliter la construction de l'interaction en cours (Machi 2019). Kobayashi & Hirose (1995) montrent que ce phénomène pragmatique fait souvent l'objet d'un transfert chez le locuteur japonophone apprenant de l'anglais qui va utiliser des hétéro-répétitions de manière bien plus fréquentes que les locuteurs natifs, ce que nous avons pu confirmer dans notre corpus et observer dans l'extrait ci-dessus pour les locuteurs japonophones apprenants du français observés dans notre étude. Enfin, on remarque ici l'emprunt d'une marque expressive d'étonnement (*he* allongé) qui pourrait être remplacé ici par les expressifs français de surprise *ah bon* ou *ah ouais*. Pour finir, voici un dernier exemple extrait du corpus pour illustrer cette section sur le diagnostic des anomalies structurelles et pragmatiques récurrentes :

Extrait 12 : Corpus Exolingue-Grenoble [00:16:05]

Les deux locuteurs natifs et les deux locutrices non natives discutent à propos villes françaises qu'ils apprécient tout particulièrement. Chacun partage ses expériences, bonnes ou mauvaises, vécues dans ces villes.

LocHFR :	et hum: euh: / vous êtes allés dans quelle ville en france vous êtes allés à paris
LocFJP2:	<u>à paris</u>
LocHFR :	à grenoble
LocFJP2 :	<u>à grenoble</u> / lyon
LocHFR :	lyon
LocFJP2 :	**うん (*un* confirmation)**
Loc HFR :	c'était bien lyon (intonation montante)
LocFJP2 :	oui
LocFJP1 :	**oui (allongé)**
LocFJP2 :	**oui j'aime je-j'aime ça**

LocHFR : tu aimes lyon (intonation montante)
LocFJP2 : **je-j'aime bien ça**
LocHFR: tu préfères lyon/ ou grenoble

Dans un premier temps, nous remarquons que la locutrice FJP2 va employer deux hétéro-répétitions en répétant les villes proposées par le locuteur HFR. Nous observons ensuite une occurrence d'emprunt à la langue japonaise d'un marqueur de confirmation *un* employé comme une confirmation en réaction à la relance *lyon* du locuteur HFR qui elle-même est faite en réaction à la proposition de la ville de Lyon de la locutrice FJP2 qui réagit, car elle voulait parler de cette ville qu'elle a visitée récemment. Ce marqueur de confirmation emprunté du japonais pourrait être remplacé par *ouais* ou *c'est ça* dans un contexte similaire en français. Un autre phénomène intéressant est cet emploi de la marque de confirmation *oui* employé avec une intensité et une longueur inhabituelle qui tend à rappeler un type de marque confirmation franche employée fréquemment en japonais oral. On peut donc se demander ici si nous n'avons pas affaire à un transfert de la langue maternelle. En effet, l'allongement de la marque de confirmation à une valeur de renforcement de l'adhésion en japonais. Enfin, nous observons un emploi assez inhabituel de la locutrice FJP2 en réponse à la question *c'était bien lyon* du locuteur HFR. Celle-ci va répondre à cette question *oui j'aime je-j'aime ça* ce qui semble constituer une erreur pour plusieurs raisons. D'une part, sur le plan pragmatique, l'expression *j'aime ça* peut avoir un sens un peu tabou en fonction du contexte et ne semble pas convenir à cette situation. D'autre part, sur le plan référentiel, l'emploi du pronom démonstratif indéfini *ça* (Maillard 1994 ; Corblin 1994 ; Rouget 2003 ; Guérin 2020) pause ici problème. En effet, le thème de la conversation en cours est manifeste (la ville de Lyon) et ne nécessite pas l'emploi d'un démonstratif pour désigner cette entité. De plus, *ça* étant un pronom indéfini qui est justement employé pour faire référence à un antécédent global et généralisant (une situation, un état des choses, un fait complexe), il nous semble ici que cet emploi est erroné, car il créer une ambigüité référentielle sur ce dont il est question. La locutrice a par exemple la possibilité dans cette situation de communication d'employer la structure *oui j'aime bien Lyon*.

Nous avons pu observer au travers de ces divers exemples extraits du corpus diagnostique que la production des LNN B1 CECRL s'accompagne fréquemment d'erreurs structurelles, pragmatiques et lexicales qui semblent être révélatrices d'un état de la compétence lexicopragmatique encore instable et dépendant des emprunts à la langue maternelle. Notre observation selon laquelle la présence de ces erreurs constitue des traces empiriques de

difficultés d'emploi permettant de diagnostiquer la compétence lexicopragmatique du locuteur japonophone du français semble ici être pertinente. Dans la partie qui suit, nous allons synthétiser les différentes analyses que nous avons faites sous forme d'une synthèse diagnostique qui nous servira de constat de départ pour le deuxième chapitre consacré à l'étude de la compétence lexicopragmatique de locuteur japonophone du français de niveau C1/C2.

3. Diagnostic et conclusion

3.1. Synthèse diagnostique

Dans un premier temps, l'analyse quantitative des résultats nous a permis de mettre certains suremplois ou sous-emplois en évidence en contrastant les corpus exolingues avec le corpus endolingue. Nous avons pu remarquer que globalement, il y a une très large différence dans les fréquences d'emploi des MDI, ainsi que pour les marques du registre oral. Les LNN ont de fortes propensions à sous-employer les marques typiques de l'oral fréquemment retrouvées dans le corpus endolingue. Sans grande surprise, la fréquence d'emploi des PhPex était très faible chez les LNN, et la plupart des marques expressives que nous avons relevées étaient faiblement lexicalisées, soit des rires des réactions vocales. Nous n'avons donc pas pu observer des spécificités d'emploi particulières des PhPex chez les LNN, si ce n'est un fort sous-emploi. Cette sous-représentation peut provenir du fait que le protocole employé, qui repose sur une conversation non-élicitée, n'est pas particulièrement propice à l'émergence de ce type d'objet chez les locuteurs de niveau B1/B2. Il aurait peut-être été plus souhaitable de cadrer un peu plus l'interaction afin de susciter l'emploi de ce type d'expression.

Par ailleurs, nous avons pu remarquer une différence intragroupe significative entre les résultats des locutrices japonophone du corpus Exolingue-Grenoble et le locuteur japonophone du corpus Exolingue-Tsukuba. En effet, le locuteur du corpus Exolingue-Tsukuba semble employer plus de formes spécifiques à l'interaction orale et son comportement au sein de l'interaction montre une certaine fluidité. Nous avons observé chez le LNN de Tsukuba des chevauchements et des prises de parole plus fréquentes et une fréquence d'usage sensiblement plus élevée des MDI. Cette différence intragroupe semble s'expliquer par la différence du contexte d'acquisition et d'usage du français, mais aussi par

l'objectif d'apprentissage. En effet, en regardant les données personnelles des LNN enregistrés, nous nous sommes aperçu que le locuteur du corpus Exolingue-Tsukuba avait bénéficié au cours de son apprentissage d'une exposition fréquente et diversifiée à la conversation quotidienne avec des locuteurs natifs, notamment dans le cadre du club de discussion « Francophonie » et dans des cours de conversation dispensés dans le cadre de sa formation universitaire. Les locutrices du corpus Exolingue-Grenoble en revanche ne semblent pas bénéficier d'une exposition aussi diversifiée que celle du locuteur Exolingue-Tsukuba. En effet, d'après nos informations, ces locutrices sont arrivées en France dans le cadre de leurs études (échange universitaire) et n'ont pas eu l'occasion au moment de l'enregistrement de s'exposer pleinement à la conversation quotidienne avec des LN. En outre, une bonne partie des français avec qui elles interagissent sont des étudiants en LEA anglais-japonais, ce qui a pu favoriser le recours au japonais comme *lingua franca*[117]. Les sources d'exposition à la langue française de ces deux locutrices ont été principalement les cours en français de l'université japonaise, cours donnés en variété standard et ne mettant pas l'accent sur la dimension conversationnelle du français. Cette différence intragroupe semble confirmer la justesse de l'hypothèse formulée dans le cadre théorique de ce chapitre selon laquelle la qualité de l'exposition à la langue française écologique et diversifiée peut avoir des effets notables sur la compétence lexicopragmatique de l'apprenant.

Notre étude quantitative a également permis de mettre en évidence 4 points d'attention pertinents pour le diagnostic de la compétence conversationnelle de l'apprenant japonophone. Ces 4 points d'attention constituent selon nous des traces empiriques et concrètement observables d'une compétence pragmatique encore instable chez le locuteur japonophone du français.

L'emploi du rire comme stratégie de contournement et de palliation

Nous avons observé que le rire n'était pas seulement employé de manière empathique ou pour exprimer de l'amusement[118]. Il est également utilisé comme stratégie de contournement et d'atténuation lors de situation d'interaction mettant en insécurité linguistique le LNN.

---

[117] En revanche, l'anglais est très peu utilisé. On pourrait supposer que la distance typologique entre l'anglais et le japonais fait que dans ce type de configuration exolingue, ce sera le japonais qui sera privilégié car connu par tous les locuteurs.

[118] Auquel cas ces rires sont généralement plus longs, plus intenses et plus expressifs, tant sur le plan facial que sur le plan intonatif.

Nous nous sommes également aperçus que les LNN de notre corpus utilisaient fréquemment le rire alors que la situation ne semblait pas évoquer de trait d'humour de la part du LN. L'emploi de ce rire présentait une fréquence plus élevée chez les LNN que les LN. Nous avons alors émis l'hypothèse d'un transfert de la langue maternelle d'une forme de rire à fonction sociale fréquemment employé par les locuteurs japonais. Cet emploi du rire propre à la langue japonaise a notamment la fonction d'ajustement social et de marque de convergence interactionnelle. Nous avons finalement remarqué que le rire était également employé pour combler les interruptions de la conversation durant lesquelles les locuteurs non natifs ne semblaient pas enclins à prendre la parole ou pour temporiser en cas d'incompréhension pour avoir le temps d'accéder au sens de la conversation. Le rire, lorsqu'il n'est pas associé à une situation cocasse, peut donc ici constituer une trace tangible qui peut tantôt être le reflet d'une stratégie culturellement dominante de convergence interactionnelle, tantôt constituer une stratégie palliative face à une lacune lexicopragmatique.

Suremploi des formes phatiques et transfert de la langue maternelle

Nous avons constaté le suremploi des formes phatiques et notamment des formes directement empruntées à la langue maternelle du LNN. En effet, les LNN observés dans notre corpus utilisent en moyenne 5 fois plus de marques phatiques que les LN, dont plus de 90% sont des transferts de la langue maternelle. Cette surabondance des marques phatiques semble être liée à la forte propension des locuteurs japonais à viser l'harmonie et l'écoute plutôt que la négociation. Cet éthos communicationnel se retrouve fortement dans les productions des LNN en français. Par ailleurs, nous avons observé une différence dans l'emploi de ces marques phatiques entre les deux groupes. Alors que les LN ont tendance à attendre que le locuteur ayant la parole termine son énoncé pour placer une marque phatique, les locuteurs non natifs ont tendance à utiliser abondamment les marques phatiques au fur et à mesure que l'interlocuteur parle, reproduisant ainsi l'emploi des marques phatiques en japonais. À l'instar de Cutrone (2005) dont les travaux ont essentiellement porté sur l'usage des marques phatiques chez les locuteurs japonais apprenant de l'anglais, nous avons pu observer que la surabondance de ces marques peut revêtir plusieurs fonctions : convivialité, palliation et temporisation. En effet, les données nous ont permis de constater que ces fonctions semblent employées afin de pallier le manque de compétence en langue cible, pour gagner du temps par manque de compréhension de ce qui est dit ou tout simplement garder

la conversation la plus harmonieuse et convergente possible. L'analyse des suremplois des marqueurs phatiques chez les LNN a donc permis de mettre en évidence une preuve concrète des difficultés rencontrées sur le plan lexicopragmatique.

Un sous-emploi des marques discursives et interactionnelles

Nous avons observé de plus près le sous-emploi des MDI chez les LNN. Nous nous sommes aperçus que ce sous-emploi ne concernait que les marques complexes servant à marquer l'intersubjectivité, à connecter les différents éléments (connecteur) de l'interaction entre eux et à moduler son assertion (marque épistémique), marques qui requièrent une compétence pragmatique et discursive approfondie en langue cible. De nombreux travaux sur l'emploi des marques discursives et interactionnelles montrent que la fréquence d'usage est corrélée proportionnellement au degré de maitrise en langue cible (Andersen 2011, Müller 2005) et que de ce fait, un sous-emploi semble être caractéristique d'un développement insuffisant de la compétence lexicopragmatique chez les LNN (Shimada 2014). Ce sous-emploi des MDI pourtant nécessaire pour fluidifier et rendre naturelle la conversation peut avoir un effet sur le comportement interaction du LN et donner une impression de communication peu naturelle. Tout comme Gilquin (2008, 2016) et De Cock (2019), nous avons pu remarquer que les LNN de notre corpus avaient tendance à utiliser des MDI empruntés à leur langue maternelle pour pallier le manque en langue cible. Nous avons remarqué dans la suite de notre analyse que les marqueurs phatiques et les confirmations étaient toutefois employés par les LNN, car elles semblent être des marques dont la maitrise semble plus accessible à un niveau intermédiaire. Nous avons observé chez les LNN peu de routines interactionnelles (*tu vois, tu sais, du coup*) ou de reformulations (*enfin, je dirais, je sais pas moi*) fréquemment employées chez les locuteurs natifs du corpus endolingue. Le lexique des locuteurs japonais est essentiellement constitué d'un contenu référentiel et explicite avec pas ou peu de sous-entendus ou de dimensions pragmatiques. Enfin, l'annotation puis la quantification des marques spécifiques au registre de l'oral nous a permis de constater que les LNN de niveau B1/B2 de notre corpus employaient extrêmement peu d'expressions spécifiques au registre familier, à l'inverse de locuteurs natifs. Nous avons cependant remarqué que lorsque le LNN tentait d'employer une tournure familière, cet emploi résultait souvent en échec pragmatique. Nous noterons par exemple l'emploi de l'expression *t'es conne* au lieu de *t'es bête* par le locuteur de Tsukuba, emploi qui semble trop familier et inapproprié étant donné que les interactants ne sont pas assez intimes pour permettre l'usage d'un terme aussi dépréciatif.

Des erreurs structurelles, pragmatiques et lexicales caractéristiques d'une compétence lexicopragmatique en cours de stabilisation

Le dernier point d'attention observé dans cette étude diagnostique concerne un ensemble d'erreurs structurelles (morphosyntaxique, formulation et grammaire), pragmatiques (usages impropres, transferts) et lexicales (absence du mot idoine, termes inadaptés, calques). En effet, l'étude quantitative des marques d'anomalie dans les corpus exolingues montre que les LNN avaient tendance à réaliser à la fois de nombreuses erreurs d'ordre structurelles et à faire des transferts de la langue maternelle qui engendraient des emplois pragmatiques inhabituels. Ainsi, l'analyse qualitative des données a permis de montrer la présence d'erreurs grammaticales et lexicales assez spécifiques au niveau B1 (flexions verbales, accords en genre et en nombre, utilisation des prépositions erronées, absence de déterminants). Des erreurs de construction syntaxiques sont également très présentes. Le discours non-natif est souvent disfluent et comporte de nombreuses traces de reformulation et de discontinuité des phrases. En outre, on observe parfois un ordre des constituants agrammatical et ayant un impact sur la compréhension du sens. L'étude de ces marques dans notre corpus diagnostique nous a permis de constater que les anomalies structurelles et lexicales semblaient aller de pair avec une compétence lexicopragmatique en cours de stabilisation. De même, nous avons observé que certains emplois pragmatiques inhabituels étaient le produit de pratiques en LM projetées dans le système de la langue cible, montrant ainsi que certaines pratiques conversationnelles en français ne sont pas encore pleinement acquises.

3.2. Conclusion de la première phase du protocole expérimental

À l'issue de cette première phase, nous avons pu observer les spécificités de la compétence lexicopragmatique de locuteurs japonophones du français de niveau A2/B1 CECRL en condition d'interaction exolingue faiblement élicitée. L'objectif de ce chapitre était de dresser un premier constat de la compétence lexicopragmatique et notamment de l'emploi des structures expressives de la conversation quotidienne chez des locuteurs japonophones en nous basant sur des marques linguistiques concrètes retrouvées dans un corpus d'interaction multimodale. L'analyse des données sur les plans quantitatifs et qualitatifs nous a permis de constater les éléments suivants :

- Les LNN de niveau intermédiaire emploient moins de MDI que les LNN dans des conditions similaires d'interaction. Les MDI les moins utilisés sont ceux impliquant le niveau intersubjectif (marqueurs épistémiques, marqueur d'attention, marques d'approbation, etc.) ou le niveau discursif (reformulation, étayage, mise en relief du propos, etc.).

- Les LNN de niveau intermédiaire emploient moins de PhPex que les LNN dans des conditions similaires d'interaction. Ce stade transitoire de développement de la compétence lexicopragmatique se traduit notamment par une surabondance des réactions expressives vocales et des rires de connivence à des endroits de la conversation propices à l'usage des PhPex.

- L'usage important de stratégies palliatives (rire, transfert, calques, surabondance de MDI simples telles que les phatiques ou les marques d'accord) semble constituer des indices fiables du diagnostic d'une lacune lexicopragmatique chez les LNN de niveau intermédiaire.

- Une des difficultés caractéristiques des LNN de niveau intermédiaire de notre corpus est la difficulté 1) du choix de l'illocutoire visé, 2) de l'adéquation du choix lexical (et de sa combinatoire) avec l'illocutoire visé (cf. par ex. l'usage de *j'aime ça* par l'une des locutrices du corpus Exolingue- Grenoble, p. 202).

- L'attitude du LNN et son engagement dans l'interaction et dans son acquisition est un facteur qui semble favoriser le développement de la compétence lexicopragmatique.

Bien que perfectible, cette approche a le mérite de proposer une première tentative de diagnostic de la compétence pragmatique d'apprenants japonais du français à l'aide de l'annotation d'une interaction orale libre et d'une analyse qui mélange à la fois approche quantitative et qualitative. L'observation de ce type d'objet linguistique à l'aide d'un protocole faiblement élicité est relativement complexe. En effet, l'observateur n'a accès ni aux données concernant la pratique du locuteur ni à son état d'esprit ou aux processus mentaux qui sous-tendent l'emploi des PhPex. L'autre problème constaté est qu'un

phénomène linguistique à priori très productif à l'oral peut ne pas apparaitre en quantité satisfaisante dans le corpus de l'analyse, d'autant plus s'il s'agit de production écologique et peu élicitée. Face à ce problème, nous avons choisi d'opérer notre diagnostic en prenant en compte d'autres catégories d'unités à fonction pragmatique et plus particulièrement des catégories plus larges de formes verbales à fonction expressive.

Cette première phase de notre étude a permis de formuler un certain nombre d'observations concernant le développement de la compétence lexicopragmatique et des difficultés d'usage des constructions pragmatiques chez les LNN japonophones. La deuxième phase de ce travail de recherche vise à étudier plus finement l'emploi des PhPex chez des locuteurs japonais du français de niveau CECRL C1/C2. Nous nous baserons sur une approche statistique puis qualitative. Nous adopterons également une approche contrastive sur des données en japonais obtenues auprès de nos informateurs à l'issue d'une activité de production.

Deuxième phase du protocole expérimental :

Test de la compétence lexicopragmatique en production et en compréhension de locuteurs japonophones du français de niveau avancé

Point étape 1 du protocole expérimental

Le chapitre précédent a constitué une première étape diagnostique préliminaire qui nous a permis de poser les hypothèses de travail qui serviront de base à l'analyse proposée dans ce chapitre. Ces hypothèses sont les suivantes :

1. Les LNN de niveau avancé produisent moins de séquences préfabriquées que les locuteurs non natifs et rencontrent plus de difficulté en production qu'en compréhension lors de la conversation quotidienne.
2. Une compréhension approfondie des PhPex n'implique pas nécessairement leur utilisation appropriée.
3. Les difficultés s'observent aux trois niveaux de la compétence lexicopragmatique (pragmalinguistique, sociopragmatique et métapragmatique)
4. Le développement de la compétence lexicopragmatique est influencé par la LM au niveau du choix du lexique, des stratégies de compréhension et de production, mais aussi de l'attitude communicationnelle transmise dans la production.
5. La mesure outillée de la compétence lexicopragmatique, en se fondant sur des traces linguistiques concrètes et collectées de manière rigoureuse dans la production des LNN, permet de caractériser les spécificités lexicales et pragmatiques de leur emploi des PhPex.

À l'aide d'un protocole expérimental fondé sur la transcription et l'annotation d'un corpus d'oral spontané, nous avons proposé une analyse à la fois quantitative et qualitative dans la phase 1 de notre expérimentation. Cette analyse s'est voulue la plus rigoureuse et la plus empirique possible, basée sur des traces linguistiques concrètes. Une des premières constatations qui a émergé de ce travail d'analyse est que la mesure de la compétence lexicopragmatique et de sa composante lexicale pouvait s'avérer complexe. En effet, son analyse implique de nombreux paramètres linguistiques, extralinguistiques et cognitifs auxquels nous n'avons pas accès dans nos données. Nous nous sommes donc essentiellement focalisé sur les phénomènes de sous et de sur représentation dans l'emploi de marques pragmatiques et lexicopragmatiques caractéristiques de la conversation orale, ainsi que sur les phénomènes d'emprunt ou de transfert de la LM. D'une part, les sous-emplois des formes pragmatiques attendues (marqueurs discursifs et PhPex, entre autres) et les suremplois des formes empruntées à la LM sont révélateurs d'un répertoire lexicopragmatique restreint ne

permettant pas au locuteur non natif de varier sa production ou de proposer des expressions pragmatiques plus adaptées. À défaut, le LNN tend à se rabattre sur des emplois « passe-partout » ou sur des formulations faiblement idiomatiques (ne présentant aucun figement sémantique ou combinatoire). D'autre part, nous avons pu constater que les transferts et les emprunts de la LM étaient très fréquents et avaient notamment une fonction palliative permettant au locuteur non natif de combler son manque lexical. Le niveau A2/B1 semble donc constituer un niveau transitionnel dans lequel le LNN développe des habitus métapragmatiques lui permettant d'identifier des situations de communication spécifiques, mais ne possède pas encore un répertoire lexical assez étendu pour réaliser l'emploi lexicopragmatique attendu.

Dans cette seconde phase, nous observons plus spécifiquement la compétence lexicopragmatique sollicitée dans la compréhension et l'emploi des PhPex. Nous observerons cette compétence chez des locuteurs japonophones avancés de niveau fin B2 à C2 à l'aide d'un test alliant compréhension et production. Alors que l'étude diagnostique du chapitre précédent s'articulait autour de la compétence lexicopragmatique et conversationnelle de manière générale, cette deuxième phase de notre thèse va se concentrer essentiellement sur la capacité du locuteur non natif à produire un PhPex approprié, à le comprendre ou à l'interpréter en fonction du contexte donné. Nous articulerons donc prioritairement notre réflexion sur les savoirs linguistiques et les savoirs procéduraux mobilisés dans l'emploi et la compréhension des PhPex. Ce test se compose de 5 activités élaborées à partir des méthodes d'élicitation et d'évaluation de la compétence pragmatique évoquées dans le deuxième volet du cadre théorique (Chap. 7). Chaque activité vise à tester l'une des trois sous-compétences de la compétence lexicopragmatique (Chap. 6, section 3). Nous proposons ci-dessous un schéma de la structure du test (Annexe 8) qui permettra au lecteur de se représenter plus facilement la logique de construction de cette deuxième phase de notre expérimentation :

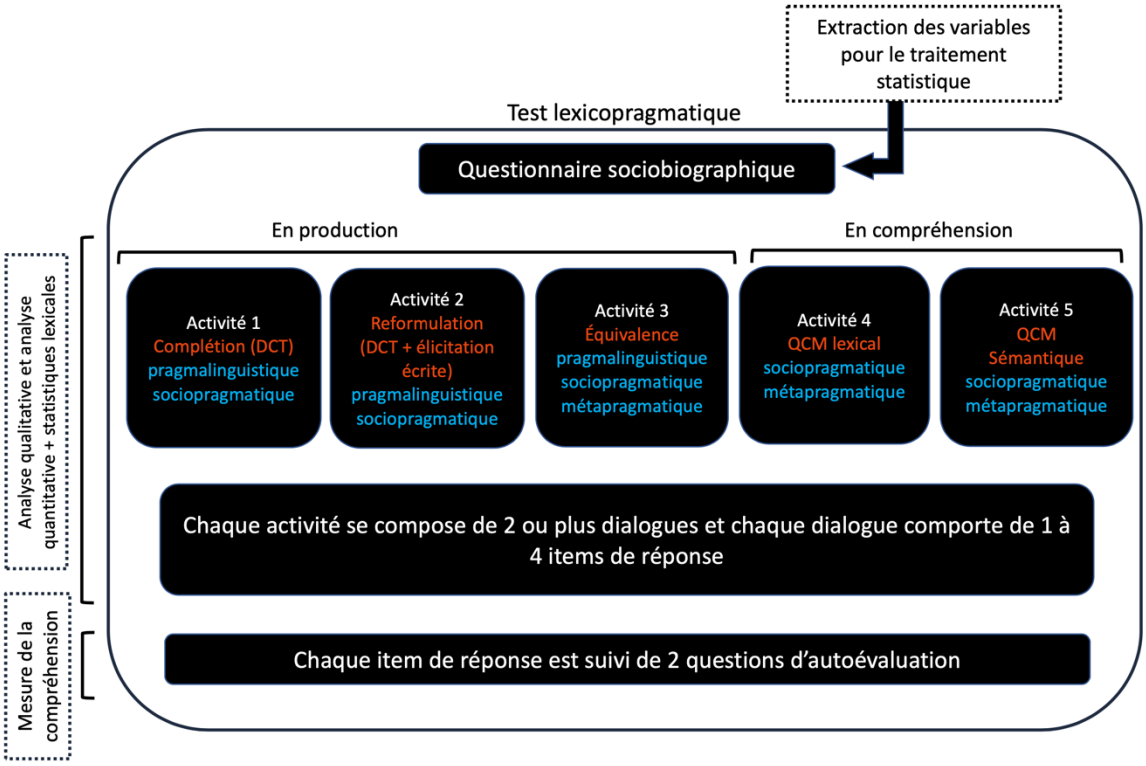


Figure 23 : Architecture du test lexicopragmatique

Dans une première partie, nous évoquons les données orales exploitées dans notre test, les spécificités des informateurs ayant participé à notre étude, ainsi que le dispositif méthodologique mobilisé lors de la construction et de la passation du test lexicopragmatique. Dans une seconde partie, nous proposons une analyse quantitative et statistique des résultats en fonction de plusieurs variables liées soit aux conditions d'apprentissage soit aux conditions d'usage du français. Nous procédons ensuite à une analyse qualitative des réponses obtenues en contrastant les réponses des LN avec les LNN. Enfin, nous ferons synthèse des résultats que nous discuterons.

Chapitre 10 – Cadrage expérimental du test lexicopragmatique

L'élaboration de ce test lexicopragmatique part du constat émis dans le chapitre précédent que l'acquisition et la maitrise des PhPex nécessitent une compétence linguistique supérieure au niveau B1 et une exposition à la conversation quotidienne plus fréquente. Nous avons également formulé en introduction de cette seconde phase quelques problématiques que nous souhaitions explorer. En nous fondant sur ces différents éléments, nous avons focalisé nos efforts sur la création d'un test linguistique qui nous permettrait de vérifier différents aspects propres à la compétence lexicopragmatique et l'usage des PhPex en situation de communication quotidienne, tout en articulant la compétence de compréhension et de production.

1. Les informateurs

Les informateurs se divisent en deux sous-groupes : un premier sous-groupe de locuteurs natifs du français (groupe contrôle) et un deuxième sous-groupe de locuteur japonais du français de niveau avancé (groupe test). Concernant le choix des locuteurs pour le groupe test, nous avons focalisé notre attention sur trois spécificités : le niveau de certification, la longueur du séjour en France et la réalisation d'un cursus universitaire en français. Nous avons ainsi pu réunir une cohorte de 35 locuteurs japonophones de niveau B2 à C2, diplômés d'un cursus en lien avec la langue et la civilisation française ou ayant étudié en France, et ayant séjourné dans l'Hexagone au moins 6 mois. En outre, nous avons privilégié les locuteurs ayant totalisé au minimum 5 ans d'apprentissage très régulier du français de manière à tester notre hypothèse de l'effet du développement de la compétence générale sur la compétence lexicopragmatique en LC. Le niveau d'étude de ce sous-groupe s'étale de la licence[119] au doctorat[120], avec environ la moitié des informateurs insérés dans un cadre professionnel où ils sont amenés à manier la langue française ou interagir avec des locuteurs natifs du français au quotidien. L'autre moitié se compose de locuteurs évoluant dans le milieu académique soit en tant que doctorants, soit en tant qu'enseignants. Le groupe contrôle totalise également 35 informateurs et se compose essentiellement de locuteurs natifs

[119] Licence acquise au moment du test.
[120] Doctorat acquis ou en cours d'acquisition au moment du test.

étudiants en licence (cursus LEA) ou en master (cursus sciences du langage). Les deux cohortes ont été soumises au même test, mais dans des conditions différentes du fait de la crise sanitaire qui est survenue au moment du recueil des données chez les LNN (point détaillé plus bas). Ci-dessous un rapide tableau récapitulatif[121] :

	LM	Effectif	Protocole	Niveau CECRL	Niveau d'étude	Critères de sélection
Groupe Test	Japonais	35	Distanciel	B2 à C2	De la licence au doctorat	Min. 5 ans d'étude du Français Min. 6 mois de séjour en France Niveau avancé Cursus universitaire en français
Groupe contrôle	Français	35	Présentiel	Natifs	De la licence au master	Étudiant à l'université Locuteurs natifs uniquement

Tableau 7 : Spécificités du groupe test et du groupe contrôle

La constitution de notre groupe test a été complexe. En effet, réunir une trentaine de locuteurs de niveau très avancé qui plus est ayant une langue maternelle typologiquement éloignée du français telle que le japonais a été chronophage. L'enseignement du français n'est pas aussi répandu au Japon que l'anglais, et de ce fait, peu d'institutions proposent des formations de français de niveau avancé. Nous avons dû passer par plusieurs campagnes de recrutement en sollicitant notamment les réseaux universitaires afin d'obtenir des contacts avec d'anciens étudiants ayant fait carrière en France ou avec des doctorants effectuant leurs travaux de recherche sur la langue française. En outre, le début de la crise sanitaire et le passage du test lexicopragmatique à distance qui s'ensuit a eu pour effet de faire émerger des contraintes de trois ordres :

- Des contraintes techniques liées au fait que nous n'étions pas équipé ni formé à ce genre d'expérimentation médiée.
- Des contraintes humaines notamment liées au retour précipité au Japon de certains des informateurs, au décalage horaire ou aux désistements pour raisons personnelles.
- Des contraintes liées à l'outil en lui-même : déconnexions intempestives, fatigue liée à ce type de communication nécessitant une concentration accrue, les interlocuteurs ne disposaient pas forcément du matériel ou du logiciel de communication adapté.

[121] Un tableau concernant les métadonnées individuelles pour chaque LNN est disponible dans l'annexe 6.

2. Le test lexicopragmatique

2.1. Étapes d'élaboration du test

La première étape d'élaboration de ce test a été le choix des expressions à traiter que nous avons effectué à l'aide d'une liste de phraséologismes oraux réactifs de l'oral établie dans le cadre du Projet Polonium *Pragmalex* (Krzyżanowska, Grossmann et Kwapisz-Osadnik 2021). Ce choix s'est notamment porté sur le motif de la colère, de l'indignation et du reproche. Ces différents sentiments pourront être observés de deux points de vue : 1) un sentiment déclenché par l'interlocuteur et dirigé directement contre lui, 2) un sentiment déclenché par un stimulus externe autre que l'interlocuteur, mais exprimé à celui-ci, par empathie notamment. Nous voulions essentiellement observer des expressions suscitant une réaction émotive forte à polarité négative, car plus complexe à employer par les LNN qui sont généralement peu amenés à exprimer un affect de ce type en français. En outre, ces motifs émotionnels se retrouvant généralement dans les deux langues-cultures, nous voulions observer les possibles interactions entre la LM et la LC.

Nous avons ensuite mené une réflexion sur les compétences linguistiques que nous souhaitions observer au moyen de ce test. Parmi les questions qui ont orienté la conception de ce protocole, celle des mécanismes sur lesquels reposent la compréhension et la production nous a paru centrale. Pour ce qui est de la compréhension, nous nous sommes demandé sur quels indices présents dans la situation d'énonciation ou dans le répertoire linguistique du LNN repose le décodage du sens et de la fonction du PhPex. Pour la production, nous avons essayé de réfléchir sur une manière d'explorer les choix lexicaux et combinatoire, ainsi que les paramètres agissant sur cette sélection. Sur le plan méthodologique, nous avons mené une étude des différents protocoles d'élicitation de la compétence pragmatique afin de sélectionner celles qui nous seraient utiles pour cette deuxième phase (cf. Partie cadre théorique, Chap. 7).

Après avoir ciblé l'objet de notre analyse, nous avons réfléchi au type de donnée à employer. Dans le cas de ce type de protocole exploratoire, deux choix semblent possibles : l'utilisation de données fabriquées (plus simple à constituer, permettant de tester une plus grande variété de situations, mais ne testant pas la compétence en conditions réelles) ou l'utilisation de données authentiques (faible disponibilité et adaptabilité, qualité sonore très variable, débit

rapide, chevauchement et discontinuités fréquentes, mais permettant de tester la compétence en conditions réelles)[122]. Nous avons finalement opté pour une solution à mi-chemin entre l'authentique et le fabriqué. Nous nous sommes basé sur la structure globale de dialogues extraits du corpus CLAPI (Baldauf-Quilliatre *et al.*, 2016) qui contenaient les PhPex ciblés dans notre étude. Nous avons ensuite remanié ces données authentiques afin d'obtenir des canevas plus adaptés aux spécificités de notre étude, puis nous avons enregistré ces canevas avec des locuteurs natifs afin d'assurer des audios de bonne qualité et ayant un débit de parole adapté. En effet, nous avons pris le parti ici non pas de tester la capacité des locuteurs à segmenter et comprendre les différentes unités de la chaine parlée en condition réelle. Nous nous sommes principalement focalisé sur la compétence de compréhension et de production du PhPex en réaction à des conditions de production habituelles. L'authenticité parfaite des enregistrements ne nous a donc pas semblé cruciale dans le cadre de cette étude[123].

---

[122] Nous renvoyons le lecteur au deuxième volet de la partie théorique et notamment au dernier chapitre consacré à la mesure de la compétence pragmatique pour une discussion sur les avantages et inconvénients des différentes méthodologies.

[123] Tout comme nous, Wang (2021) remarque que les données brutes d'un corpus oral ne sont pas adaptées à une utilisation des LNN du fait de la qualité audio variable et du débit parfois très rapide et peu articulé des production de LN. De ce fait, elle a procédé à tout un travail d'adaptation des données (nettoyage et ralentissement des bandes audios, sous-titres guidés, traitement des vidéos) afin de les rendre exploitable dans le cadre de son étude la compréhension de quelques phraséologismes pragmatiques de l'oral chez des locuteurs sinophones de niveau avancé.

2.2. Structure et spécificités du test

Le test employé pour explorer les spécificités de la compétence lexicopragmatique des locuteurs japonophones de niveau avancé en japonais se divise en 5 activités (A) comprenant chacune plusieurs dialogues (D) dans lesquels se trouvent un ou plusieurs items (I) qui constituent les PhPex qui sont traités dans le cadre de l'activité par les répondants :

			Expression testée	Tps réponse	Type d'activité	Compétence visée[124]
Activité 1	D1	I1	C'est + [réponse au choix]	1 min	**Activité de complétion** dans laquelle il est demandé à l'informateur de continuer un extrait de dialogue en employant ses connaissances linguistiques.	Compétence de production. L'informateur mobilise sa compétence pragmalinguistique et sociopragmatique pour continuer un dialogue en cours de déroulement.
	D2	I1	Sérieux ? + [réponse au choix]			
Activité 2	D1	I1	C'est le pied !	40 sec	**Activité de reformulation** dans laquelle il est demandé aux informateurs de proposer une paraphrase de l'expression soulignée.	Compétence de production et de compréhension. L'informateur mobilise sa compétence pragmalinguistique et sociopragmatique pour interpréter le sens et la fonction de l'expression soulignée et proposer ensuite une reformulation adéquate.
	D2	I1	C'est pas possible ! (distracteur[125])			
	D3	I1	Faut pas pousser !			
		I2	C'est abusé !			
	D4	I1	C'est pas possible !			
		I2	J'y suis pour rien !			
Activité 3	D1	I1	J'y crois pas !		**Activité d'équivalence** dans laquelle il est demandé aux informateurs de proposer un équivalent fonctionnel en français à l'expression	Compétence de production. L'informateur mobilise dans un premier temps sa compétence métapragmatique et sociopragmatique pour interpréter le sens et
	D2	I1	C'est pas vrai ?			
		I2	C'est chaud !			
	D3	I1	C'est pas gagné !			

[124] Il est intéressant de noter que les trois premières activités, bien que mobilisées pour tester la production, impliquent également la compétence de compréhension en tâche de fond. En effet, le LNN va au préalable interpréter les éléments du contexte et/ou l'expression proposée avant de proposer une réponse. Cette étape d'inférence fait partie du processus de production.
[125] Un distracteur est un dispositif mis en place afin d'éviter d'éventuelles réponses aléatoires. Il s'agit de l'emploi d'un PhPex dans sa forme littérale.

		I2	Tu plaisantes ?	40 sec	soulignée. Pour les non-natifs, il est demandé de produire un équivalent fonctionnel dans la LM. À la différence de l'activité de reformulation, il faut que l'équivalent proposé ait les mêmes propriétés lexicopragmatiques, c'est-à-dire que l'expression soit une séquence préfabriquée ayant la même dimension illocutoire et discursive que l'expression de départ.	la fonction de l'expression soulignée. Il mobilise ensuite sa compétence pragmalinguistique pour produire un équivalent lexicopragmatique dans la langue cible ou dans la langue maternelle.
Activité 4	D1	I1	Choix entre : [c'est génial ! /c'est le pied ! /c'est le bouquet ! / c'est une horreur !]	30 sec	**Activité de complétion à choix multiple (QCM lexical)**. Il est demandé aux informateurs de choisir l'expression idoine en fonction du contexte en choisissant parmi 4 propositions de réponse possible.	Compétence de compréhension. L'informateur mobilise sa compétence sociopragmatique et métapragmatique pour choisir parmi une proposition de 4 expressions celle qui est la plus appropriée au contexte d'emploi proposé. Cette activité nécessite un effort cognitif supplémentaire puisqu'elle implique pour l'informateur de discriminer une expression parmi d'autres, c'est-à-dire d'identifier le sens et la fonction de chaque expression par rapport au contexte et de choisir l'expression la plus appropriée en éliminant celles qui présentent un degré d'adéquation plus faible.
	D2	I1	Choix entre : [c'est génial !/ c'est pas gagné ! / ça tombe bien ! / c'est une horreur !]			
	D3	I1	Choix entre : [Tu crains ! / Tu parles ! / T'inquiète ! / Tu l'as dit !]			
Activité 5	D1	I1	T'es sérieux ?	40 sec	**Activité de reconnaissance sémantique à choix multiple (QCM sémantique)**. Pour chaque expression soulignée dans le dialogue, il est demandé aux informateurs de choisir la proposition de sens la plus appropriée parmi plusieurs choix.	Compétence de compréhension. L'informateur mobilise sa compétence sociopragmatique et métapragmatique pour choisir parmi les propositions de sens et de fonction la proposition la plus appropriée. Tout comme l'activité précédente, cette activité implique un effort cognitif supplémentaire en raison de la nécessité de discriminer le sens le plus approprié.
		I2	C'est pas possible (distracteur)			
	D2	I1	J'y crois pas !			
		I2	Les bras m'en tombent !			
		I3	C'est une honte !			
		I4	C'est le pied. (distracteur)			

Tableau 8 : Structure détaillée et contenu du test lexicopragmatique

Le test a été pensé au départ pour être mené sur un support papier, dans une salle dédiée et sous la supervision d'un surveillant. Une vidéo de guidage a été éditée afin d'accompagner les informateurs tout au long du test en proposant pour chaque activité la consigne de manière orale puis un extrait audio contenant un dialogue. Cette vidéo comporte également un timer pour chaque réponse. Ce timer a été mis en place afin d'éviter les phénomènes de surinterprétation chez les informateurs. La vidéo a aussi été conçue afin de rendre plus dynamique le test et maintenir l'engagement et la concentration cognitive des répondants :

Figure 24 : Extrait de la vidéo de guidage du test

Le test papier (Annexe 8) commence par un rapide questionnaire sociolinguistique sur les informations personnelles de chaque informateur. Ces informations sont utilisées par la suite pour élaborer un fichier de métadonnée que nous utilisons pour le traitement statistique ou pour affiner l'analyse qualitative. Viennent ensuite les consignes de l'activité accompagnées d'un exemple concret permettant de contextualiser ce qui est attendu puis le dialogue sur lequel porte la question :

Consigne : Observez les dialogues suivants et reformulez les expressions soulignées. Par exemple :

Léa, Ivan et Cédric discutent à propos de la sœur de Léa.

Léa : *Ma sœur, pour le travail saisonnier, elle est allée chercher aux impôts*
Ivan : *Ah ouais ?*
Léa : *Et elle a été prise ! Elle a quand même une chance elle, un truc international !*
Cédric : *C'est pas faux !*
Léa : *donc elle a été prise aux impôts, donc elle était trop contente, mais c'était super dur hein, les impôts !*
Ivan : *Ah bon ?*
Léa : *ouais, et c'était une trop sale ambiance en plus*
Ivan : ***tu m'étonnes = [Tu as raison, c'est sûr !]***
Cédric : *ouais, y a moyen, j'imagine pas*

❖ **Dialogue 1**

Alice, Billie et Chris discutent à propos d'un salon de massage.

Alice : Tu te rappelles du chèque cadeau que mon copain m'a offert ?
Billie : Ah oui pour ton anniversaire, c'est ça ?
Alice : Et bah je l'ai utilisé hier dans le salon de massage qui est vers la mairie.
Billie : Ah super !
Chris : Alors, c'est comment ?
Alice : Ah bah **c'est le pied !** = _____ Ils te font la totale de la tête aux pieds avec les huiles essentielles !
Chris : Faudra que j'offre ça à ma copine alors !

Figure 25 : Exemple d'une question proposée dans le test papier proposé aux LN

Chaque dialogue comporte des didascalies qui précisent les différents éléments contextuels utiles à la compréhension puis le script du dialogue qui sera entendu dans le guide vidéo et enfin un espace dédié à la réponse. Au total, 22 questions ont été soumises aux informateurs pour une durée de 30 minutes de test. Ce test a volontairement été conçu de manière à être le plus court possible pour éviter la fatigue cognitive liée à ce type d'exercice ou un effet de *speeding*[126]. Cette première version a été distribuée aux locuteurs natifs.

En raison de la crise sanitaire qui a débuté en janvier 2020, nous avons dû revoir l'intégralité du test pour le transposer dans une modalité à distance avant de le soumettre au groupe des LNN. Nous avons ainsi transformé le support papier initial en un questionnaire en ligne dans le lequel nous avons intégré la vidéo de guidage élaboré auparavant :

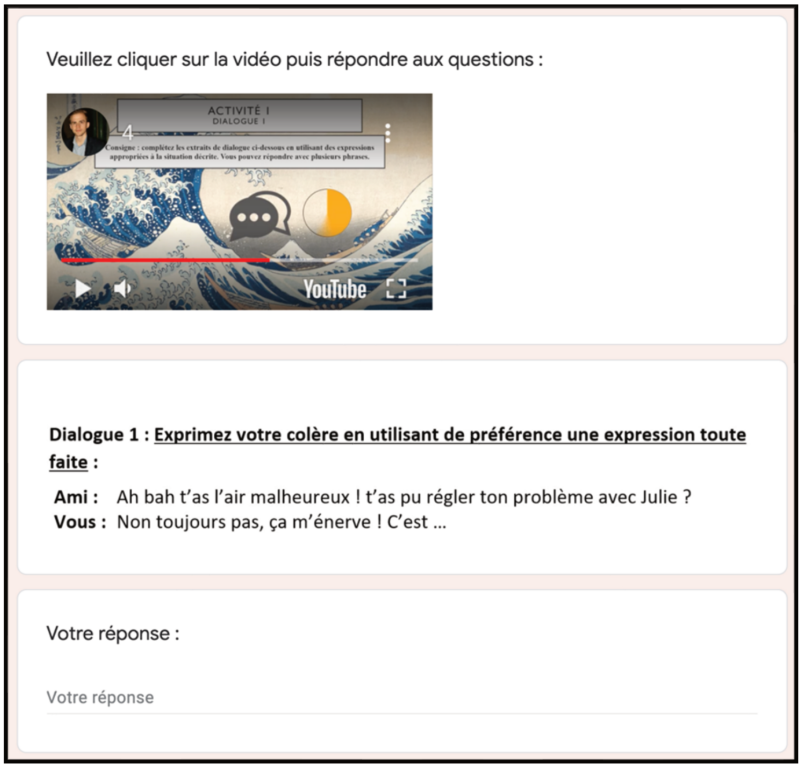


Figure 26 : Exemple d'une question proposée dans le test en ligne proposé aux LNN avec l'intégration du guide vidéo

[126] Effet de précipitation provoqué par l'apparition d'items identiques dans un questionnaire qui amène le répondant à répondre à une question de la même manière qu'avec les précédents items sans considérer la question en elle-même. On rencontre cet effet lors de l'utilisation de l'échelle de Likert ou de questions dont la structure est répétitive par exemple (Andreadis & Kartsounidou, 2020, p33)

Nous avons conservé la structure du test papier à ceci près que nous avons dû ajouter quelques informations pratiques concernant l'utilisation de la plateforme de réponse en ligne. Pour chaque question, nous avons ajouté une partie d'autoévaluation :

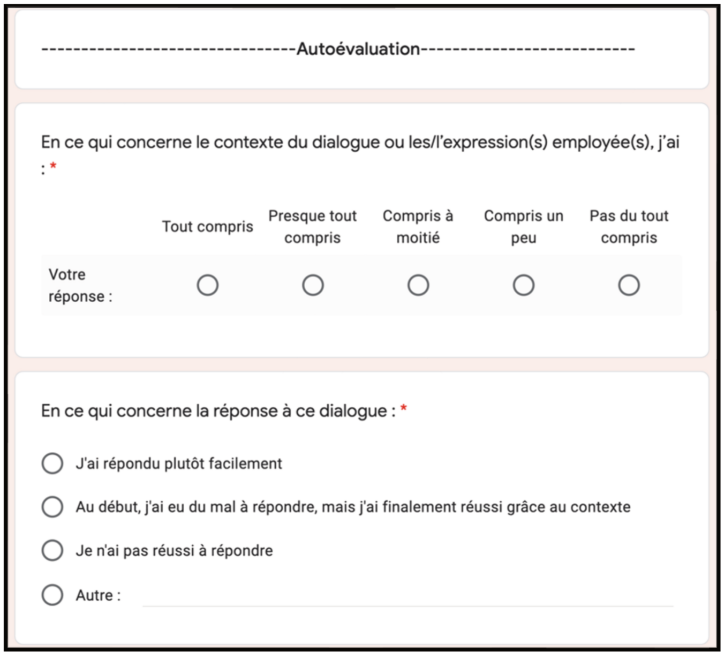


Figure 27 : Exemple d'une question d'autoévaluation

Cette autoévaluation se divise en deux questions : une première question interroge l'informateur sur sa compréhension du contexte tandis que la deuxième question interroge sur le niveau de compréhension de l'emploi et le niveau de difficulté ressenti pour répondre à la question. Le but de cette autoévaluation est d'avoir accès à des traces de la compréhension métapragmatique et réflexive des informateurs concernant les expressions traitées dans le test. Nous avons également choisi d'ajouter à la fin du test un *entretien d'élicitation métapragmatique semi-directif* visant à mener une discussion avec l'informateur autour des réponses qu'il a produites ou à propos de l'emploi d'une expression en particulier. Pour finir, un questionnaire à réponse libre[127] a été distribué à chaque informateur quelques semaines après l'expérimentation afin d'éviter toute influence du test sur les réponses à ce questionnaire. Cette enquête questionne la pratique du français et les problèmes rencontrés lors de la communication quotidienne en français et/ou avec un locuteur français. Ce questionnaire a été élaboré en japonais de manière à faciliter les

[127] Disponible à cette adresse : https://forms.gle/8MfCAANFy7kcS5eY8

réponses aux locuteurs et obtenir un maximum de détails. Nous reviendrons en précision sur ces deux derniers dispositifs dans la troisième phase de notre étude.

Le test à distance se déroule par webcam interposée de manière à s'assurer que les répondants n'ont pas recours à des ressources externes. Le passage du test et l'entretien qui s'ensuit sont enregistrés à l'aide du logiciel de capture vidéo d'écran *Camstasia*[128] pour avoir une trace consultable d'éventuelles interactions ou de questions qui nous seraient utiles pour notre analyse par la suite. Afin de garantir la diffusibilité de nos résultats et le cadre juridique de l'utilisation de nos données, nous avons demandé à chaque informateur de signer un consentement éclairé[129] (Annexe 3 et 4) nous autorisant à utiliser les données recueillies à des fins scientifiques.

3. Traitement des données

3.1. Formalisation des données

La formalisation des données a consisté en leur nettoyage, leur anonymisation et leur stockage en vue de l'analyse. La très grande quantité de données à la fois écrites (tests lexicopragmatiques) et orales (entretiens) nous a contraint à faire appel à un système d'archivage fonctionnant sur la base de l'utilisation d'identifiants. Ces identifiants présents tout au long du traitement et de l'analyse permettent de garder la traçabilité des données employées. Nous avons également constitué un index des métadonnées concernant les informateurs LN et LNN établi à l'aide des informations personnelles renseignées lors de la passation du test (Annexes 6 et 7). Enfin, nous avons réalisé un comptage par item et une synthèse des réponses à l'autoévaluation pour faciliter la comparaison de celle-ci avec les réponses effectives du test.

---

[128] https://www.techsmith.fr/camtasia.html

[129] Le consentement éclairé a été rédigé en français et en japonais afin d'une part de garantir que les locuteurs ont pleinement compris le contenu du document qui leur a été soumis, d'autre part, de garantir que le document soit utilisable en France ou au Japon.

3.2. Validation des résultats du test

La seconde étape de traitement des données constitue la validation des résultats recueillis. Chaque proposition de réponse des LN et des LNN a été soumise à un jugement d'acceptabilité qui détermine si la proposition est adéquate ou non, permettant ainsi de calculer un score de réussite au test. Afin d'éviter tout biais subjectif lors de ce jugement d'acceptabilité, nous avons évalué l'acceptabilité des réponses à l'aide d'une procédure d'accord interjuge. Du fait de la présence de réponses en français et en japonais dans l'activité 3 (équivalence), nous avons mis en place deux procédures d'accord interjuge avec des LN de chaque langue. Ces accords interjuge consistent à soumettre les résultats du test à plusieurs juges locuteurs natifs du français et/ou du japonais afin qu'ils évaluent la pertinence des réponses proposées pour chaque activité.

3.2.1. Procédure d'accord interjuge des réponses en français

Afin de nous assurer que tous les juges vont évaluer les résultats sur la base de mêmes critères, nous avons préparé un guide qui détaille précisément les critères qui sont à évaluer pour chaque activité et les points d'attention qu'il faut absolument observer lors du traitement des données.

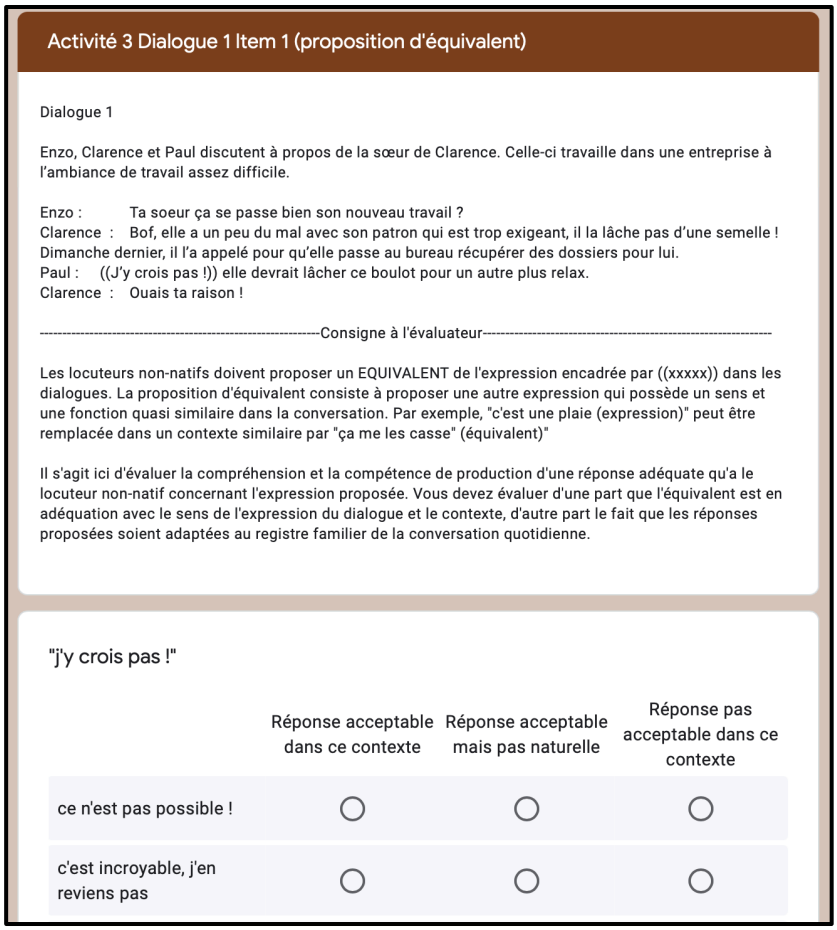
Nous leur avons demandé de rester le plus spontané possible dans le jugement d'acceptabilité afin d'éviter tout biais cognitif de sur-interprétation. Nous avons également rappelé que l'orthographe n'était pas ici évaluée, uniquement l'acceptabilité des réponses dans le dialogue donné, en supposant qu'il s'agit d'une interaction orale habituelle. Nous avons également donné la possibilité aux évaluateurs de consulter le PDF du test pour avoir accès à plus de précisions concernant la structure du test, la consigne demandée ou les informations extralinguistiques mises à la disposition des informateurs.

Le groupe d'évaluateurs ayant participé à cette procédure pour le français se compose de 6 locuteurs natifs[130] (3 hommes, 3 femmes), dont nous-même. Parmi ces 6 évaluateurs, 2 sont des chercheurs en sciences du langage spécialistes de la phraséologie, 2 sont chercheurs en sciences du langage respectivement spécialistes de sociolinguistique et de linguistique

[130] Nous remercions ici ces évaluateurs qui ont joué un rôle crucial dans l'évaluation de nos résultats et l'analyse qui en a été faite.

descriptive ; et les deux derniers locuteurs sont des non-linguistes. La décision de ne pas choisir que des évaluateurs linguistes experts de la phraséologie est essentiellement motivée par le souci de représentativité. La compétence du linguiste spécialiste est certes un atout pour évaluer les réponses selon des critères précis et consensuels, mais peut être aussi à l'origine de la mise à l'écart de certaines réponses qui bien qu'inhabituelles, rendent compte d'emplois existants. En outre, il nous a semblé que d'impliquer que des évaluateurs ayant une forte culture académique, et de fait, l'habitude de communiquer dans une langue propre à ce domaine, pourrait dans une certaine mesure constituer un biais de sur-correction lors de l'évaluation des réponses.

Les formulaires d'évaluation comportent dans un premier temps le dialogue associé à l'item traité dans l'activité. À la suite de ce dialogue se trouvent les consignes de l'évaluateur. Ces consignes spécifient de manière explicite ce qui est attendu. Nous avons également précisé la compétence visée par l'activité et les points d'attention à prendre en compte lors de l'évaluation de l'acceptabilité. Voici un exemple de ce formulaire :



Activité 3 Dialogue 1 Item 1 (proposition d'équivalent)

Dialogue 1

Enzo, Clarence et Paul discutent à propos de la sœur de Clarence. Celle-ci travaille dans une entreprise à l'ambiance de travail assez difficile.

Enzo : Ta soeur ça se passe bien son nouveau travail ?
Clarence : Bof, elle a un peu du mal avec son patron qui est trop exigeant, il la lâche pas d'une semelle ! Dimanche dernier, il l'a appelé pour qu'elle passe au bureau récupérer des dossiers pour lui.
Paul : ((J'y crois pas !)) elle devrait lâcher ce boulot pour un autre plus relax.
Clarence : Ouais ta raison !

----------Consigne à l'évaluateur----------

Les locuteurs non-natifs doivent proposer un EQUIVALENT de l'expression encadrée par ((xxxxx)) dans les dialogues. La proposition d'équivalent consiste à proposer une autre expression qui possède un sens et une fonction quasi similaire dans la conversation. Par exemple, "c'est une plaie (expression)" peut être remplacée dans un contexte similaire par "ça me les casse" (équivalent)"

Il s'agit ici d'évaluer la compréhension et la compétence de production d'une réponse adéquate qu'a le locuteur non-natif concernant l'expression proposée. Vous devez évaluer d'une part que l'équivalent est en adéquation avec le sens de l'expression du dialogue et le contexte, d'autre part le fait que les réponses proposées soient adaptées au registre familier de la conversation quotidienne.

"j'y crois pas !"

	Réponse acceptable dans ce contexte	Réponse acceptable mais pas naturelle	Réponse pas acceptable dans ce contexte
ce n'est pas possible !	○	○	○
c'est incroyable, j'en reviens pas	○	○	○

Figure 28 : Exemple du formulaire d'accord interjuge pour les réponses en français

L'évaluateur peut également avoir accès à l'enregistrement audio associé au dialogue en se munissant du PDF fourni au préalable. En dessous des consignes, l'expression cible de l'évaluation est une nouvelle fois mentionnée, puis la liste des réponses enregistrées dans le test est soumise aux évaluateurs. Pour chacune de ces réponses, trois choix d'évaluation sont possibles. Les deux extrêmes constituent l'acceptabilité ou la non-acceptabilité de la réponse. Le choix médian est un ajout volontaire permettant d'évaluer les réponses qui sont acceptables, mais ne constituent pas les réponses préférentiellement associées au contexte de communication proposé. Nous avions remarqué lors de l'observation préliminaire des données que beaucoup de réponses des LNN étaient acceptables, mais nous laissaient une impression d'étrangeté[131]. L'hypothèse que nous avions faite à ce moment-là est que bien qu'avancée, la compétence lexicopragmatique des répondants ne permettait pas dans certains cas d'exprimer la nuance d'emploi la plus appropriée. C'est notamment pour vérifier cette hypothèse auprès des évaluateurs que nous avons ajouté cette valeur médiane. Une fois les évaluations des 6 juges récoltées, nous avons eu accès une synthèse quantitative directement dans l'interface du formulaire :

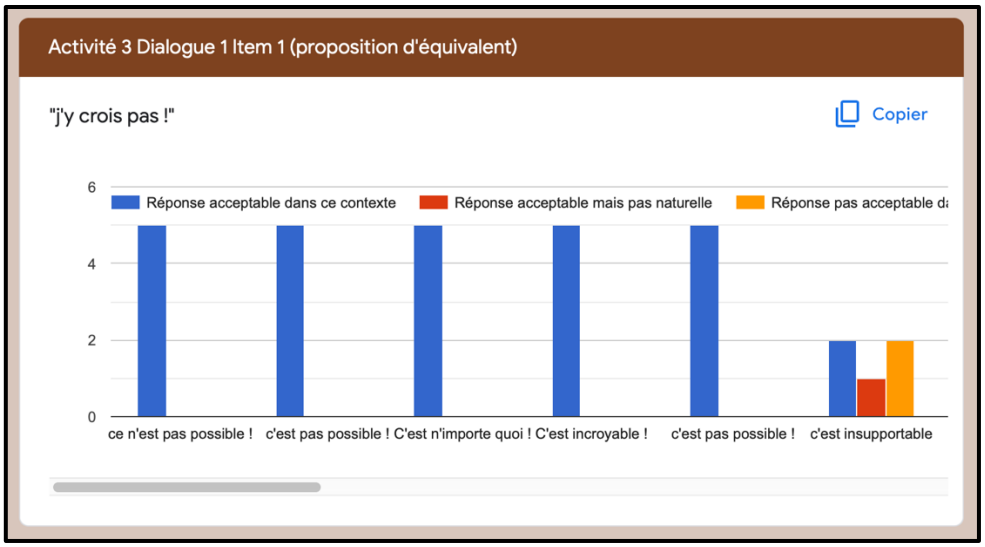


Figure 29 : Exemple de synthèse des accords interjuge pour les réponses en français

La méthode de calcul de l'acceptabilité est la suivante : l'évaluation prise en compte est celle totalisant 4 ou plus choix des évaluateurs. En cas d'égalité entre les deux extrêmes (3 choix acceptables *vs* 3 choix non acceptables), nous choisissons d'enregistrer une évaluation non

[131] Nous verrons dans l'analyse qualitative qu'il peut s'agir par exemple de la modulation de l'intensité expressive véhiculée par le choix lexical opéré lors de la production.

acceptable de la réponse. Lorsqu'il y a égalité entre les trois valeurs de l'évaluation, nous choisissons d'enregistrer une valeur acceptable. L'item en question est ensuite marqué sur le fichier de synthèse des réponses au test afin d'être observé plus en détail lors de l'analyse qualitative. Les résultats de cette évaluation interjuge ont été ensuite reportés sur les réponses des informateurs. Nous avons ensuite effectué un comptage des réponses acceptables par informateur que nous avons transformé en pourcentage pondéré en fonction du total d'items présents dans le test pour obtenir un taux de réussite. Globalement, nous avons observé peu de discordances entre les évaluations des juges sur l'ensemble du test. Cependant, nous avons pu remarquer que l'activité 1 (complétion) a suscité le plus de désaccord entre les juges étant donné que ce type d'activité permet le recours à une plus grande variété de réponses possibles et tout autant d'interprétations.

3.2.2. Procédure d'accord interjuge des réponses en japonais

La même procédure que précédemment a été appliquée aux équivalents fonctionnels en japonais proposés dans l'activité 3 (équivalence). Toutefois, compte tenu de la dimension contrastive des données de l'activité 3, nous avons modifié quelques points. La spécificité de cette procédure d'évaluation interjuge est qu'il faut à la fois évaluer et comprendre le dialogue et le PhPex proposés en français, puis en fonction de cette première interprétation, évaluer l'acceptabilité des équivalents donnés en japonais. Ce travail nécessite donc un temps de réflexion plus important que pour l'évaluation interjuge classique.

Étant donné la grande difficulté de la tâche, nous avons demandé à 4 enseignants-chercheurs japonophones natifs (4 femmes) qui sont des linguistes spécialistes du japonais et extrêmement compétentes en français de réaliser cet accord interjuge[132]. Les spécialités des évaluatrices sont la linguistique et la didactique du japonais, le FLE et la linguistique contrastive français-japonais. Les universités d'appartenances des évaluatrices sont en France, en Belgique et au Japon. La sélection des juges est dans ce cas soumise à de fortes contraintes. Il nous fallait trouver des locuteurs experts dans les deux langues, natifs du japonais et étant compétents en sciences du langage pour l'aspect contrastif du jugement. De ce fait, nous n'avons pas pu varier les profils des juges comme nous l'avons fait dans la

[132] Nous remercions les quatre évaluatrices qui ont donné de leur temps pour cette évaluation très exigeante alors même qu'elles étaient déjà très sollicitées par leurs obligations universitaires respectives.

procédure initiale. En outre, n'étant pas un locuteur natif du japonais, nous avons choisi de ne pas tenir compte de notre évaluation des réponses réalisée au préalable afin de garantir l'impartialité des évaluations.

Afin d'éviter au maximum les biais dans l'évaluation des résultats, nous avons à nouveau élaboré un guide à l'attention des 4 juges détaillant précisément les points à évaluer concernant la consigne de l'activité :

----------Consigne à l'évaluateur----------

Dans cette activité, les locuteurs non-natifs doivent proposer un ÉQUIVALENT en français et en japonais de l'/les expression(s) soulignée(s) et en gras dans les dialogues proposés (cf. aperçu des consignes ci-dessous).

La proposition d'équivalent consiste à proposer une expression qui possède un sens et une fonction quasi similaire à une expression employée dans la conversation, mais qui peut avoir en revanche une forme tout à fait différente par rapport à l'expression en langue source. Par exemple, "c'est une plaie" (expression idiomatique de départ) peut être remplacée dans un contexte similaire par "ça me les casse" (équivalent) en français, ou par 「面倒くさいなー」(équivalent) en japonais.

Il s'agit ici d'observer si les informateurs ont bien compris le contexte ainsi que le sens de l'expression proposée. Vous devez évaluer les points suivant :

a) l'équivalent proposé a le statut d'expression idiomatique
b) le sens de l'équivalent proposé est en adéquation avec le sens de l'expression du dialogue
c) l'équivalent proposé s'emploie dans un contexte similaire au dialogue
d) l'équivalent proposé correspond à un usage au registre familier de la conversation quotidienne
e) l'équivalent proposé correspond à la portée expressive et discursive de l'expression désignée dans le dialogue

Nous vous suggérons lors de votre évaluation de prendre en compte le fait que les réponses données étant écrites, les informateurs ne pouvaient pas rendre compte complétement du trait intonatif de l'expression. Pour la plupart, ils ont utilisé une ponctuation expressive (... /? / !). N'hésitez pas à essayer d'imaginer plusieurs intonations pour aider votre jugement.

Figure 30 : Consignes de l'accord interjuge pour les réponses en japonais

Nous avons également précisé le déroulement de cette procédure d'évaluation pour nous assurer que les 4 évaluatrices réaliseraient la tâche de manière homogène et comprendrons bien ce qui est attendu. Cette dernière précaution est selon nous très importante étant donné la grande complexité de la tâche :

----------Déroulement de l'évaluation----------

Dans ce qui suit, vous aurez la possibilité d'écouter le dialogue tel qu'il a été proposé aux informateurs le jour du passage du test. Chaque vidéo contient à la fin un timer qui était destiné aux informateurs pour cadrer le temps de réponse, je vous prie de bien vouloir mettre la vidéo en pause lorsque le dialogue est terminé pour éviter que ce timer vous gêne. Vous disposez également du script du dialogue pour accompagner l'écoute de la bande sonore.

Comme la qualité audio de Google Form est plutôt faible, je vous conseille de vous équiper d'un casque audio et de bien monter le volume. Vous pouvez relancer la vidéo à tout moment pour réécouter le dialogue en cas de doute.

Pour chaque expression désignée dans les dialogues, il vous sera proposé une définition de l'expression pour vous aider à situer les différents paramètres de son usage et ajuster votre évaluation si besoin. Bien évidemment, je vous invite à suivre votre intuition en premier lieu, puis à vous référer à cette définition en cas de doute.

L'évaluation consiste en une liste dans laquelle vous pourrez examiner toutes les formes retrouvées dans les réponses au test que les locuteurs japonophones apprenants du français ont fourni. Vous devrez cocher parmi trois valeurs celle qui correspond le plus à l'adéquation de l'équivalent fonctionnel proposé avec l'expression d'origine du dialogue :

–> Valeur 1 "Équivalent adéquat dans ce contexte" = l'équivalent proposé est parfaitement adapté et répond à tous les critères d'équivalence en termes de sens, de registre, d'idiomaticité et d'expressivité.

–> Valeur 2 "Équivalent acceptable, mais pas naturel" = l'équivalent proposé est acceptable, mais ne semble pas tout à fait naturel ou peu adapté au contexte. Cela peu concerner par exemple une expression dont le sens semble acceptable, mais qui ne réunit pas les autres critères établis plus haut pour une équivalence parfaite.

–> Valeur 3 "Réponse inadéquate dans ce contexte" = l'équivalent proposé n'est pas adapté ou constitue un contresens.

Figure 31 : Informations additionnelles proposées dans l'accord interjuge pour les réponses en japonais

Les évaluatrices avaient, tout comme la procédure initiale, accès au PDF du test distribué et aux enregistrements audios des dialogues. Dans le formulaire d'évaluation, nous avons intégré au dialogue une définition du PhPex dont il est question afin de permettre d'avoir une base de référence commune dans leur évaluation :

L'expression "J'y crois pas !"

J'Y CROIS PAS ! est utilisé lorsque le locuteur veut exprimer son étonnement ou son indignation, en réaction à une action, une situation ou un comportement qui vient de se produire. Sa dimension expressive permet d'exprimer son indignation et plus rarement son contentement. Sur le plan discursif, l'expression permet de désapprouver ou manifester son incompréhension. Le locuteur fait comme s'il ne pouvait pas croire ce qu'il vient d'entendre, ce qui est une manière pour lui de dire que ce à quoi il réagit est inconcevable.

Script du dialogue

Dialogue 1

Enzo, Clarence et Paul discutent à propos de la sœur de Clarence. Celle-ci travaille dans une entreprise à l'ambiance de travail assez difficile.

Enzo : Ta sœur ça se passe bien son nouveau travail ?
Clarence : Bof, elle a un peu du mal avec son patron qui est trop exigeant, il la lâche pas d'une semelle ! Dimanche dernier, il l'a appelé pour qu'elle passe au bureau récupérer des dossiers pour lui.
Paul : J'y crois pas ! elle devrait lâcher ce boulot pour un autre plus relax.
Clarence : Ouais ta raison !

Figure 32 : Définition du PhPex et dialogue ciblé dans l'évaluation

Puis nous avons proposé les occurrences à évaluer de la même manière que dans la procédure de départ, avec les mêmes critères d'évaluation :

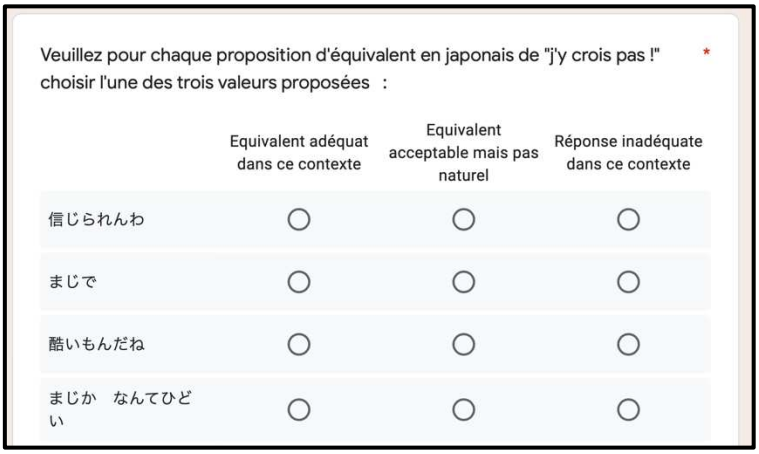


Figure 33 : Exemple du formulaire proposé dans l'accord interjuge pour les réponses en japonais

Une fois les évaluations des 4 juges récoltées, nous avons consulté la synthèse quantitative directement intégrée à l'outil de formulaire. Nous avons laissé la possibilité aux évaluatrices de laisser un commentaire si besoin :

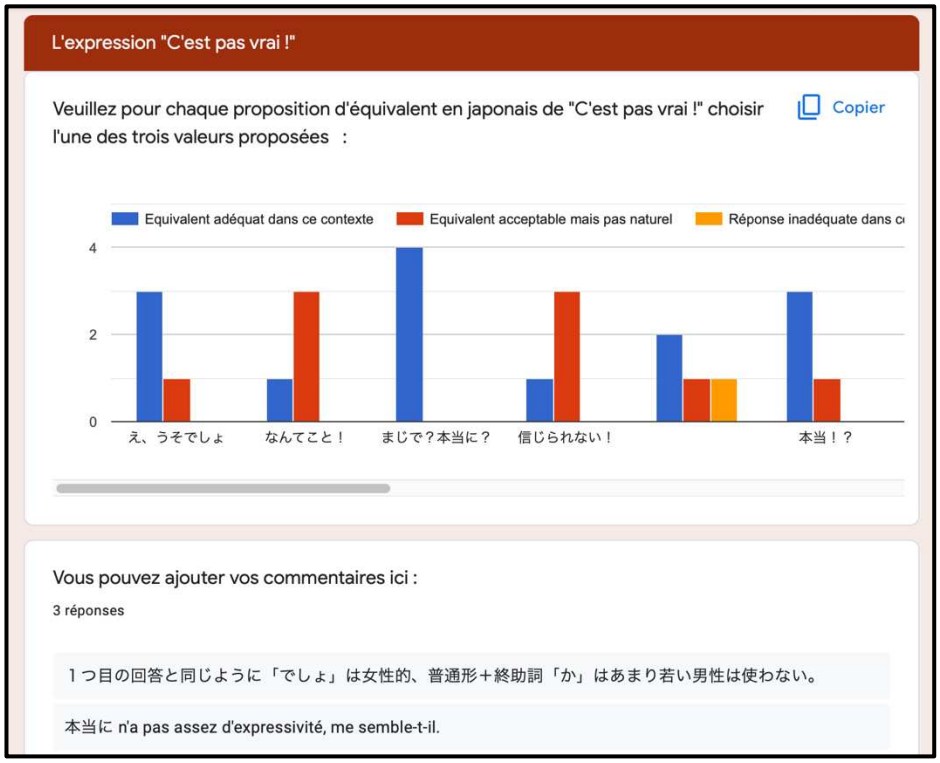


Figure 34 : Exemple de synthèse des accords interjuge pour les réponses en japonais

La méthode de calcul de l'acceptabilité est la même que pour la procédure sur le français : l'évaluation prise en compte est celle totalisant 3 ou plus choix des évaluateurs. En cas de dispersion des résultats entre les deux extrêmes (2 choix acceptables *vs* 2 choix non acceptables), nous enregistrons une évaluation non acceptable de la réponse. Comme la procédure d'accord faite sur le français, nous avons observé peu de discordances entre les évaluations des juges. Cependant, nous avons pu remarquer que les expressions *c'est pas gagné !* et *tu plaisantes ?* de l'activité 3 (équivalence) ont suscité le plus de désaccord entre les juges. Nous verrons dans l'analyse que les résultats enregistrés pour ces deux expressions ont également posé problème aux LNN, notamment pour la tâche d'équivalence en japonais, car certaines des expressions proposées dans le test ne semblent pas avoir d'équivalents directs en japonais.

3.2.3. Pourquoi évaluer les résultats des LN ?

Le lecteur de ce travail de thèse ne manquera pas de constater que nous avons aussi évalué les réponses des locuteurs natifs. Il s'agit d'un choix volontaire de notre part de nous reposer exclusivement sur la procédure d'accord interjuge pour valider les résultats de notre test. En effet, l'évaluation de la production du LNN pose 3 questions :

- La première question est celle de la légitimité du statut[133] que l'on va donner à la production du LN recueillie dans les données de contrôle.

- La seconde question est de savoir ce qui constitue une interaction réussie. Sur quelles traces linguistiques fonder ce jugement ? (Larsen-Freeman, 2014)

- La troisième question est celle de la norme. Est-elle la même pour tous est dans tous les cas ? Est-ce que ne pas s'exprimer dans le standard la langue cible constitue un échec pragmatique ?

---

[133] Cette question a été soulevée par Agnès Tutin (LIDILEM) lors de la relecture de ce manuscrit, nous l'en remercions.

Dans le cas de notre travail de recherche qui porte sur la conversation quotidienne, il nous a semblé que cette approche comparative centrée sur l'idée d'une compétence native idéale peut paraitre prescriptive et comporte certains aspects critiquables.

La première critique concerne le fait que la comparaison des données de LNN avec celles des LN tend à pérenniser une situation où une acquisition réussite constitue nécessairement une production quasi identique à celle des natifs, rendant ainsi la production du LNN nécessairement déficiente, et ce peu importe le niveau d'expertise (cf. *Illusion Comparative*[134], Bley-Vroman 1983 ; Selinker 2014, notre traduction). La seconde critique concerne le fait que cette approche présume que le LNN aspire nécessairement à atteindre une production similaire au LN. Toutefois, les objectifs des LNN peuvent grandement varier d'un individu à l'autre. Alors qu'un LNN se fixera comme objectif ultime la capacité d'interagir avec aisance totale dans un milieu académique, un autre pourra très bien aspirer à pouvoir être en mesure de pouvoir communiquer sans encombre dans la vie de tous les jours en langue cible. De même le standard attendu lors d'un entretien officiel, ne sera pas le même que celui attendu entre amis ou lorsque la langue cible constitue une *lingua franca* purement utilitaire.

Dans ce travail de recherche, la production du LNN n'est pas perçue comme une version déviante de la production du LN, mais comme une variété de la langue cible qui représente l'état de développement de la compétence lexicopragmatique du LNN et constitue notre objet d'analyse. Cette approche implique donc une certaine prise de recul quant aux données comparatives et à l'interprétation qu'elles donnent à voir. La prise en compte d'objectifs d'acquisition des LNN semble ici nécessaire. En outre, l'étalon d'évaluation ne porte sur la capacité du locuteur à égaler la production du LNN, mais sur sa capacité réaliser de manière effective un acte de langage, tout en étant compris de son interlocuteur et en respectant les règles de formation du système linguistique de la langue cible[135]. Le corpus contrôle sera principalement mobilisé pour rendre compte des spécificités habituellement observées lors de la communication orale quotidienne entre LN. La comparaison n'aura pas de caractère

---

[134] *Comparative fallacy* (Bley-Vroman 1983)

[135] Nous précisons ici que le respect « des règles de formation du système de la langue cible » ne fait pas ici référence au standard de la langue, mais aux règles qui président à la construction d'un énoncé qui sera jugé intelligible et acceptable, que ce soit à l'oral ou à l'écrit. Ainsi, un énoncé comportant une élision ne sera pas jugé comme standard, mais sera jugé bien construit du point de vue de la norme de l'oral.

prescriptif, mais permettra de mettre en relief les différences entre la variété native et non native. Ces différences constitueront des indices qui nous permettront de caractériser la variété non native.

Un troisième argument à la non-pertinence de l'utilisation du corpus contrôle pour l'évaluation de l'acceptabilité est que nous nous sommes aperçu que les emplois inadéquats pouvaient apparaitre chez les LNN[136]. Ces erreurs d'interprétation sont plus rares, mais méritent d'être mentionnées. Partant des principes que les PhPex sont des objets linguistiques complexes à définir même pour un linguiste et que l'interprétation des PhPex peut être très variable en fonction des individus, nous avons émis l'hypothèse que cette complexité pouvait être source d'emplois inadéquats chez le locuteur natif également, d'autant plus en condition de test. Nous verrons par ailleurs plus loin dans la section dédiée à l'analyse qualitative que cette évaluation des réponses des LN nous a également permis de repérer des biais externes liés à une difficulté présente dans la conception du test en elle-même que nous n'avions pas décelée lors de son élaboration.

Ce qui semble finalement central dans ce travail de recherche, c'est de comprendre ce que le LNN mobilise sur le plan linguistique pour répondre à un besoin communicatif dans la conversation quotidienne, les indices sur lesquels il se fonde pour comprendre ou choisir la forme adéquate et les difficultés rencontrées. Il ne s'agira donc pas d'établir si les LNN observés sont de « bons » locuteurs ou des locuteurs compétents, mais d'observer objectivement quelles sont les traces linguistiques visibles du développement et des caractéristiques de leur compétence lexicopragmatique. C'est pour cette raison que nous avons jugé utile de fonder le taux de réussite au test sur la procédure d'accord interjuge. Cette procédure a notamment le mérite d'écarter au maximum les jugements subjectifs ou les représentations du standard[137]. La validation de l'acceptabilité se fonde sur la capacité de la réponse à répondre au besoin communicatif rencontré, et non pas sur sa capacité à égaler un standard idéal. Cette validation repose elle-même sur des critères et une définition

---

[136] D'autres études telles que Taguchi (2011 : 923) montrent que les LN n'obtiennent pas forcément le score maximal à un tâche linguistique sollicitant la compétence pragmatique. Cela ne parait pas si surprenant si l'on prend en compte le fait que les échecs pragmatiques et communicatifs peuvent très bien se produire entre deux LN et ne sont pas l'apanage des LNN.

[137] Cette représentation du standard est propre au chercheur et peut influencer le jugement d'acceptabilité. Par exemple un chercheur dont la sensibilité à la langue standard est élevée aura tendance à juger un énoncé inadéquat lorsque celui-ci s'écarte trop de la norme, et ce même si cet énoncé est utilisable en contexte.

commune partagée entre les différents évaluateurs de l'item à évaluer. Ce positionnement à l'égard des résultats des LNN et des LN sera celui adopté tout au long de l'analyse.

3.3. Mesure des spécificités lexicales des réponses des LN et des LNN

Afin de pouvoir mesurer des indicateurs concrets qui nous permettront de caractériser de manière empirique les spécificités de la compétence lexicopragmatiques des LNN, nous avons élaboré une série de mesures quantitatives. Ces mesures ont été élaborées afin d'explorer les hypothèses suivantes :

- Les LNN semblent moins enclins à diversifier le lexique pragmatique employé lors de la communication quotidienne en langue cible => mesure de la *variabilité lexicale.*

- Les LNN rencontrent des difficultés à adapter leur production au registre attendu dans la conversation et semblent employés plus souvent un registre standard => mesure de la *variabilité du registre*

- Les LNN semblent moins enclins à utiliser un lexique phraséologique lors de la communication quotidienne en langue cible => mesure de l'*indice phraséologique*.

- Les LNN ont bien souvent une compréhension avancée de la forme pragmatique attendue, mais ne possèdent pas une palette lexicopragmatique assez large pour l'exprimer en langue cible => mesure du *taux de compréhension* comparé au score de réussite.

Ces mesures sont calculées à la fois dans les résultats du groupe contrôle et du groupe test. Elles nous permettent de faire ressortir des tendances que nous confronterons à un examen concret des données. Comme évoqué plus haut, la comparaison des mesures de spécificités observées dans les deux cohortes nous permettra de caractériser à la fois les spécificités et les difficultés propres la variété non native observée.

3.3.1. Variabilité lexicale

La *variabilité lexicale* mesure le nombre de séquences lexicales différentes proposées par les informateurs du test rapporté au nombre total de réponses enregistrées pour chaque

cohorte, en ne comptant pas les variantes d'une même expression (ex. : *c'est pas possible !* = *ce n'est pas possible !*) ou les incises (ex. : *c'est dingue ! = c'est vraiment dingue !*). La mesure est élaborée en calculant le pourcentage de formes différentes observées sur le total de réponses enregistrées pour chaque cohorte, et ce à l'aide d'un produit en croix[138].

La mesure de la variabilité du lexique part du principe que l'emploi des PhPex n'est pas contraint comme le serait celui des pragmatèmes. Pour un même schéma conversationnel, on peut observer tout un paradigme de PhPex mise à la disposition du locuteur par la langue. Par exemple, une situation ou un propos suscitant l'indignation peut déclencher l'utilisation de PhPex tels *ça va pas la tête !*, *c'est pas possible !, c'est scandaleux !, mais où va-t-on !, j'y crois pas !*. Le choix de l'expression qu'opère le locuteur au sein de ce paradigme dépend de plusieurs facteurs parmi lesquels le registre, la relation avec l'interlocuteur, l'intensité expressive ou l'intention communicative. Une mesure de variabilité faible (Variabilité < 20% soit 7 formes différentes sur 35 réponses) peut indiquer une forte spécialisation de certaines séquences ou une contrainte particulière propre au contexte. À l'inverse, une trop grande variabilité du lexique (Variabilité > 80% soit 28 formes différentes sur 35 réponses) peut indiquer chez les LNN une difficulté de compréhension. Dans le cas de l'activité 3 (équivalence), cette forte variabilité peut également indiquer l'absence d'équivalent fonctionnel dans la langue maternelle.

3.3.2. Variabilité du registre

La notion de registre renvoie à la nomenclature selon laquelle les productions linguistiques sont évaluées et catégorisées par les locuteurs d'une langue (Ure 1982 ; Biber 1991 ; Gadet 1996 ; Biber et Conrad 2019). Le registre se caractérise par des traits spécifiques tels que la complexité ou la fréquence du lexique employé ; la combinatoire, la syntaxe et la longueur des énoncés ; l'intonation et l'attitude associées, ainsi que le type d'illocutoire ou de stratégie communicative visés. Il est souvent organisé en niveau (par ex. soutenu, courant, familier, populaire, vulgaire, etc.). Ce partitionnement diffère selon les disciplines et les dictionnaires (pour une discussion sur ce point, cf. Wachs 2005). La notion de registre ou de niveau de langue est très discutée (Biber et Finegan 1994 ; Gadet 1996). Nous choisissons ici le terme

[138] Pour cette mesure et celle qui seront présentées ensuite, nous avons systématiquement écarté les absences de réponse et les réponses non-pertinentes (réponses beaucoup trop vulgaires ou surjouées).

de *registre* plutôt que de *niveau* de langue afin d'éviter de hiérarchiser un registre par rapport à un autre ou de placer le standard comme norme absolue. Au contraire, les différents registres employés dans notre travail de thèse se trouvent sur un continuum et ne sont pas cloisonnés. Tout comme Biber et Conrad (2019), nous considérons le registre comme une variété d'une langue donnée associée à des situations et des objectifs particuliers de communication. L'identification de ces registres repose sur des éléments linguistiques qui s'actualisent en fonction de cette situation ou de ces objectifs de communication. Notre travail de recherche portant essentiellement sur le niveau lexical, nous prendrons en compte des critères liés à la construction du lexique, à la morphosyntaxe et à la grammaire. Nous adoptons ainsi une partition en quatre catégories : *formel, courant, familier* et *très familier*.

- **Registre formel** = Structure syntaxique généralement complexe du propos. Exhaustivité de l'information. Politesse linguistique prédominante, distance énonciative forte. Emploi de lexique complexe voir spécialisé. Emplois phraséologiques ou idiomatiques orientés vers des tournures rituelles, de politesse ou spécialisées. Hypercorrection forte et standardisée. (Ex. : *quelle déception !*)

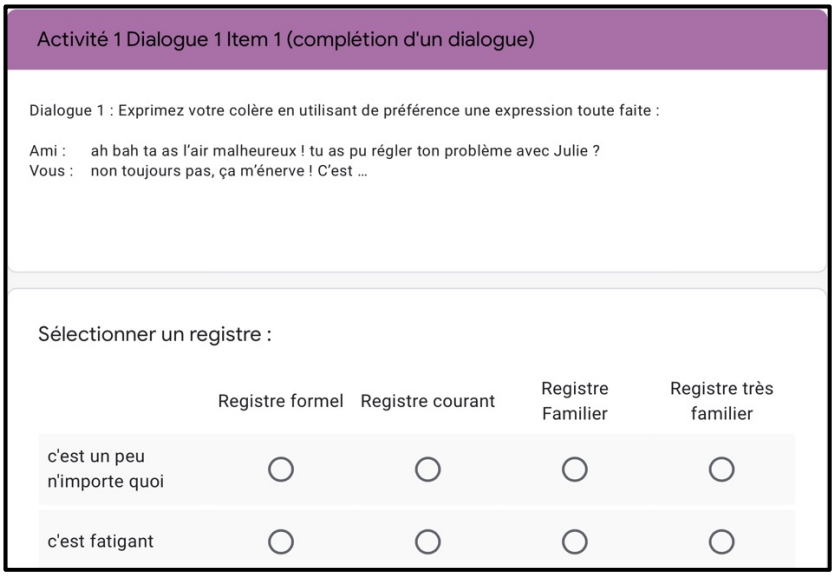
- **Courant** = le registre de base ayant le minimum de variation. Structure syntaxique standard du propos. L'interaction n'appelle ni un effacement ni un renforcement de la distance énonciative. Politesse minimale. Peu voire pas d'emploi du lexique spécialisé. Le lexique est plus générique. Hypercorrection moins forte. (Ex. : *c'est dommage !*)

- **Familier** = Relâchement des contraintes linguistiques autorisant des écarts à la norme (grammaire, syntaxe, ellipse fréquente, amalgames morphologiques, etc.). Structuration synthétique du propos. Énoncés peu complexes sur le plan syntaxique. Plus grande variabilité des formes. Effacement de la distance énonciative et spontanéité. Hypercorrection faible. (Ex. : *je suis dégouté !*)

- **Très familier** = Similaire au familier, mais comporte fréquemment l'emploi du lexique vulgaire. Utilisable uniquement avec un effacement complet de la distance énonciative. (Ex. : *j'ai trop la mort putain !*).

Les 4 paliers envisagés se situent dans la lignée de la catégorisation proposée par Wachs (2005) et de celle proposée par Mekki *et al.* (2021). Bien évidemment, le but de la catégorisation que nous proposons ici n'est pas de créer de nouveaux cloisonnements, mais de tenter d'établir des critères observables et comparables pour l'analyse de nos résultats.

La *variabilité du registre* mesure la répartition des réponses enregistrées entre ces quatre registres. L'indice de variabilité du registre est élaboré en calculant le pourcentage de résultats classés dans chaque catégorie de registre rapporté au total de réponses enregistrées pour chaque cohorte, et ce à l'aide d'un produit en croix. Le registre formel n'étant présent dans aucun des résultats observés dans le test lexicopragmatique, nous avons réduit les données affichées dans les tableaux de synthèses qui seront proposés dans l'analyse qualitative aux registres courant, familier et très familier. Le classement des réponses dans chaque catégorie de registre a été effectué selon une procédure similaire à l'accord interjuge. Cette procédure de classement a été mise en œuvre pour les réponses en français et en japonais. Les résultats en français ont été catégorisés par 3 locuteurs natifs du français : 2 locuteurs doctorants en linguistique (linguistique descriptive) et un locuteur enseignant de français en collège. Nous avons sélectionné les locuteurs pour leur expertise en linguistique et en langue, ainsi que pour leur capacité à suivre rigoureusement la procédure de classification soumise. Pour les résultats en japonais, les résultats ont été classés par 3 locuteurs japonais : 2 locuteurs doctorants en linguistique (linguistique contrastive et sociolinguistique) et un locuteur enseignant du japonais (japonais langue étrangère). Les locuteurs ont également été sélectionnés en fonction de leur expertise en linguistique et en langue japonaise. La procédure a été réalisée à l'aide d'un formulaire en ligne très similaire à celui de l'accord interjuge :



Activité 1 Dialogue 1 Item 1 (complétion d'un dialogue)

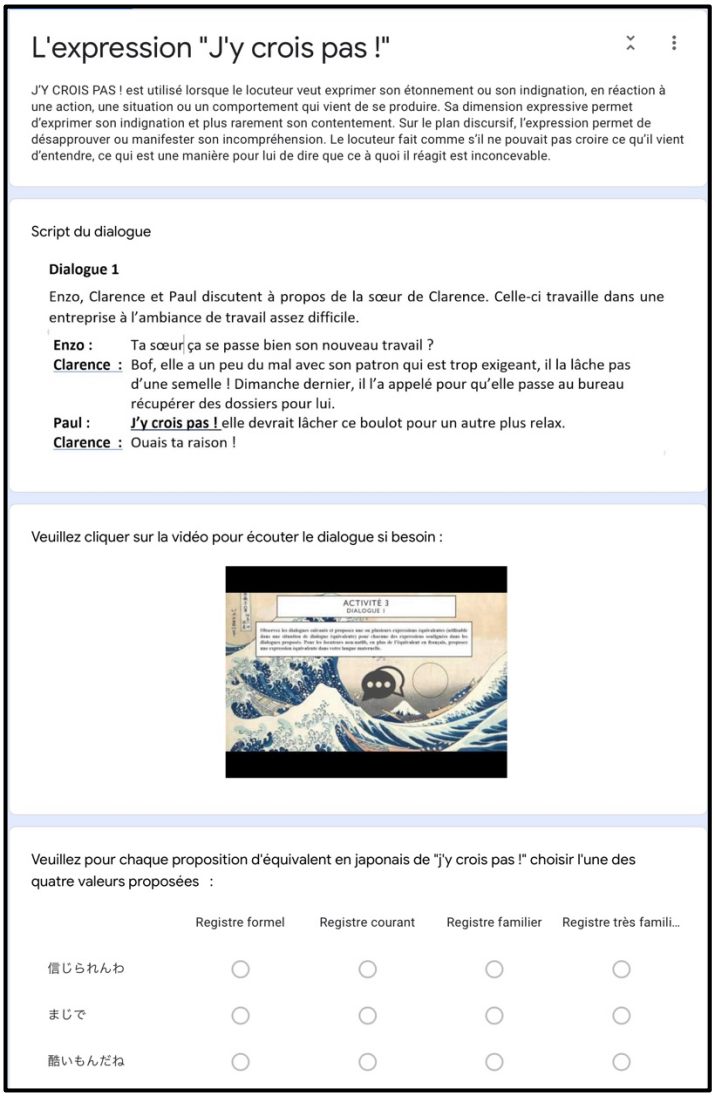
Dialogue 1 : Exprimez votre colère en utilisant de préférence une expression toute faite :

Ami : ah bah ta as l'air malheureux ! tu as pu régler ton problème avec Julie ?
Vous : non toujours pas, ça m'énerve ! C'est …

Sélectionner un registre :

	Registre formel	Registre courant	Registre Familier	Registre très familier
c'est un peu n'importe quoi	○	○	○	○
c'est fatigant	○	○	○	○

Figure 35 : Exemple du formulaire proposé dans le classement du registre pour les réponses en français



L'expression "J'y crois pas !"

J'Y CROIS PAS ! est utilisé lorsque le locuteur veut exprimer son étonnement ou son indignation, en réaction à une action, une situation ou un comportement qui vient de se produire. Sa dimension expressive permet d'exprimer son indignation et plus rarement son contentement. Sur le plan discursif, l'expression permet de désapprouver ou manifester son incompréhension. Le locuteur fait comme s'il ne pouvait pas croire ce qu'il vient d'entendre, ce qui est une manière pour lui de dire que ce à quoi il réagit est inconcevable.

Script du dialogue

Dialogue 1

Enzo, Clarence et Paul discutent à propos de la sœur de Clarence. Celle-ci travaille dans une entreprise à l'ambiance de travail assez difficile.

Enzo : Ta sœur ça se passe bien son nouveau travail ?
Clarence : Bof, elle a un peu du mal avec son patron qui est trop exigeant, il la lâche pas d'une semelle ! Dimanche dernier, il l'a appelé pour qu'elle passe au bureau récupérer des dossiers pour lui.
Paul : **J'y crois pas !** elle devrait lâcher ce boulot pour un autre plus relax.
Clarence : Ouais ta raison !

Veuillez cliquer sur la vidéo pour écouter le dialogue si besoin :

Veuillez pour chaque proposition d'équivalent en japonais de "j'y crois pas !" choisir l'une des quatre valeurs proposées :

	Registre formel	Registre courant	Registre familier	Registre très famili…
信じられんわ	○	○	○	○
まじで	○	○	○	○
酷いもんだね	○	○	○	○

Figure 36 : Exemple du formulaire proposé dans le classement du registre pour les réponses en japonais

Les différents types de registres ont préalablement été définis de manière à harmoniser les critères de classement entre tous les évaluateurs. Nous avons ajouté dans les deux langues des exemples concrets afin de faciliter la compréhension des différences entre chaque type :

Classification du registre lié aux réponses du test en français

Dans le cadre de mon travail de thèse, je souhaiterais solliciter votre expertise de locuteur natif du français pour classer les réponses des informateurs ayant répondu à un test sur les expressions idiomatiques de la conversation orale quotidienne. Ce classement porte sur le registre de langue relatif aux réponses recueillies lors de ce test.

Dans un premier temps, je vous invite à lire le dialogue proposé au début de chaque rubrique, puis pour chaque proposition de réponse (en colonne sur le côté à gauche), choisir la valeur (en ligne en haut à droite) qui correspond le plus à la catégorie dont relève la réponse proposée.

Les caractéristiques des registres sont les suivantes :

=> Registre Formel = Réponses généralement longues et complexes. Politesse importante, distance relationnelle. Utilisation de termes compliqués voir spécialisés. Tournures très standards et grammaticalement proche de l'écrit.

Ex. : Il y a une éternité que nous ne nous sommes pas vus ! comment allez-vous ?

=> Courant = Lexique quotidien et habituel. Réponse de longueur standard, peu complexe. pas de politesse ou de relation particulière. Peu voire pas d'emploi de termes spécialisé. Tournures grammaticalement standards, mais étant proche de l'oral.

Ex. : ça fait longtemps ! comment vas-tu ?

=> Familier = Relâchement des contraintes grammaticales. Enoncés courts et très idiomatiques. Énoncés peu complexes sur le plan syntaxique. Grande variété d'emplois. Effacement de la distance relationnelle et spontanéité forte. Correction de la langue faible, peu d'usage standard.

Ex. : ça fait un bail ! ça va ?

=> Très familier = Similaire au familier, mais comporte fréquemment l'emploi de termes vulgaires. Utilisable uniquement avec une relation proche, sinon peut être ressenti comme une manière de parler vulgaire, voire dans certains cas insultante.

Ex. : putain, ça fait une plombe ! comment va ?

Ne réfléchissez pas trop et soyez le plus spontané possible dans votre catégorisation. Attention, les réponses qui comportent des erreurs d'orthographe ne sont pas nécessairement à classer dans le registre vulgaire.

Il est fortement recommandé de vous munir de l'exemplaire PDF du questionnaire ayant été utilisé pour récolter les réponses que vous allez évaluer aujourd'hui et qui est disponible via le lien suivant :

https://drive.google.com/file/d/11M2LCmN9q6lEw7qi2KMd-JAS50tkIYtD/view?usp=sharing

Vous pourrez vous référer à ce document pour avoir plus d'informations sur les dialogues et les consignes proposées. Chaque activité comporte un dialogue et chaque dialogue peut comporter de 1 à 4 items. Je précise pour chaque activité, le type compétence qui est évalué (production, reformulation, équivalent, choix de l'expression, etc...).

Figure 37 : Exemple des consignes et des définitions des 4 types de registres proposées aux évaluateurs français

Classification du registre lié aux réponses du test en japonais

Dans le cadre de mon travail de thèse, je souhaiterais solliciter votre expertise de locuteur natif du japonais pour classer les réponses des informateurs ayant répondu à un test sur les expressions idiomatiques de la conversation orale quotidienne. Ce classement porte sur le registre de langue relatif aux réponses recueillies lors de ce test.

Dans un premier temps, je vous invite à lire le dialogue proposé au début de chaque rubrique, puis pour chaque proposition de réponse (en colonne sur le côté à gauche), choisir la valeur (en ligne en haut à droite) qui correspond le plus à la catégorie dont relève la réponse proposée.

Les caractéristiques des registres sont les suivantes :

=> Registre Formel = Réponses généralement longues et complexes. Politesse importante, distance relationnelle. Utilisation de termes compliqués voir spécialisés. Tournures très standard et grammaticalement proche de l'écrit.

Ex. : ご無沙汰しております！お変わりありませんでしょうか？

=> Courant = Lexique quotidien et habituel. Réponse de longueur standard, peu complexe. pas de politesse ou de relation particulière. Peu voire pas d'emploi de termes spécialisé. Tournures grammaticalement standards, mais étant proche de l'oral.

Ex. : お久しぶりです！お元気ですか？

=> Familier = Relâchement des contraintes grammaticales. Enoncés courts et très idiomatiques. Énoncés peu complexes sur le plan syntaxique. Grande variété d'emplois. Effacement de la distance relationnelle et spontanéité forte. Correction de la langue faible, peu d'usage standard.

Ex. : 久しぶり！元気?

=> Très familier = Similaire au familier, mais comporte fréquemment l'emploi de termes vulgaires. Utilisable uniquement avec une relation proche, sinon peut être ressenti comme une manière de parler vulgaire, voire dans certains cas insultante.

Ex. : やば, めっちゃ久しぶりじゃん! 元気?

Ne réfléchissez pas trop et soyez le plus spontané possible dans votre catégorisation. Attention, les réponses qui comportent des erreurs d'orthographe ne sont pas nécessairement à classer dans le registre vulgaire.

Il est fortement recommandé de vous munir de l'exemplaire PDF du questionnaire ayant été utilisé pour récolter les réponses que vous allez évaluer aujourd'hui et qui est disponible via le lien suivant :

https://drive.google.com/file/d/11M2LCmN9q6lEw7qi2KMd-JAS50tkIYtD/view?usp=sharing

Vous pourrez vous référer à ce document pour avoir plus d'informations sur les dialogues et les consignes proposées. Chaque activité comporte un dialogue et chaque dialogue peut comporter de 1 à 4 items. Je précise pour chaque activité, le type compétence qui est évalué (production, reformulation, équivalent, choix de l'expression, etc...).

Il faut prévoir 30 min environ pour finaliser cette évaluation des résultats.

よろしくお願いします。

Figure 38 : Exemple des consignes et des définitions des 4 types de registres proposées aux évaluateurs japonais

La procédure pour les locuteurs japonais a été enrichie avec des précisions supplémentaires pour nous assurer que la procédure soit réalisée dans les meilleures conditions. Le système de synthèse des résultats des classifications est similaire à celui des accords interjuge. Nous avions dans les deux procédures de classement un nombre impair d'évaluateurs. Nous avons ainsi fonctionné de la manière suivante pour valider le choix du registre : la majorité (2/3 évaluateurs) valide le choix du registre. En cas d'égalité entre 3 registres, nous tranchons nous-même. Ce dernier cas n'a jamais été observé. De manière globale, nous avons retrouvé très peu de disparités entre les évaluateurs :

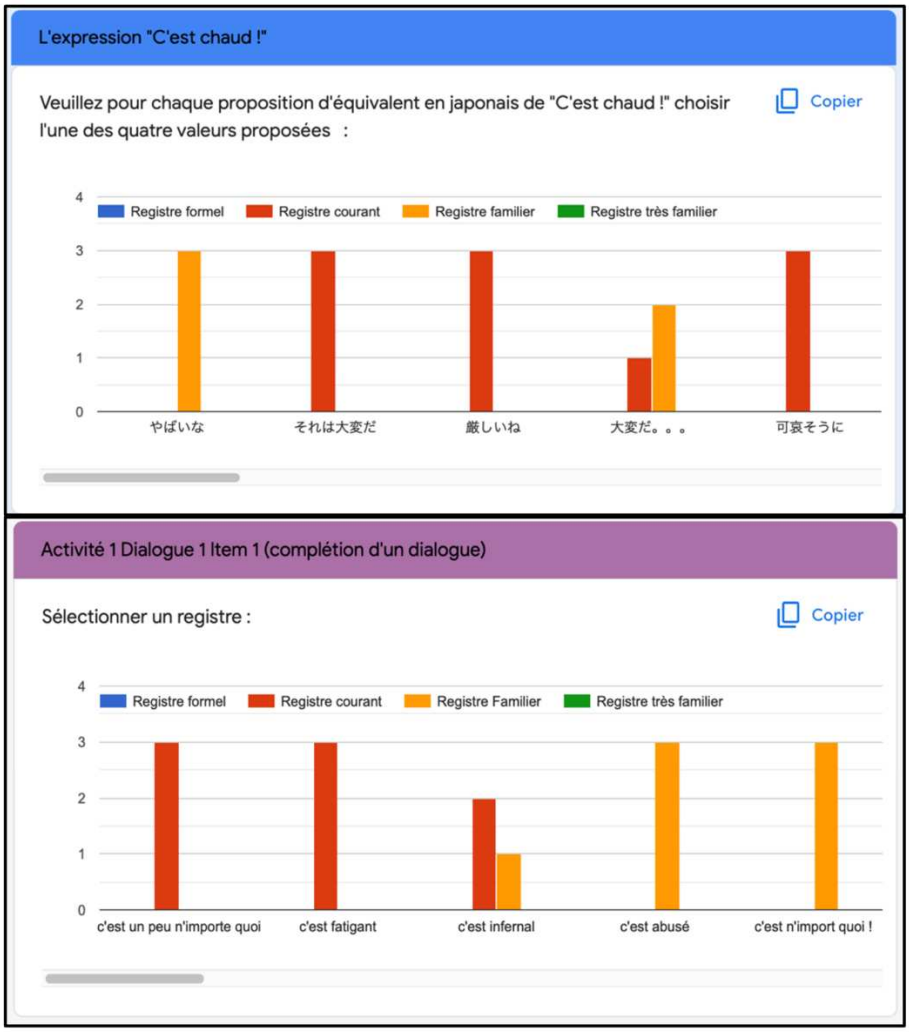


Figure 39 : Exemple des synthèses de classement dans les formulaires proposés aux évaluateurs français et japonais

Nous avons pu observer que les registres principalement identifiés sont le registre courant et familier pour les réponses en japonais, alors qu'il s'agit principalement du registre familier

pour le français. Nous observerons de manière plus approfondie ces tendances dans l'analyse qualitative du chapitre 3 de cette deuxième phase de notre étude.

3.3.3. Indice Phraséologique

L'*indice phraséologique* est une mesure du nombre de réponses présentant les caractéristiques d'un énoncé phraséologisé. Les critères retenus pour la mesure sont :

- La fixité de la forme (figement et semi-figement morphosyntaxique)
- Opacité ou la non-prédictibilité du sens et de la fonction (figement sémantique et pragmatique)
- Usage conventionnel et quotidien
- Fonction particulière dans la communication qui n'est pas explicitement exprimée dans le contenu linguistique.

Nous avons volontairement écarté certains critères trop complexes à traiter pour faciliter le travail de classement. Si l'on prend un exemple concret dans nos données, la séquence *c'était vraiment bien* sera considérée comme une séquence libre tandis que la séquence *c'est pas gagné !* sera considérée comme phraséologisée. Tous comme la mesure de la variabilité du registre, le classement des résultats en fonction de leur caractère préfabriqué ou non a été effectué à partir d'une procédure d'accord interjuge. Les groupes d'évaluateurs en français et en japonais sont les mêmes que pour le classement des résultats en fonction du registre. La procédure est également identique, mais avec une clarification des différents critères de la préfabrication de manière à expliciter le travail de classification :

Classification des réponses préfabriquées du test en japonais

Dans le cadre de mon travail de thèse, je souhaiterais solliciter votre expertise de locuteur natif du français pour classer les réponses des informateurs ayant répondu à un test sur les expressions idiomatiques de la conversation orale quotidienne. Ce classement porte sur le caractère préfabriqué et idiomatique des réponses recueillies lors de ce test.

Dans un premier temps, je vous invite à lire le dialogue proposé au début de chaque rubrique, puis pour chaque proposition de réponse (en colonne sur le côté à gauche), choisir la valeur (en ligne en haut à droite) qui correspond le plus à la catégorie dont relève la réponse proposée.

Les caractéristiques des expressions préfabriquées sont les suivantes :

=> Les expressions préfabriquées sont fixes. Les unités lexicales qui composent l'expression sont difficilement interchangeables. Le remplacement d'une unité lexicale composant l'expression peut par exemple changer son sens ou changer sa fonction.

Exemple : お気の毒 (préfabriqué) / お魂の毒 (pas préfabriqué, sens différent).

=> Les expressions préfabriquées ont un sens et une fonction non-prédictibles et difficile à interpréter sans le contexte. L'interprétation des termes qui composent l'expression préfabriquée ne suffit pas pour trouver son sens ou sa fonction.

Exemple : ムカつく！ (les mots composant l'expression ne permettent pas d'interpréter le sens d'indignation, de réprobation et de reproche).

=> Les expressions préfabriquées sont utilisées dans la vie de tous les jours de manière fréquente. Leur usage et leur sens est connu est conventionnel, c'est-à-dire qu'il est connu de tous les individus de la communauté de locuteur.

Exemple : やばい ! (fréquemment employé dans la conversation quotidienne pour exprimer une situation dangereuse, quelque chose de répréhensible ou au contraire, quelque chose d'amusant, de surprenant ou d'enthousiasmant.

=> Les expressions préfabriquées ont une fonction particulière dans la conversation quotidienne. Cette fonction n'est pas explicitement interprétable avec les unités lexicales qui composent l'expression.

Exemple : "冗談でしょう" est utilisé fréquemment pour exprimer son étonnement pour ce qui est dit ou fait, ainsi que son indignation.

Figure 40 : Exemple des consignes et des définitions des caractéristiques des structures préfabriquées proposées aux évaluateurs japonais

Classification des réponses préfabriquées du test en français

Dans le cadre de mon travail de thèse, je souhaiterais solliciter votre expertise de locuteur natif du français pour classer les réponses des informateurs ayant répondu à un test sur les expressions idiomatiques de la conversation orale quotidienne. Ce classement porte sur le caractère préfabriqué et idiomatique des réponses recueillies lors de ce test.

Dans un premier temps, je vous invite à lire le dialogue proposé au début de chaque rubrique, puis pour chaque proposition de réponse (en colonne sur le côté à gauche), choisir la valeur (en ligne en haut à droite) qui correspond le plus à la catégorie dont relève la réponse proposée.

Les caractéristiques des expressions préfabriquées sont les suivantes :

=> Les expressions préfabriquées sont fixes. Les unités lexicales qui composent l'expression sont difficilement interchangeable. Le remplacement d'une unité lexicale composant l'expression peut par exemple changer son sens ou changer sa fonction.

Exemple : Quelle galère ! (préfabriqué) / Quelle barque ! (pas préfabriqué, sens différent).

=> Les expressions préfabriquées ont un sens et une fonction non-prédictible et difficile à interpréter sans le contexte. L'interprétation des termes qui composent l'expression préfabriquée ne suffisent pas pour trouver son sens ou sa fonction.

Exemple : ça va pas la tête ? (les mots composant l'expression ne permettent pas d'interpréter le sens d'indignation, de réprobation et de reproche).

=> Les expressions préfabriquées sont utilisées dans la vie de tous les jours de manière fréquente. Leur usage et leur sens est connu est conventionnel, c'est-à-dire qu'il est connu de tous les individus de la communauté de locuteur.

Exemple : ça craint ! (fréquemment employé dans la conversation quotidienne pour exprimer une situation dangereuse, quelque chose de répréhensible ou quelque chose de ringard.

=> Les expressions préfabriquées ont une fonction particulière dans la conversation quotidienne. Cette fonction n'est pas explicitement interprétable avec les unités lexicale qui composent l'expression.

Exemple : "Tu m'en diras tant !" est utilisé fréquemment pour exprimer son manque d'intérêt pour ce qui est dit ou son incrédulité.

Figure 41 : Exemple des consignes et des définitions des caractéristiques des structures préfabriquées proposées aux évaluateurs français

L'indice phraséologique est élaboré en calculant le pourcentage de formes préfabriquées retrouvées sur le total des réponses enregistrées pour chaque cohorte, et ce à l'aide d'un produit en croix. Un indice phraséologique faible (indice < 40% soit 14 formes différentes sur 35 réponses) peut indiquer qu'un nombre important des informateurs disposent d'une palette lexicopragmatique limitée. Dans le cas de l'activité 3 (équivalence), cela peut également évoquer l'absence d'équivalent dans la langue maternelle. Tout comme les précédentes mesures, nous n'avons pas observé de discordances fortes entre les résultats. Toutefois, nous avons remarqué que l'accord interjuge pour les résultats en japonais a donné

lieu à moins de consensus total (3 évaluateurs d'accord sur la catégorisation) que celui des résultats en français. Nous supposons ici que la notion de préfabrication est moins évidente à observer en japonais en raison des spécificités morphosyntaxiques et sémantiques propres à la langue japonaise qui rendent moins visibles les marques de cette préfabrication.

3.3.4. Fréquence des formes employées dans le test

La fréquence des formes employées est une mesure permettant d'indiquer les 3 formes les plus fréquentes classées par ordre décroissant. Le nombre d'occurrences est affiché à côté de la forme en question. Pour l'activité 3 (équivalence), les résultats en japonais les plus fréquents sont également indiqués. Cette fréquence est réalisée par simple comptage et classement des formes retrouvées dans les réponses des LN et des LNN. Le but de cette mesure est de pouvoir comparer les tendances d'usage ou d'observer les différences de stratégies pragmatiques entre les deux cohortes. Cette mesure est d'autant plus intéressante lorsqu'elle est observée en parallèle de l'indice phraséologique ou la variabilité du registre.

3.3.5. Taux de compréhension

Le taux de compréhension correspond à une mesure établie à partir du questionnaire d'autoévaluation systématiquement proposé aux LNN après la passation du test. Nous avons essayé de proposer ici une mesure de la capacité de compréhension des PhPex et de leur contexte d'usage. Les questions d'autoévaluation proposées sont les suivantes :

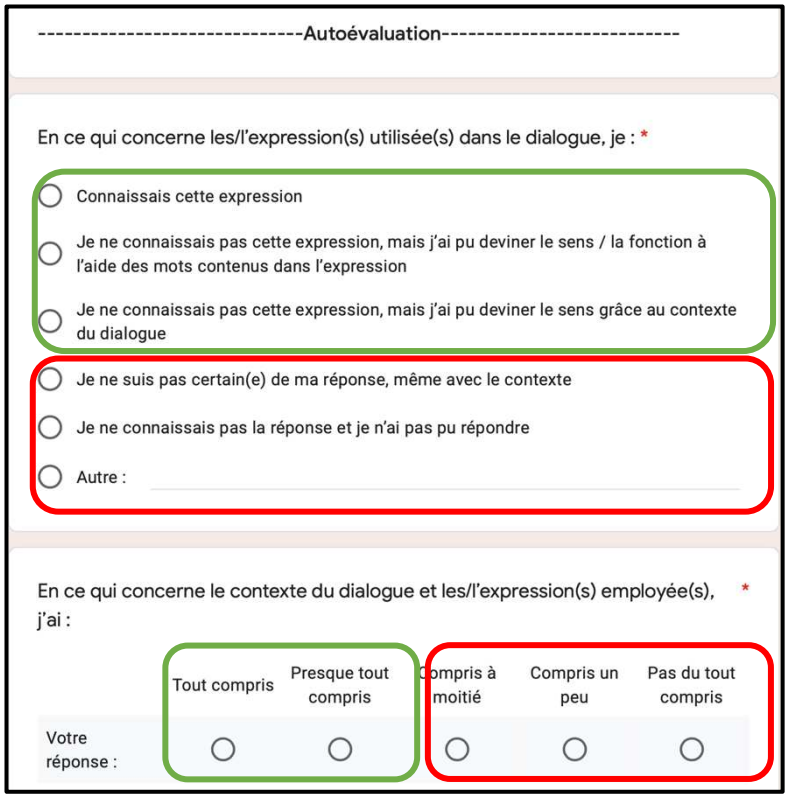


Figure 42 : Exemple des questions d'autoévaluations proposées dans le test lexicopragmatique

Nous avons ensuite établi un seuil à partir duquel la réponse choisie correspond à la compréhension de l'expression et du contexte. Les trois premiers choix de réponse de la première question et les deux premiers choix de l'échelle de Likert sont considérés comme relevant de la compréhension (en vert). Le reste des items proposés sont considérés comme relevant de la non-compréhension (en rouge). En suivant ce principe, nous avons établi des ratios du nombre d'informateurs appartenant à chaque catégorie sur le nombre total de participants à l'expérimentation. Ce ratio a été calculé de deux points de vue. D'une part un ratio individuel (par informateur) et un ratio collectif (toute la cohorte) ont été calculés. D'autre part, le ratio calculé pour ces deux points de vue a trois niveaux de granularité : un ratio par item, par activité et sur l'ensemble du test. Cette mesure est tout à fait intéressante lorsque comparée à la mesure du taux de réussite. Nous le verrons dans l'analyse qualitative, un décalage significatif entre le score obtenu au test et le taux de compréhension peut permettre de rendre compte de l'état de développement de la palette lexicopragmatique du LNN. Une confrontation de ces indicateurs avec l'indice phraséologique et la variabilité lexicale peut également donner des résultats intéressants dans la caractérisation de la variété non native observée dans ce travail de recherche.

Dans ce premier chapitre de la seconde phase de notre étude, nous avons évoqué les spécificités du protocole mis en œuvre et les modalités d'évaluation et de calcul des différentes mesures qui seront proposées dans l'analyse qualitative. Dans le chapitre qui suit, nous allons proposer une analyse statistique en deux temps. Le premier temps permettra d'observer les spécificités quantitatives des résultats entre les LN et LNN. Dans un second temps, nous tenterons de mesurer quels seraient les facteurs externes au test qui pourraient avoir un effet sur le développement de la compétence lexicopragmatique des LNN observés.

Chapitre 11 - Analyse quantitative et statistique des données

Dans ce chapitre, nous allons proposer un travail statistique à deux niveaux. D'une part, il s'agira d'étudier les spécificités quantitatives des résultats au test observés entre les LN et les LNN. D'autre part, nous tenterons d'approfondir la caractérisation des spécificités de la variété linguistique non native observée en étudiant de manière statistique la relation entre les résultats au test observés et les différents critères sociobiographiques associés à la cohorte des LNN. La première partie quantitative de ce chapitre se fonde sur la mesure d'acceptabilité des résultats obtenue lors de l'accord interjuge. La seconde partie constituera un travail de statistique inférentielle fondé sur les données récoltées dans le questionnaire sociobiographique proposé avant la passation du test. Voici un aperçu de ce questionnaire sociobiographique :

Informations sociobiographiques

1 - Age :
2 - Sexe : Femme | Homme | Autres
3 - Occupation : Salarié | Étudiant | Autre
4 - Si vous êtes étudiant, quels sont votre niveau d'étude et votre spécialité ?
5 - Si vous êtes salarié, dans quel domaine travaillez-vous ?
6 - Avez-vous déjà obtenu un diplôme universitaire en France ? Oui | Non
7 - Si oui, précisez quel type de diplôme.
8 - Depuis combien de temps étudiez-vous le français ?
9 - Avez-vous déjà fait évaluer votre niveau ? Oui | Non
10 – Si oui, quel test avez-vous passé ? TCF | DELF | DALF | Autres (Précisez :)
11 - Si oui, précisez votre niveau : A1 | A2 | B1 | B2 | C1 | C2 | Natif
12 - Avez-vous déjà séjourné / séjournez-vous en France ? Oui | Non
13 - Si oui, pourriez-vous indiquer la durée de votre séjour ou depuis combien de temps vous vivez en France ?
14 - De quelle manière avez-vous appris principalement le français ?
Au lycée ou à l'Université | Tout seul | En vivant en France |Centre culturel | Institut Franco-japonais | école de langues ou cours privé
15 - Discutez-vous souvent en Français ? Oui | Non
16 - Si oui, combien de fois par semaine / jour en moyenne ?

	Par jour	Par semaine	Par mois	Par an
En moyenne je discute une fois				
En moyenne je discute deux ou trois fois				
En moyenne je discute plusieurs fois				

17 - Avec qui parlez-vous français ? (Plusieurs choix possibles)
Professeurs | Camarades de cours | Amis français | Conjoint(e) | Amis japonais | Personnels de magasin ou de services quand je vais en ville | Collègues de travail
18 - Dans quel contexte discutez-vous le plus souvent en français ?
Entre amis ou proches | En contexte professionnel | En contexte scolaire ou académique

Figure 43 : Questionnaire sociobiographique proposé aux informateurs[139]

[139] Il s'agit d'une retranscription du formulaire proposé en ligne.

Les réponses de ce questionnaire ont ensuite été formalisées dans un tableau (Annexe 6). À partir de la formalisation de ces réponses, nous avons établi un jeu de variable qui sera employé lors des mesures inférentielles. Les informations relatives aux variables et à la procédure de calcul statistique seront détaillées plus bas dans la deuxième partie de ce chapitre. Nous allons à présent dans ce qui suit observer les aspects quantitatifs de nos résultats.

1. Étude quantitative des résultats obtenus au test lexicopragmatique

Nous rappelons que nous parlerons de *score de réussite* pour exprimer le nombre de réponses jugées acceptables sur le nombre total de questions proposées dans le test (par ex. : Locuteur X = 20/22 réponses au test sont acceptables). Ce score de réussite est ensuite transformé en pourcentage à l'aide d'un produit en croix afin d'obtenir un *taux de réussite* (par ex. : Locuteur X = 90 % de réussite). Voici un tableau de synthèse reprenant les taux de réussite par cohorte :

<table>
<tr><th colspan="3"></th><th colspan="3">LNN</th><th colspan="3">LN</th></tr>
<tr><th colspan="3"></th><th>Item</th><th>Activité</th><th>Test</th><th>Item</th><th>Activité</th><th>Test</th></tr>
<tr><td rowspan="2">Activité 1 (complétion)</td><td>Dialogue 1</td><td>C'est + [au choix]</td><td>83%</td><td rowspan="2">76%</td><td rowspan="19">64%</td><td>100%</td><td rowspan="2">99%</td><td rowspan="19">89%</td></tr>
<tr><td>Dialogue 2</td><td>Sérieux ? +[au choix]</td><td>69%</td><td>97%</td></tr>
<tr><td rowspan="6">Activité 2 (reformulation)</td><td>Dialogue 1</td><td>C'est le pied !</td><td>77%</td><td rowspan="6">65%</td><td>97%</td><td rowspan="6">95%</td></tr>
<tr><td>Dialogue 2</td><td>C'est pas possible !*</td><td>83%</td><td>100%</td></tr>
<tr><td rowspan="2">Dialogue 3</td><td>Faut pas pousser !</td><td>34%</td><td>94%</td></tr>
<tr><td>C'est abusé !</td><td>60%</td><td>94%</td></tr>
<tr><td rowspan="2">Dialogue 4</td><td>C'est pas possible !</td><td>49%</td><td>86%</td></tr>
<tr><td>J'y suis pour rien !</td><td>89%</td><td>97%</td></tr>
<tr><td rowspan="5">Activité 3 (équivalence)</td><td>Dialogue 1</td><td>J'y crois pas !</td><td>80%</td><td rowspan="5">62%</td><td>97%</td><td rowspan="5">91%</td></tr>
<tr><td rowspan="2">Dialogue 2</td><td>C'est pas vrai !</td><td>77%</td><td>100%</td></tr>
<tr><td>C'est chaud !</td><td>37%</td><td>63%</td></tr>
<tr><td rowspan="2">Dialogue 3</td><td>C'est pas gagné !</td><td>43%</td><td>94%</td></tr>
<tr><td>Tu plaisantes ?</td><td>74%</td><td>100%</td></tr>
<tr><td rowspan="3">Activité 4 (QCM lexical)</td><td>Dialogue 1</td><td>Choix entre : [c'est génial ! /c'est le pied ! /c'est le bouquet ! / c'est une horreur !]</td><td>17%</td><td rowspan="3">41%</td><td>57%</td><td rowspan="3">80%</td></tr>
<tr><td>Dialogue 2</td><td>Choix entre : [c'est génial !/ c'est pas gagné ! / ça tombe bien ! / c'est une horreur !]</td><td>51%</td><td>89%</td></tr>
<tr><td>Dialogue 3</td><td>Choix entre : [Tu crains ! / Tu parles ! /</td><td>54%</td><td>94%</td></tr>
</table>

		T'inquiète ! / Tu l'as dit !]						
Activité 5 (QCM sémantique)	Dialogue 1	Tu es sérieux ?	63%	73%		60%	84%	
		C'est pas possible*	60%			69%		
	Dialogue 2	J'y crois pas !	80%			83%		
		Les bras m'en tombent !	94%			100%		
		C'est une honte !	71%			94%		
		C'est le pied.*	71%			100%		
% locuteurs en dessous de la moyenne des résultats			82%			43%		

Figure 44 : Tableau de synthèse des taux de réussite au test lexicopragmatique pour les deux cohortes[140]

Le tableau ci-dessus se lit sur trois niveaux pour chaque cohorte. Le premier niveau correspond au taux de réussite pour chaque question du test. Le deuxième niveau est celui du taux de réussite sur l'activité et le dernier niveau le taux de réussite sur l'ensemble du test. Les résultats enregistrés lors du dépouillage des réponses au test montrent à première vue une différence dans le taux de réussite au test lexicopragmatique avec un taux de 89% chez les LN contre 64% pour les LNN soit une différence de 25 %. Si l'on observe les résultats obtenus pour chaque activité, on s'aperçoit également d'une différence de 20% en moyenne entre les résultats des natifs et des non natifs. On remarque qu'une partie importante des taux de réussite au test sont inférieurs à la moyenne générale chez les LNN (82%). Cela laisse supposer une forte dispersion des résultats autour de la moyenne (des scores très faibles et des scores très élevés) avec une valeur médiane plus faible chez les locuteurs japonais. De même chez les LNN, nous observons qu'un peu moins de la moitié des taux de réussite ne semblent pas resserrés autour de la moyenne générale. Ci-dessous, un histogramme de ces différents résultats :

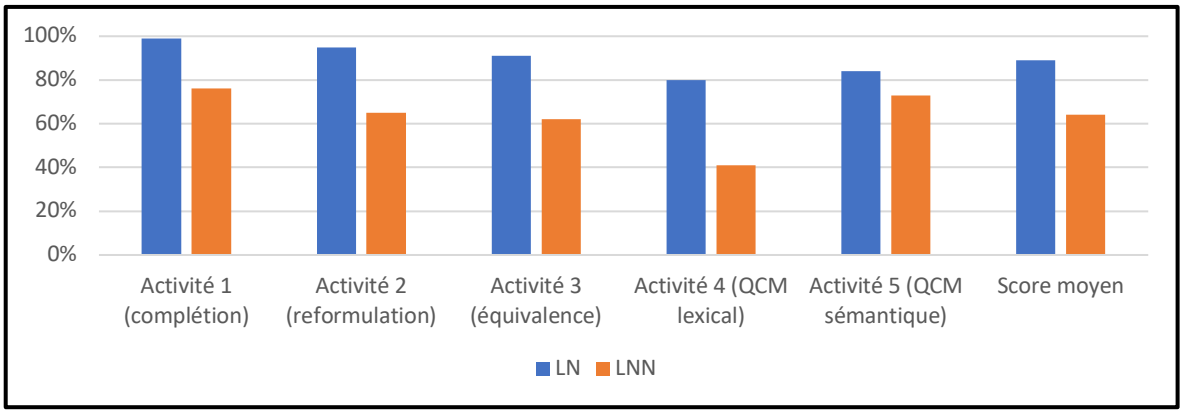


Figure 45 : Comparaison des taux de réussite par activité entre les LN et les LNN

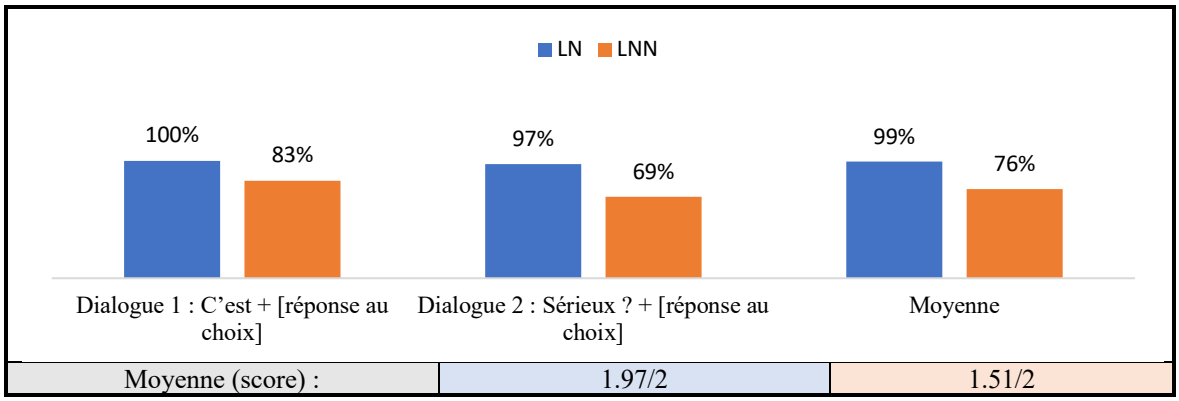
[140] Les astérisques mentionnent les expressions qui ont joué le rôle de distracteur.

L'activité 1 (production libre) a généré le plus haut taux de réussite (LN = 99% / LNN = 76%), l'activité 4 (complétion à choix multiple) a produit les plus faibles taux de réussite (LN = 80% / LNN = 41%). Deux hypothèses semblent expliquer cette disparité. Concernant l'activité de complétion, nous pensons que cela est impliqué par la grande flexibilité de réponses possibles et le peu de contraintes de la consigne. Cette flexibilité semble être un facteur de diminution des chances de produire une réponse inadéquate (aucune réponse précise ni équivalence n'étant attendue). Concernant l'activité de complétion à choix multiple, la difficulté semble liée à la tâche en elle-même. En effet, l'activité impose un inventaire restreint de choix possible dont il faut simultanément évaluer l'acceptabilité pour chaque contexte proposé. La charge cognitive et la complexité de la tâche demandée peuvent, selon nous, constituer des contraintes qui ont influencé le taux de réussite des LN et des LNN dans cette dernière activité[141]. Il y aura donc un biais externe lié à la complexité de la tâche à prendre en compte dans l'interprétation de nos résultats.

1.1. Analyse par activité

L'observation des taux de réussite par activité nous permet d'affiner notre traitement quantitatif et d'observer que certaines expressions posent particulièrement des difficultés pour les natifs et pour les non-natifs. L'identification de ces expressions nous permet ensuite d'isoler des zones de difficulté que nous traiterons plus en détail dans l'analyse qualitative.

1.1.1. Activité 1 (complétion)



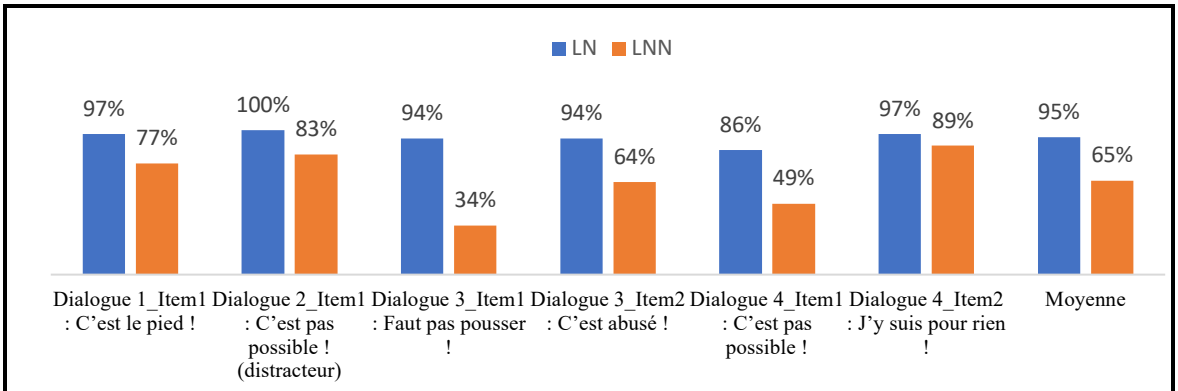
[141] Wang (2021) qui a employé un protocole similaire au notre a également constaté un faible taux de réussite pour les deux cohortes étudiées (français natifs et locuteurs chinois apprenants du français) dans ce type d'activité.

Écart type :	0.169(dispersion faible)	0.562(dispersion moyenne)

Figure 46 : Taux de réussite et écart type par item et par cohorte pour l'activité 1

Comme expliqué précédemment, la première activité (complétion) n'a pas posé de réelles difficultés aux informateurs des deux cohortes. Le taux d'acceptabilité des réponses pour le deuxième item de l'activité est un peu plus faible que le premier item. Nous le verrons dans l'analyse qualitative, ce taux plus bas est lié à une différence de stratégie pragmatique employée par les LNN qui a impliqué certaines anomalies illocutoires dans le contexte proposé. On peut également observer une dispersion moyenne des résultats autour de la moyenne des LNN alors que cette dispersion est beaucoup plus faible chez les LN.

1.1.2. Activité 2 (reformulation)



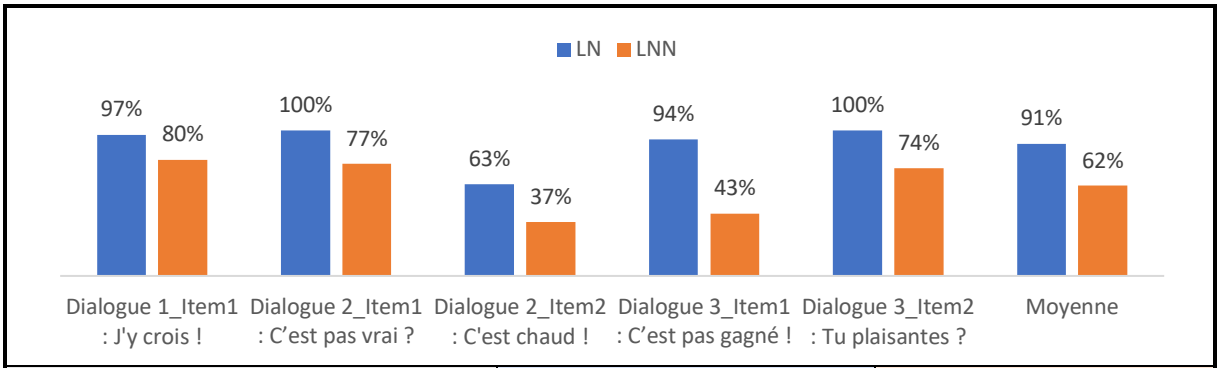
Moyenne (score) :	5.69/6	3.91/6
Écart type :	0.471 (faible dispersion)	1.36 (dispersion moyenne)

Figure 47 : Taux de réussite et écart type par item et par cohorte pour l'activité 2

On remarque pour cette activité 2 (reformulation) que quelques items ont engendré des scores plus bas que la tendance observable sur les autres items. Il s'agit notamment des expressions *Faut pas pousser !* et *C'est pas possible !*. Pour la première expression, la forte idiomaticité de l'expression et la faible prédictibilité de sa fonction illocutoire semblent être une des causes du taux de réponses inadéquates important chez les LNN. Nous nous posons également la question de savoir s'il existe des emplois de ce type en japonais sur lesquels les LNN auraient pu appuyer l'interprétation du PhPex. Pour la seconde expression, il pourrait vraisemblablement s'agir d'une confusion avec le sens littéral de *c'est pas possible* proposé dans la même activité en tant que distracteur. Toutefois, les LN ont aussi enregistré un score plus faible que la normale sur cette deuxième expression, ce qui nous laisse penser

que la difficulté pourrait bien résulter d'un autre facteur comme par exemple un manque d'attention, un effet de *speeding* ou un manque de clarté du contexte. Un examen plus approfondi dans l'analyse qualitative nous permettra d'observer cela plus en détail. Un autre détail important est la dispersion des résultats par rapport à la valeur médiane relevés chez les LNN.

1.1.3. Activité 3 (équivalence)

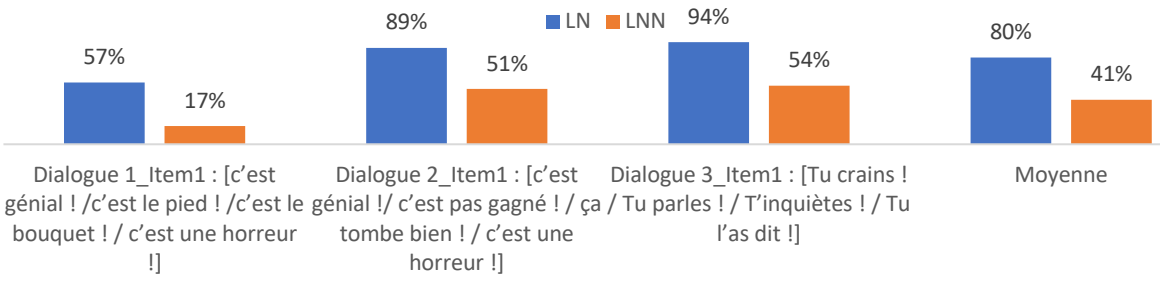


Moyenne (score) :	4.54/5	3.11/5
Écart Type :	0.611 (faible dispersion)	1.21 (dispersion moyenne)

Figure 48 : Taux de réussite et écart type par item et par cohorte pour l'activité 3

La troisième activité a donné lieu à des scores élevés du côté des LN, mais plus mitigés du côté des LNN. Nous remarquons que les expressions *c'est chaud !* et *c'est pas gagné !* ont tout particulièrement posé certaines difficultés. Concernant le PhPex *c'est chaud !*, les deux cohortes affichent un score faible en comparaison de la moyenne affichée. Une des hypothèses qui peut être avancée ici est que ce PhPex est à la fois très polysémique et sa fonction est non prédictible. L'emploi de ce PhPex peut selon nous avoir donné lieu à une très grande variété d'interprétation dans les deux cohortes. L'expression *c'est pas gagné !* a quant à elle posé de grandes difficultés aux LNN. Nous pensons ici que cela est dû à plusieurs facteurs. D'une part, l'expression présente une dimension pragmatique complexe, mais peu prédictible (expression de l'incertitude, déception voire indignation). D'autre part, il nous a semblé que cette expression ne disposait d'aucun équivalent fonctionnel en japonais sur lequel s'appuyer lors de l'interprétation. Ces difficultés se remarquent également dans l'écart type enregistré pour les LNN qui indique une grande dispersion des scores autour de la moyenne de la cohorte.

1.1.4. Activité 4 (QCM lexical)

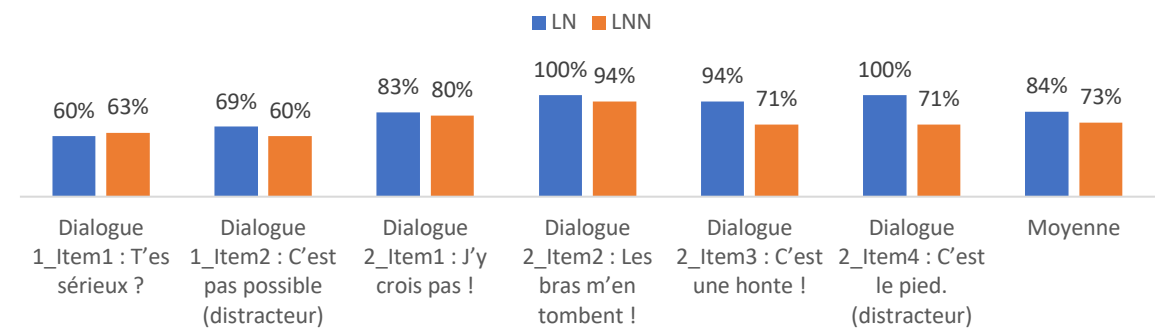


Moyenne (taux /score) :	2.40/3	1.23/3
Écart type :	0.651 (dispersion moyenne)	0.973 (dispersion forte)

Figure 49 : Taux de réussite et écart type par item et par cohorte pour l'activité 4

L'activité 4 affiche des taux de réussite plus bas que la moyenne constatée jusqu'à présent. Les taux de réussite obtenus pour chaque dialogue montrent que le premier dialogue a particulièrement posé problème aux deux cohortes avec un taux de réussite très inférieur à la moyenne enregistrée. Les deux autres dialogues ont également été source de difficulté pour les LNN. Les écarts types enregistrés pour les deux cohortes semblent assez discordants et évoquent une dispersion hétérogène des résultats autour de la moyenne enregistrée pour chaque groupe. L'analyse qualitative des résultats nous permettra d'explorer plus finement les raisons de ces taux anormalement bas en comparaison de ceux retrouvés dans les activités précédentes. Cependant, nous pouvons d'ores et déjà formuler l'hypothèse d'une complexité liée à la tâche demandée dans le cadre de cette activité. Cette complexité peut avoir constitué une contrainte supplémentaire qui a grandement influencé les résultats des deux cohortes.

1.1.5. Activité 5 (QCM sémantique)



Moyenne (taux /score) :	84%/5.09	73%/4.40

Écart type:	0.981 (dispersion faible)	1.29 (dispersion moyenne)

Figure 50 : Taux de réussite et écart type par item et par cohorte pour l'activité 5

Concernant l'activité 5, les résultats semblent assez élevés. Le premier dialogue montre cependant des difficultés importantes dans les deux cohortes. Ce dialogue comportait deux expressions : *t'es sérieux ?* et *c'est pas possible* (ici, un distracteur). Une hypothèse plausible pour ce qui concerne les résultats des LN serait un effet de la fatigue cognitive ressentie en fin de test alliée à un phénomène de précipitation dans les réponses. En effet, les deux expressions ne comportaient aucune ambigüité et leur sens ainsi que leur fonction étaient clairement identifiables. En outre, il semble curieux que les informateurs aient rencontré autant de difficultés étant donné la relative intuitivité des expressions proposées. Nous soupçonnons ici aussi un effet des définitions proposées dans le QCM sémantique de cette activité. Il est possible qu'elles n'étaient pas assez explicites pour permettre de sélectionner clairement le sens adéquat. Cette hypothèse s'est confirmée lors des entretiens post-test où 3 informateurs ont évoqué une difficulté à comprendre certaines des affirmations proposées dans le QCM sémantique. Pour ce qui est des scores des LNN, il peut s'agir soit d'une maitrise encore imparfaite de la première expression soit d'une difficulté à discerner le sens adéquat du distracteur. Enfin, mentionnons que les résultats statistiques des LNN pour cette activité affichent un écart type qui indique une dispersion des résultats autour de la moyenne. Cette dispersion est souvent le signe d'une forte hétérogénéité des résultats obtenus au test, et peut être une trace de la disparité entre les niveaux d'expertise des informateurs LNN.

En résumé, on constate de manière surprenante que les natifs qui *à priori* devraient disposer d'une connaissance approfondie de leur LM ont enregistré pour certains items un taux de réussite assez faible. Avec une certaine prise de recul, on s'aperçoit néanmoins que cela n'est pas si improbable si l'on considère les deux hypothèses évoquées précédemment : 1) la charge cognitive liée au type d'activité et 2) le fait que pour les phraséologismes pragmatiques, il semble courant que le natif ne puisse pas spontanément proposer un équivalent adapté ou une définition précise du sens et de l'emploi du phraséologisme en question[142]. Pour cette deuxième hypothèse, il s'agit notamment des spécificités d'emploi des PhPex qui, comme nous l'avons évoqué dans le premier chapitre sur les

[142] Le locuteur natif peut en revanche savoir employer ou interpréter ces expressions, sans pour autant être capable de les reformuler.

phraséologismes, ont tendance à être polyfonctionnels et à dépendre de plusieurs facteurs contextuels[143]. C'est notamment cette forte dimension actionnelle dépendante de certains schémas conversationnels et la relative complexité de leur emploi expressif et discursif qui complexifient la description du sens et de l'emploi des PhPex. Ce faisant, la reformulation ou l'équivalence ne constituent pas un exercice qui peut être facilement réalisé dans un court laps de temps. Le tableau ci-dessous décrit les indicateurs statistiques généraux pour les taux de réussite individuels obtenus par chacune des cohortes à l'ensemble du test :

	LNN	**LN**
Moyenne	64.4	89.3
Médiane	59	91
Écart-type	15.8	5.87
Variance	251	34.4
Minimum	32	77
Maximum	91	100

Tableau 9 : Statistiques générales descriptives pour les scores de réussite de chaque cohorte

En outre, nous présentons en complément deux diagrammes modélisant la différence de dispersion et d'étendue des plages de résultats entre chaque cohorte :

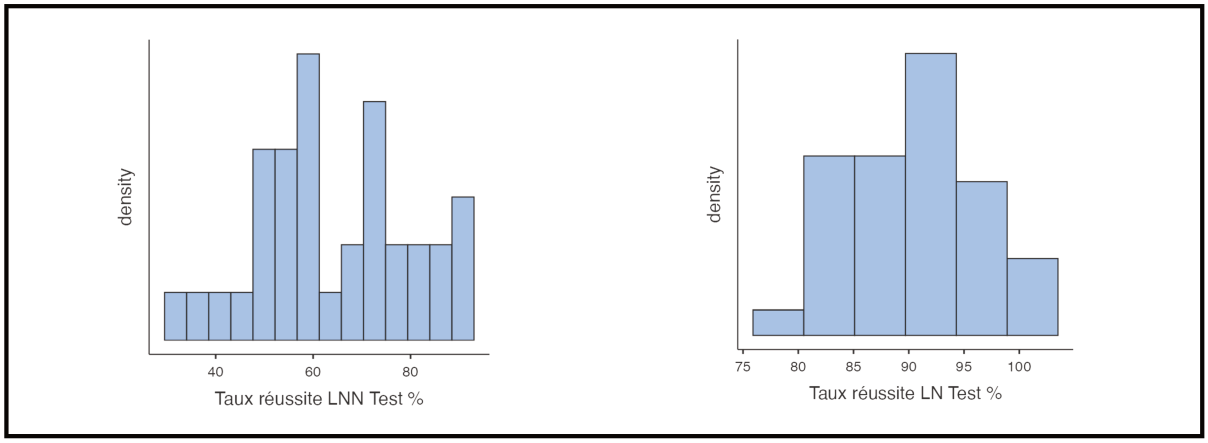


Figure 51 : Dispersion des résultats au test lexicopragmatique des LNN

La densité exprime le nombre d'informateurs ayant le taux de réussite exprimé en abscisse. On observe que l'étendue de la plage de résultats des LNN (17%-94%) est bien plus

[143] Cette complexité inhérente à la définition des phraséologismes pragmatiques a été évoquées à plusieurs reprise lors des journées d'étude PHIOE 2022 (Phraséologie des interactions orales et écrites) qui se sont déroulées le 14 et 15 juin à l'université de Grenoble (Laboratoire LIDILEM et CRIFUQ).

importante que celle des LN (77%). L'épaisseur du bâton[144] montre que la répartition des locuteurs de la cohorte se fait en sous-groupes de même score plus resserrés autour de la moyenne exprimée dans le tableau plus haut. À contrario, plus les bâtons sont fins et nombreux, plus les résultats seront dispersés et regroupés en sous-groupe de score plus épars. La lecture de tous ces éléments permet de confirmer la forte dispersion des résultats des LNN par rapport à ceux des LN, et de valider l'hypothèse d'une certaine hétérogénéité des niveaux de compétence lexicopragmatique chez les LNN. Nous verrons toutefois dans l'analyse qualitative que ces résultats de statistique descriptive restent à nuancer à l'aide de l'analyse qualitative.

Après avoir discuté dans cette partie les éléments statistiques généraux concernant les résultats obtenus au test par les LN et les LNN, nous allons dans la section qui suit tenter de déterminer s'il existe un effet des variables liées à l'apprentissage et à l'usage du français sur les taux de réussite enregistrés chez les LNN.

2. Étude statistique des variables liées à l'apprentissage et les variables liées à la pratique du français

Dans cette section, nous allons observer les résultats d'une analyse statistique corrélative appliquée à des variables indépendantes spécifiques aux LNN et réparties en deux sous-catégories : les *variables liées à l'apprentissage* et les *variables liées à la pratique* du français. Ces variables indépendantes (ou *explicatives*) ont été identifiées grâce à des informations personnelles collectées auprès des informateurs non natifs puis consignées dans un fichier de métadonnées (Annexe 6). Nous rappelons que cette collecte d'informations personnelles concernant les LNN a été effectuée via un questionnaire sociobiographique proposé avant la passation du test. La variable dépendante (ou *à expliquer*) se compose des scores de réussite relevés dans le test, le principe étant de déterminer s'il est possible d'établir une influence des variables indépendantes observées sur les scores de réussite obtenus. Nous employons ainsi deux mesures d'inférence statistique qui nous permettent d'estimer le risque d'inférer une tendance ou une relation constatée dans notre échantillon à toute une

[144] En termes plus statistiques, les surfaces des bâtons sont proportionnelles aux effectifs.

population supposée ayant les caractéristiques de notre échantillon. La marge de risque ou *p-value* conventionnellement admise en SHS correspond à une probabilité de valider l'hypothèse nulle[145] inférieure ou égale à 0,05 ($p \leq 0.05$ ou 5 %). Si la p-value est inférieure ou égale à 0,05, on pourra alors estimer que l'on peut rejeter l'hypothèse nulle et donc qu'il y a une possible influence de la variable explicative sur la variable à expliquer. Il est alors admis de parler de *significativité* des résultats dans le cas de $p \leq 0.05$ ou de *non-significativité* dans le cas de $p > 0.05$. Mentionnons que ce seuil du p-value peut varier en fonction de la discipline ou du phénomène observé, et notamment du degré maximum de risque d'erreur accepté.

Pour chaque variable indépendante, nous comparons plusieurs sous-groupes constitués à partir de la cohorte principale des LNN. Ces sous-groupes sont organisés en fonction des différentes valeurs afférentes (ou modalités) de la variable[146]. Nous comparons ensuite les scores de réussite obtenus au test entre les individus de ces sous-groupes pour observer s'il y a une différence significative des résultats. Afin de mesurer la significativité de ces différences de résultats au test, nous avons appliqué deux tests statistiques qui permettent de vérifier la significativité des différences de moyenne entre deux ou plusieurs sous-groupes, c'est-à-dire de généraliser une différence de moyennes de l'échantillon à la population étudiée (dans notre cas les apprenants japonophones du français de niveau avancé). En principe, la mesure de ce type de relation statistique passe par des tests dits *paramétriques* qui partent de la condition d'une distribution *normale*[147] des données avec symétrie de la distribution des valeurs de la variable de part et d'autre de la médiane et *kurtosis* (mesure de l'aplatissement de la distribution) normal et peu aplati. Lorsque la condition de normalité est réalisée, le test T de Student est employé pour mesurer l'effet d'une variable indépendante à deux modalités sur une variable dépendante, et le test ANOVA est quant à lui utilisé pour mesurer l'effet d'une variable indépendante à deux ou plus modalités.

Comme vu précédemment (cf. *Figure 51,* p. 258), la distribution des valeurs de notre variable dépendante est fortement dispersée autour de la moyenne. Afin de nous assurer de

---

[145] L'hypothèse nulle correspond au cas où le résultat observé ne subit aucun effet de la variable appliquée et est le fruit du hasard.
[146] Par exemple, les valeurs afférentes / modalités de la variable *niveau de certification* sont les valeurs *B1, B2, C1, C2.*
[147] *Hypothèse de normalité* : la distribution des valeurs de l'échantillon suit une courbe gaussienne lors de la modélisation en histogramme avec des queues de distribution qui tendent vers zéro de façon asymptotique, comme si elles rencontraient l'abscisse à l'infini.

la normalité des distributions de nos variables, nous avons regardé, en plus de la dispersion, la symétrie et l'aplatissement (kurtosis) de nos données. Le tableau ci-dessous reprend les données de distribution de notre variable dépendante :

	LNN
Moyenne	64.4
Écart-type	15.8
Variance	251
Asymétrie	0.001
Kurtosis	-0.712

Tableau 10 : Statistiques descriptives de la distribution pour la variable des scores de réussite chez les LNN

Nous avons également généré une modélisation en histogramme de cette distribution avec la courbe gaussienne que devrait suivre en principe cette distribution si elle validait l'hypothèse normale :

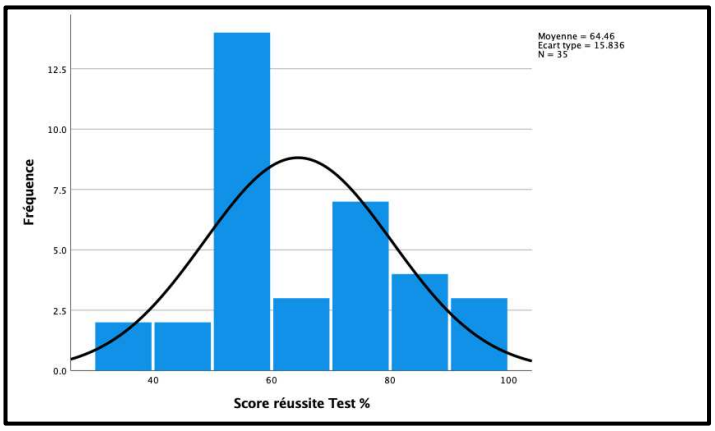


Figure 52 : Distribution des résultats au test lexicopragmatique des LNN

Les résultats affichés dans le tableau 10 montrent, en plus d'une forte dispersion, une asymétrie à gauche et un aplatissement anormal des valeurs de la variable dépendante étudiée. Le graphique montre quant à lui assez clairement que la distribution ne suit pas la courbe gaussienne théorique de la variable avec un pic de fréquence à gauche de la médiane qui dépasse très largement la courbe théorique. La distribution de la variable des scores de réussite au test chez les LNN ne semble donc pas suivre une distribution normale. Nous avons pratiqué cette analyse de la distribution sur tous les sous-groupes liés aux modalités des variables étudiées, analyse qui a confirmé que la distribution de ces sous-groupes ne

validait pas non plus l'hypothèse de normalité. Les spécificités de la distribution des données de notre variable ne semblent donc pas nous permettre d'employer les tests paramétriques cités supra.

Dans le cas où la distribution des valeurs de la variable étudiées ne suivrait pas la loi de normalité, il existe des tests statistiques alternatifs dits *non-paramétriques*. Ce faisant, nous avons appliqué à nos mesures un test *Test U de Mann-Whitney* pour les variables à deux modalités et un test *Anova one-way* (test Kruskall-Wallis) pour les variables à trois modalités ou plus. Le test U de Mann-Whitney permet d'étudier l'effet d'une variable indépendante nominale[148] ou ordinale[149] à deux modalités sur une variable dépendante quantitative[150]. Le test ANOVA One way (test Kruskall-Wallis) fonctionne sur le même principe que le test U de Mann-Whitney, mais permet d'étudier l'effet d'une variable indépendante nominale ou ordinale à 3 ou plus modalités sur une variable dépendante quantitative. Notons que notre effectif compte 35 informateurs, ce qui constitue un seuil acceptable pour l'utilisation de tests inférentiels non-paramétriques. Les tests ont été réalisés sur l'environnement *R*[151] à l'aide de l'interface open source *Jamovi*[152] pour faciliter la formalisation du jeu de données et son traitement. Voici un premier tableau récapitulant les variables observées :

---

[148] Variable attribuant une étiquette, par exemple le genre « homme » ou « femme »
[149] Variable attribuant un ordre mais pas de valeur numérique, par exemple un niveau de certification CECRL : A1<A2<B1<B2<C1<C2.
[150] Variable attribuant une valeur numérique à un objet, par exemple dans notre étude, un score de réussite par informateur.
[151] https://www.r-project.org/
[152] https://www.jamovi.org/

		Niveaux de la variable	Hypothèse de la variable[153]
Variables liées aux conditions d' apprentissage	**Variable niveau d'étude**	Licence Master Doctorat	Partant du fait que tous les cursus suivis par les informateurs sont fortement liés à la langue française, existe-t-il une influence du niveau d'étude universitaire sur le score de réussite au test lexicopragmatique ?
	Variable lieu d'université	Université française Université japonaise	Le fait d'être diplômé[154] d'une université française ou japonaise peut-il avoir un effet sur le score de réussite au test lexicopragmatique ?
	Variable spécialité d'étude	Études françaises/littéraires Sciences Humaines Sciences du langage	Existe-t-il un effet de la spécialité d'étude sur le score de réussite au test lexicopragmatique ? Être spécialisé dans des disciplines où la maitrise du français joue un rôle central, telles que la littérature ou la linguistique, a-t-il un effet sur les résultats ?
	Variable niveau certification CECRL	B2 C1 C2	Existe-t-il un effet du niveau de certification en français sur le score de réussite au test ? Nous souhaitons observer ici si un niveau d'expertise attesté est nécessairement synonyme d'une compétence lexicopragmatique élevée.
	Variable durée d'apprentissage du français	1 à 5 ans 6 à 10 ans 11 à 14 ans 15 ans et plus	Existe-t-il un effet de la durée d'apprentissage du français sur le score de réussite au test ? Un apprentissage long de la langue cible est-il nécessairement signe d'une compétence lexicopragmatique développée ?
	Variable cadre d'apprentissage du français	Autonome Guidé académique Guidé non académique	Le cadre d'apprentissage peut-il avoir un effet sur le score de réussite au test ? Le fait d'être guidé ou non lors de l'apprentissage peut-il avoir un effet sur le développement de la compétence lexicopragmatique ?
Variables liées aux conditions d' usage	**Variable statut socioprofessionnel**	Académique Professionnel	Le statut socioprofessionnel dans lequel l'informateur évolue peut-il avoir un effet sur le score de réussite au test ?
	Variable du lieu de résidence	France Japon	Le fait de vivre en France ou au Japon peut-il avoir un effet sur le score au test ? Le fait d'être en immersion prolongée en France peut-il avoir un impact sur le développement de la compétence lexicopragmatique ?
	Variable de la durée du séjour en France	Moins d'un an Entre 1 et 3 ans Entre 4 et 7 ans 8 ans et plus	La durée du séjour en France peut-elle avoir un effet sur le score de réussite au test ? Un séjour de longue durée est-il nécessairement synonyme d'une compétence lexicopragmatique avancée en langue cible ?
	Variable du type de contexte de conversation en français	Académique Personnel Professionnel	Le contexte principal d'usage du français peut-il avoir un effet sur le score de réussite au test ? Quel serait le contexte le plus propice à l'obtention d'un score élevé ?
		Plusieurs fois par mois et moins	

[153] Nous rappelons ici que ces hypothèses et ces résultats ne s'appliquent uniquement qu'au profil de locuteur de la cohorte étudiée, mais pourront être généralisés en reproduisant le protocole employé à d'autres langues.
[154] Seul le dernier diplôme obtenu est pris en compte dans notre étude statistique.

	Variable de la fréquence de la conversation en français	Plusieurs fois par semaine Plusieurs fois par jour	La fréquence d'interaction avec des locuteurs natifs du français peut-elle avoir un effet sur le score de réussite au test ? Une exposition intensive génère-t-elle nécessairement une compétence lexicopragmatique avancée ?
	Variable du type d'interlocuteur en français	Familier / proche Professionnel / académique Locuteur lambda	La nature des interlocuteurs auxquels l'informateur est fréquemment confronté lors de la conversation quotidienne peut-elle avoir un effet sur les scores de réussite au test ?

Tableau 11 : Inventaire des variables observées sur les résultats au test des LNN

Le tableau suivant récapitule les résultats obtenus après validation statistique :

		Niveau de la variable	Effectif par niveau de la variable (/35)	Moyenne Score par groupe (/22)	Écarts types pour chaque groupe	Valeur Test de significativité (P < 0.05 = significatif)
Variables liées aux conditions d' apprentissage	**Variable niveau d'étude**	Licence	8	15.38	3.628 (faible dispersion)	P = 0.322 (Test Kruskall-Wallis, non significatif)
		Master	16	14.50	3.378 (faible dispersion)	
		Doctorat	11	12.82	3.307 (faible dispersion)	
	Variable lieu d'université	Université française	16	13.81	3.391 (faible dispersion)	P = 0.683 (Test U de Mann-Whitney, non significatif)
		Université japonaise	19	14.47	3.596 (faible dispersion)	
	Variable spécialité d'étude	Études françaises/littéraires	14	15.21	3.093 (faible dispersion)	P = 0.245 (Test Kruskall-Wallis, non significatif)
		SHS	6	14.33	2.658 (faible dispersion)	
		ScL	15	13.13	3.944 (faible dispersion)	
	Variable niveau certification CECRL	B2	7	13.42	3.599 (faible dispersion)	P = 0.241 (Test Kruskall-Wallis, non significatif)
		C1	13	13.30	2.463 (faible dispersion)	
		C2	15	15.26	4.026 (faible dispersion)	
	Variable durée d'apprentissage du français	1 à 5 ans	7	13.42	3.599 (faible dispersion)	**P = 0.051 (Test Kruskall-Wallis, significatif)**
		6 à 10 ans	12	13.25	2.832 (faible dispersion)	
		11 à 14 ans	7	13.14	2.854 (faible dispersion)	
		15 ans et plus	9	16.77	3.734 (faible dispersion)	
	Variable cadre d'apprentissage du français	Autonome	19	13.84	3.184 (faible dispersion)	P = 0.163 (Test Kruskall-Wallis, non significatif)
		Guidé académique	12	13.66	3.750 (faible dispersion)	
		Guidé non académique	4	17.25	3.096 (faible dispersion)	

Variables liées aux conditions d' usage	**Variable statut socioprofessionnel**	Académique	23	13.17	3.353 (faible dispersion)	**P = 0.016 (Test U de Mann-Whitney, significatif)**
		Professionnel	12	16.08	2.937 (faible dispersion)	
	Variable du lieu de résidence	France	16	13.75	3.416 (faible dispersion)	P = 0.567 (Test U de Mann-Whitney, non significatif)
		Japon	19	14.53	3.565 (faible dispersion)	
	Variable de la durée du séjour en France	Moins d'un an	4	11.25	1.258 (faible dispersion)	**P = 0.010 (Test Kruskall-Wallis, significatif)**
		Entre 1 et 3 ans	10	13.30	3.129 (faible dispersion)	
		Entre 4 et 7 ans	11	15.27	2.760 (faible dispersion)	
		8 ans et plus	10	15.00	4.447 (faible dispersion)	
	Variable du type de contexte de conversation en français	Académique	8	13.75	5.435 (faible dispersion)	P = 0.714 (Test Kruskall-Wallis, non significatif)
		Personnel	19	14.68	3.128 (faible dispersion)	
		Professionnel	8	13.38	5.069 (dispersion moyenne)	
	Variable de la fréquence de la conversation en français	Plusieurs fois par mois et moins	10	13.78	3.143 (faible dispersion)	P = 0.958 (Test Kruskall-Wallis, non significatif)
		Plusieurs fois par semaine	9	14.10	4.295 (faible dispersion)	
		Plusieurs fois par jour	16	14.44	3.366 (faible dispersion)	
	Variable du type d'interlocuteur en français	Familiers / proches	10	14.70	3.199 (faible dispersion)	P = 0.064 (Test Kruskall-Wallis, non significatif)
		Pro / académique	16	15.06	3.202 (faible dispersion)	
		Locuteur lambda	9	12.00	3.435 (faible dispersion)	

Tableau 12 : Relevé statistique pour chaque variable étudiée sur les résultats au test des LNN

2.1. Variables liées aux conditions d'apprentissage

Les variables liées aux conditions d'apprentissage regroupent toutes les variables relatives aux différentes modalités caractérisant les conditions de l'apprentissage du français chez les LNN observés dans notre étude. Ces conditions d'apprentissage ne sont pas forcément synchroniques de la passation du test. Par cela, nous entendons que les informations relatives aux conditions d'apprentissage concernent à la fois des locuteurs qui continuent de suivre un apprentissage régulier et des locuteurs qui ont effectué cet apprentissage par le passé. En effet, certains de nos informateurs ne sont plus dans le milieu académique, et de fait, ne suivent plus un apprentissage suivi. Ils ont donc fourni des informations relatives à l'époque où ils étaient étudiants. Nous souhaitons observer s'il existe un lien statiquement significatif entre les taux de réussite enregistrés et les conditions d'apprentissage du français des informateurs. Cette démarche nous permettra d'affiner notre analyse lors de l'exploration qualitative des données en apportant des éléments contextuels sur les informateurs analysés.

2.1.1. Variable niveau d'étude

La première variable du niveau d'étude n'a pas été validée statistiquement (P = 0.299 pour un seuil de significativité à 0.05) et ne semble donc pas évoquer une influence directe du niveau d'étude des LNN sur leur taux de réussite au test lexicopragmatique. Les moyennes par groupe évoquent malgré tout une différence assez importante entre le groupe de licence (moy. = 15.38) et le groupe de doctorat (moy. = 12.82). L'écart se creuse un peu moins entre le groupe de master (moy. = 14.30) et les deux autres niveaux d'étude. Les résultats indiquent une distribution inverse à celle de notre hypothèse de départ[155] pour cette variable : Licence > Master > Doctorat. L'hypothèse selon laquelle un haut niveau d'étude dans un cursus lié à la langue française favoriserait une compétence lexicopragmatique développée ne semble donc pas vérifiable dans notre jeu de données.

[155] Le niveau de compétence en français augmente à mesure que le niveau d'étude lui-même augmente.

2.1.2. Variable de la localisation de l'université

La variable du lieu d'université n'a pas été validée statistiquement (P = 0.582 pour un seuil de significativité à 0.05), le fait d'étudier ou non en France ne semble donc pas avoir d'influence significative sur les résultats obtenus au test par les LNN. On observe que la moyenne enregistrée pour les étudiants diplômés d'une université japonaise (moy. = 14.47) est légèrement supérieure à ceux diplômés en France (moy. = 13.81) ce qui semble montrer une tendance inverse à l'hypothèse de départ selon laquelle une formation académique en France peut être un facteur d'amélioration de la compétence lexicopragmatique des LNN.

2.1.3. Variable de la spécialité d'étude

La variable de la spécialité d'étude n'a pas donné de résultats significatifs sur le plan statistique (P = 0.325 pour un seuil de significativité à 0.05). Au niveau des moyennes enregistrées, on observe la distribution suivante : Études françaises/Littérature (moy. = 15.21) > SHS (moy. = 13.33) > ScL (moy. = 13.13). Cette distribution indique ici une tendance plus ou moins cohérente avec notre hypothèse de départ avec des résultats élevés enregistrés du côté des disciplines littéraires ou liées à l'apprentissage du français. En revanche, en dépit du lien étroit qu'entretiennent les disciplines des sciences du langage avec la compréhension de la structure linguistique du français [156], nous observons avec étonnement que les résultats les plus faibles de cette variable se retrouvent dans cette catégorie.

2.1.4. Variable du niveau de certification

La variable du niveau CECRL n'a pas été validée sur le plan statistique (P = 0.318 pour un seuil de significativité à 0.05) et n'est donc pas généralisable. On observe toutefois une distribution des taux de réussite au test qui semble aller dans le sens de l'hypothèse d'une influence du niveau de certification sur la compétence lexicopragmatique : C2 > B2 > C1. Le niveau C2 (moy. = 15.26) affiche la moyenne de score la plus élevée ce qui semble évoquer, tout du moins pour la cohorte étudiée, une possible augmentation des résultats liée

---

[156] Nous précisons ici que tous les informateurs déclarés comme étant spécialistes de sciences du langage travaillent sur la langue française dans le cadre de leur recherche.

au niveau d'expertise. On observe des résultats similaires entre le niveau B2 (moy. = 13.42) et C1 (moy. = 13.30), ce qui ne semble pas étonnant étant donné qu'il s'agit ici du palier de transition entre la compétence intermédiaire et la compétence avancée.

2.1.5. Variable de la durée d'apprentissage

La variable de la durée d'apprentissage du français a été validée sur le plan statistique avec une valeur de P égal au seuil de significativité (P = 0.051) pour un seuil de significativité à 0.05). Cette significativité statistique nous permet de postuler l'hypothèse d'un possible effet de la durée de l'apprentissage sur le score de réussite au test dans notre cohorte. La distribution des résultats est la suivante : 15 ans et plus > 1 à 5 ans > 6 à 10 ans >11 à 14 ans. On remarque que les moyennes des scores de réussite restent stables de 1 à 14 ans d'apprentissage (moy. =13.14 ~ 13.42), puis on observe un pic significatif dans les résultats à partir de 15 ans et plus (moy. = 16.77). En observant la grille des résultats individuels au test, on remarque que la totalité des LNN ayant commencé un apprentissage précoce dès l'enfance enregistre des scores de réussite proches de ceux des locuteurs natifs. La longueur d'apprentissage du français semble apparaitre comme un paramètre pertinent dans la caractérisation du niveau d'expertise des LNN (sans pour autant être le seul ou le principal).

2.1.6. Variable du cadre d'apprentissage

La variable du cadre d'apprentissage n'a pas été validée sur le plan statistique (P = 0.190 pour un seuil de significativité à 0.05). La distribution des moyennes est la suivante : Guidé non académique > Autonome > guidé académique. L'apprentissage en milieu guidé non académique (moy. = 17.25), c'est-à-dire les cours privés en présentiel ou en ligne, les écoles de langue, les instituts français et autre cadre d'apprentissage hors centre académique, semble se détacher des autres cadres d'apprentissage avec une moyenne 3 point supérieure à celles enregistrées pour les deux autres cadres. Le cadre guidé académique (moy. = 13.66), qui englobe l'apprentissage en université, dans un cursus FLE ou dans le cadre d'une LV2 ou LV3 dans un collège ou lycée, a donné des résultats très proches de ceux retrouvés pour le cadre d'apprentissage en autonomie (moy. = 13.84) qui concerne le cas d'un apprentissage seul en immersion ou un apprentissage en autodidacte. Il nous semble important de noter ici que les locuteurs ayant déclaré un apprentissage en autodidacte ont aussi déclaré avoir effectué un cursus universitaire liée à la langue française, ce qui implique *de facto* un

apprentissage guidé académique. Ces résultats semblent donc à prendre avec une certaine précaution au vu de la complexité du parcours de locuteur de chacun des LNN ayant participé à cette étude.

Après avoir observé les résultats statistiques obtenus à partir de variables liées aux conditions d'apprentissage des LNN, nous allons dans la section qui suit évoquer les résultats concernant les variables liées aux conditions d'usage du français.

2.2. Variables liées à l'usage

Les variables liées à l'usage regroupent toutes les variables relatives aux conditions d'emplois et aux enjeux stratégiques et sociaux associés à l'usage du français chez les LNN. L'objectif de cette seconde partie de notre travail statistique est de déterminer s'il existe un lien statistiquement significatif entre les résultats obtenus au test et les différentes modalités liées à l'emploi du français chez les LNN. Nous souhaitons ainsi déterminer à l'instar de la partie précédente s'il existe des facteurs susceptibles de favoriser la compétence lexicopragmatique.

2.2.1. Variable du statut socioprofessionnel

La première variable étudiée est celle du statut socioprofessionnel. Le test U de Mann-Whitney indique une significativité de la relation entre la variable et le score de réussite au test (P = 0.016 pour un seuil de significativité à 0.05). Nous observons ainsi que le statut socioprofessionnel et par extension le fait d'être exposé ou non à un milieu impliquant une activité académique ou professionnelle peut influer sur les scores enregistrés. La distribution des résultats est la suivante : statut professionnel (moy. = 16.08) > statut académique (moy. = 13.17). Les résultats semblent indiquer que les informateurs ayant un statut académique (étudiant, chercheur, enseignant de langue) ont un taux de réussite plus faible que les informateurs insérés dans une entreprise. Une première interprétation possible est que le cadre académique impliquant un usage plus standard et normé, le locuteur est *de facto* moins exposé aux variétés familières du français oral. À l'inverse, les informateurs évoluant dans un milieu professionnel standard ou transactionnel sont plus souvent confrontés à des situations de communication authentiques avec des interactions souvent informelles avec les

collègues, ou très diversifiées (négociation, explication, plaisanteries, requêtes, refus, etc.) avec les clients. Une autre hypothèse[157] possible serait que les milieux académiques semblent accueillir un nombre de locuteurs non natifs plus important, diminuant ainsi les occasions pour les locuteurs japonophones évoluant dans ce milieu de pouvoir interagir avec des locuteurs natifs. Cela semble d'autant plus plausible qu'il est fréquent d'observer des phénomènes d'inertie dans le milieu académique au cours desquels les locuteurs d'une même langue maternelle ou ayant une langue tierce en commun ont tendance à se regrouper de manière préférentielle, au détriment d'une interaction avec des LN de la langue cible. En revanche, les milieux professionnels semblent plus propices à l'interaction avec des locuteurs natifs qui ne partagent bien souvent pas la langue maternelle du locuteur japonophone, ce qui pourrait favoriser l'émergence d'une compétence lexicopragmatique permettant l'emploi d'un lexique plus naturel et adapté aux enjeux de la communication quotidienne. De même, en observant le tableau des métadonnées, on s'aperçoit que les informateurs qui travaillent en France sont bien plus souvent en couple avec un/une locuteur.rice français.e natif.ve que les informateurs en milieu académique (souvent doctorants ou lecteur.rices en langue), ces premiers semblent donc plus exposé à la communication quotidienne dans un cadre familial francophone.

2.2.2. Variable du lieu de résidence

La variable du lieu de résidence n'a pas été validée par le test de significativité (P = 0.518 pour un seuil de significativité à 0.05). Le fait de résider en France ou au Japon ne semble donc pas avoir d'effet significatif sur les résultats obtenus au test lexicopragmatique. La distribution des moyennes de score indique la distribution suivante : résident Japon (moy. = 14.53) > résident France (moy. = 13.75). Ce faisant, l'hypothèse selon laquelle l'immersion dans le contexte authentique de la langue cible favoriserait l'acquisition d'une certaine aisance dans la conversation quotidienne, et donc l'émergence d'une compétence pragmatique et lexicopragmatique étendue n'est pas vérifiable dans notre jeu de données.

---

[157] Hypothèse soulevée par Agnès Tutin lors d'une réunion de recherche, nous l'en remercions.

2.2.3. Variable de la durée du séjour

La variable de la durée du séjour a été validée statistiquement (P = 0.010 pour un seuil de significativité à 0.05). Le test Kruskall-Wallis employé pour tester la significativité de la variable a montré que les résultats obtenus n'étaient pas dus au hasard et traduisaient bien un effet de la variable sur les scores de réussite. La distribution des résultats est la suivante : Entre 4 et 7 ans > 8 ans et plus > Entre 1 et 3 ans > moins d'un an. Cette distribution décrit une trajectoire croissante cohérente avec l'hypothèse selon laquelle un séjour prolongé en France peut favoriser l'émergence d'une compétence lexicopragmatique avancée. On observe une stabilisation des résultats à partir de 4 ans de résidence en France. On pourrait émettre l'hypothèse que les progrès deviennent de plus en plus limités au fur et à mesure que la compétence s'accroit et qu'à partir d'un certain niveau d'expertise, l'acquisition porte sur des pratiques linguistiques plus complexes qui demandent un temps d'acquisition plus important. En outre, nous précisons ici que cette immersion seule ne suffit pas et doit être associée au facteur motivation à interagir dans la langue cible.

2.2.4. Variable du contexte principal de conversation

La variable du contexte principal de conversation en français n'a pas d'influence significative dans notre jeu de donnée (P = 0.647 pour un seuil de significativité à 0.05). Il ne semble donc pas exister d'effet du contexte principal d'interaction en français sur les scores obtenus au test. La distribution des résultats est la suivante : Personnel (moy. = 14.68) > Académique (P = 0.647 pour un seuil minimal de significativité à 0.07)> Professionnel (moy. = 13.38). L'inventaire des moyennes obtenues au test pour chaque groupe montre un meilleur taux de réussite chez les informateurs utilisant le français fréquemment en contexte personnel (famille, amis). Les valeurs associées aux contextes en entreprise et académiques sont sensiblement au même niveau. Nous observons toutefois dans la grille de métadonnées que certains informateurs ayant déclaré un statut académique ont déclaré le contexte professionnel comme contexte principal d'usage du français. Il s'agit notamment de quelques doctorants qui travaillent aussi pour l'université comme vacataire et qui sont donc amenés à interagir avec les autres membres du personnel en tant qu'employés et non pas en tant qu'étudiants.

2.2.5. Variable de la fréquence d'interaction

La variable de la fréquence d'interaction en français n'a pas donné de résultats concluants sur le plan de la significativité (P = 0.919 pour un seuil de significativité à 0.05). Notre jeu de donnée ne nous permet donc pas d'attester d'un effet de la fréquence à laquelle les informateurs interagissent avec des locuteurs natifs du français sur les résultats obtenus au test. La distribution des moyennes obtenues est la suivante : Plusieurs fois par jour (moy. = 14.44) > Plusieurs fois par semaine (moy. = 14.10) > Plusieurs fois par mois et moins (moy. = 13.78). Cette distribution semble indiquer une trajectoire régulière avec une augmentation du score en fonction de la fréquence d'interaction des informateurs avec des locuteurs français en langue française. Cette tendance retrouvée dans notre groupe d'informateur ne peut être généralisée du point de vue statistique, mais montre tout de même une certaine régularité au niveau intragroupe dans l'augmentation des résultats en fonction de la fréquence d'interaction.

2.2.6. Variable du type d'interlocuteur

La dernière variable étudiée est celle du type de LN du français avec lesquels les informateurs ont l'habitude d'interagir fréquemment. Le test statistique indique une non-significativité (P = 0.097 pour un seuil de significativité à 0.05). Nous notons néanmoins qu'à la différence des autres résultats peu significatifs, le test Kruskall-Wallis affiche ici un résultat proche du seuil minimum de significativité. Les moyennes par groupe se distribuent de la manière suivante : Locuteurs professionnels (moy. = 15.06) / académiques > Locuteurs familiers / proches (moy. = 14.70) > Locuteurs lambdas (moy. = 12.00). Il nous semble ici un peu prématuré de postuler l'hypothèse d'un éventuel effet de cette variable sur les résultats obtenus au test, car certaines inconnues restent à vérifier. En effet, même si un informateur déclare interagir fréquemment avec des locuteurs familiers, rien ne nous indique qu'il n'interagit pas non plus avec des locuteurs du milieu professionnel ou académique. De même, rien ne nous indique que la compétence lexicopragmatique s'est développée principalement au contact du type principal d'interlocuteur et non pas au contact d'interlocuteurs moins fréquents. Il nous semble donc important de prendre certaines précautions dans la prise en compte de cette dernière variable.

3. Synthèse de l'analyse quantitative et statistique des données

L'analyse statistique descriptive des résultats obtenus par les LN et les LNN au test lexicopragmatique montre globalement un écart assez important entre les taux de réussite des deux groupes. Cet examen comparatif des résultats entre le groupe test et le groupe de référence semble aller dans le sens d'une différence de compétence notable, et ce malgré le niveau C2 d'une grande partie des LNN interrogés (15/35 informateurs). L'observation des scores obtenus par individu montre une grande dispersion résultats autour de la moyenne chez les LNN alors que les résultats des natifs sont très homogènes et centrés autour du taux moyen de réussite de la cohorte. Au vu des scores de réussite obtenus par les LNN, il semble pertinent de penser que le niveau C2 n'implique pas forcément une maitrise étendue des expressions idiomatiques et des stratégies discursives spécifiques à la communication comme semble le suggérer le tableau de compétence interactionnelle ou la section dédiée à la compétence pragmatique publiés dans le volume complémentaire du CECRL (2018 : 86, 138) :

INTERACTION ORALE	
INTERACTION ORALE GÉNÉRALE	PROSIGN
C2	Possède une bonne maîtrise d'expressions idiomatiques et de tournures courantes avec une conscience du sens connotatif. Peut exprimer avec précision des nuances fines de signification, en utilisant assez correctement une gamme étendue de modalités. Peut revenir sur une difficulté et la restructurer de manière si habile que l'interlocuteur s'en rende à peine compte.
C1	Peut s'exprimer avec aisance et spontanéité, presque sans effort. Possède une bonne maîtrise d'un vaste répertoire lexical lui permettant de surmonter facilement des lacunes par des périphrases, avec apparemment peu de recherche d'expressions ou de stratégies d'évitement. Seul un sujet conceptuellement difficile est susceptible de gêner le flot naturel et fluide du discours.
B2	Peut utiliser la langue avec aisance, correction et efficacité dans une gamme étendue de sujets d'ordre général, académique, professionnel et concernant les loisirs, en indiquant clairement les relations entre les idées. Peut communiquer spontanément avec un bon contrôle grammatical sans donner l'impression d'avoir à restreindre ce qu'il/elle souhaite dire et avec le degré de formalisme adapté à la circonstance.
	Peut communiquer avec un niveau d'aisance et de spontanéité tel qu'une interaction soutenue avec des locuteurs de la langue cible soit tout à fait possible sans entraîner de tension d'une part ni de l'autre. Peut mettre en valeur la signification personnelle de faits et d'expériences, exposer ses opinions et les défendre avec pertinence en fournissant explications et arguments.

Figure 53 : Tableau de synthèse des compétences liées à l'interaction orale issu du volume complémentaire du CECRL (2018)

Nous verrons toutefois dans l'analyse qualitative que ces résultats statistiques restent à nuancer. Les scores obtenus par les non-natifs, bien que relativement mitigés, n'indiquent en rien que la compétence pragmatique en compréhension est proportionnellement faible.

En effet, les questions d'autoévaluations systématiquement proposées après la réponse à chaque item semblent indiquer que les LNN observés comprennent dans la plupart des cas les différents éléments caractérisant la situation de communication, ont une idée du type d'expression à utiliser au moment T, mais ne disposent pas tout le temps du matériel lexical approprié (en compréhension ou en production) pour accéder à la fonction pragmatique visée :

	PhPex traité	**Type de protocole**	**Compétence observée**	**Score de réussite**	**Score de comp.**
Activité 1	C'est + [réponse au choix] Sérieux ? + [réponse au choix]	Complétion de dialogue	Production -Pragmalinguistique -Sociopragmatique	76%	90%
Activité 2	C'est le pied ! C'est pas possible ! (distracteur) Faut pas pousser ! C'est abusé ! C'est pas possible ! J'y suis pour rien !	Reformulation de PhPex	Production -Pragmalinguistique -Sociopragmatique	65%	88%
Activité 3	J'y crois pas ! C'est pas vrai ? C'est chaud ! C'est pas gagné ! Tu plaisantes ?	Production d'équivalence de PhPex	Production -Pragmalinguistique -Sociopragmatique -Métapragmatique	62%	88%
Activité 4	Choix entre : [c'est génial ! /c'est le pied ! /c'est le bouquet ! / c'est une horreur !] Choix entre : [c'est génial !/ c'est pas gagné ! / ça tombe bien ! / c'est une horreur !] Choix entre : [Tu crains ! / Tu parles ! / T'inquiète ! / Tu l'as dit !]	QCM lexical	Compréhension -Sociopragmatique -Métapragmatique	41%	73%
Activité 5	Tu es sérieux ? C'est pas possible (distracteur) J'y crois pas ! Les bras m'en tombent ! C'est une honte ! C'est le pied. (distracteur)	QCM sémantique	Compréhension -Sociopragmatique -Métapragmatique	73%	89%

Tableau 13 : Tableau de comparaison entre les modalités propres à chaque activité, le taux de réussite et le taux de compréhension.

Dans le tableau comparatif ci-dessus, nous observons une mise en relation des modalités de chaque activité avec les taux de réussite et de compréhension. Nous pouvons remarquer que les taux de compréhension demeurent relativement élevés malgré le taux de réussite faible de l'activité 3 ou 4 par exemple. L'analyse quantitative ne nous permet pas de déterminer les facteurs à l'origine de ces différences entre compréhension et production. En revanche,

elle nous a permis de constater cette tendance qui aurait été invisible dans un travail purement qualitatif.

L'analyse quantitative nous a permis de constater que certains items ont suscité des scores de réussite faibles dans les deux cohortes. Ces scores faibles sont le résultat de deux facteurs : la forte idiomaticité et polyfonctionnalité des expressions et une difficulté liée à la complexité de la tâche demandée. Nous tenterons de déterminer plus précisément dans l'analyse qualitative proposée dans le prochain chapitre les difficultés à l'origine de ces baisses dans le taux de réussite. Dans la partie consacrée à l'étude des variables, nous avons pu constater à l'aide de tests statistiques paramétriques (Test U de Mann-Whitney et Test Kruskall-Wallis) que finalement peu de résultats étaient statistiquement significatifs si ce n'est la variable concernant la durée d'apprentissage du français, celle du statut socioprofessionnel dans lequel évoluent les informateurs et celle de la durée du séjour en France. Cette faible significativité des mesures statistiques mises en œuvre peut également être résultat d'un effectif trop petit et d'une trop forte dispersion des résultats. De manière générale, nous n'avons pu constater que certaines tendances propres aux spécificités du groupe de LNN étudié :

		Significativité	
		Oui	**Non**
Variables liées aux conditions d' apprentissage	**Variable niveau d'étude**		X
	Variable lieu d'université		X
	Variable spécialité d'étude		X
	Variable niveau certification CECRL		X
	Variable durée d'apprentissage du français	X	
	Variable cadre d'apprentissage du français		X
Variables liées aux conditions d' usage	**Variable statut socioprofessionnel**	X	
	Variable du lieu de résidence		X
	Variable de la durée du séjour en France	X	
	Variable du type de contexte de conversation en français		X
	Variable de la fréquence de la conversation en français		X
	Variable du type d'interlocuteur en français		X

Tableau 14 : Synthèse de la significativité des variables observées chez les LNN

Par ailleurs, nous avons remarqué que les variables qui semblaient à première vue être conventionnellement perçues comme des facteurs d'amélioration de la compétence pragmatique (et de la compétence linguistique en général) ne produisaient finalement pas d'effets significatifs dans notre cohorte. Le niveau ou la spécialité d'étude par exemple n'ont pas semblé avoir d'influence particulière sur le taux de réussite au test. De surcroit, certains des informateurs qui sont doctorants en sciences du langage ont produit de moins bons scores de réussite que des informateurs titulaires d'une licence d'études françaises. À priori, on pourrait penser que le fait d'avoir un très haut niveau d'étude et une expertise en linguistique française serait un avantage certain pour analyser, comprendre et employer plus facilement les phraséologismes pragmatiques spécifiques à la conversation en langue cible. Pourtant, les résultats de notre étude montrent que cette variable n'entre pas spécifiquement en jeu dans le cas de notre groupe test et que l'acquisition de la compétence lexicopragmatique nécessaire à la compréhension et l'usage des PhPex semble dépendre d'autres facteurs qui dépassent la connaissance du système linguistique ou du standard. Il en va de même pour la variable du niveau CECRL qui n'a pas montré de lien spécifique entre un niveau certifié avancé en français et une meilleure maitrise de l'emploi des PhPex de notre test lexicopragmatique.

4. Discussion des résultats quantitatifs et statistiques

Bien que le champ exploré par notre étude soit relativement novateur, nous allons ici tenter de discuter nos résultats à la lumière d'études acquisitionnelles présentant des traits communs avec notre recherche. Nous avons ainsi mené un rapide examen d'études statistiques en acquisition L2 menées sur d'autres types de LNN, apprenant l'anglais L2 pour la plupart, de manière à relever certaines convergences et divergences avec nos résultats. Parmi les variables les plus explorées dans ces études et qui pourraient se recouper avec nos résultats : la longueur du séjour dans le pays de la langue cible, l'expertise du LNN en langue cible, l'intensité à laquelle il communique avec des LN et la diversité des situations de communication et des types de locuteurs auxquels il s'expose.

4.1. Effet de la longueur d'exposition sur la compétence lexicopragmatique

Notre étude statistique a montré que la longueur avait un effet positif sur les résultats obtenus par les LNN de notre cohorte au test lexicopragmatique (P = 0.010, significatif) avec un pic après 4 ans d'exposition. Les LNN observés ont donc plus de chance de développer leur compétence lexicopragmatique en s'exposant de manière prolongée à la LC[158]. Une revue d'autres études concernant cette variable a montré que dans la majorité des cas, la longueur du séjour avait un impact positif sur le développement de la compétence pragmatique, corroborant ainsi nos résultats.

Ainsi, Kecskés (2000, 2003) a observé qu'une immersion d'au moins deux ans dans la communauté linguistique cible permettait à un LNN de développer un usage des expressions idiomatiques présentant certaines similarités avec l'emploi que font les LN. Roever (2005) montre dans son étude à grande échelle (267 informateurs) que la compréhension des expressions idiomatiques et des routines situationnelles semblait plus importante chez les informateurs ayant séjourné dans le pays cible sur une longue durée plutôt que sur les informateurs dotés d'une certification élevée. Concernant ce dernier point, nos résultats semblent aussi montrer que l'expertise en langue cible n'était pas directement liée au développement de la compétence lexicopragmatique.

Plus récemment, Alcón-Soler & Sánchez-Hernández (2017) dans leur étude sur les routines pragmatiques ont montré que la progression de la compétence était plus significative au niveau de la compréhension que de la production, ce que nous avons pu également constater dans notre étude lors de l'observation des taux de compréhension par type d'activité. Enfin, Liu et Yang (2020) montrent que l'exposition à la langue cible sur une durée prolongée et avec une intensité modérée avait un effet sur l'augmentation de l'emploi des structures phraséologiques en production sur des locuteurs de niveau avancé.

Le lien entre durée d'exposition et développement de la compétence lexicopragmatique semble donc avoir été observé de manière fréquente dans le champ de la pragmatique L2. Toutefois, nous faisons deux constatations. La première est que la très grande majorité de ces études ont été menées a) dans le cadre d'un échange académique ou sur des locuteurs engagés dans un processus d'apprentissage guidé, b) principalement sur l'anglais. Notre étude tente une approche différente en se focalisant sur le couple de langue français-japonais

---

[158] Pour rappel, LC = langue cible.

pour lequel il n'existe à notre connaissance aucune étude de ce type. Elle se fonde également sur une population de LNN diversifiée en type d'exposition (étudiant, employés, conjoints de couple internationaux, chercheurs) et de différentes générations (de la vingtaine à la soixantaine).

4.2. Effet de l'expertise en langue cible sur la compétence lexicopragmatique

L'étude des variables liées à l'expertise en langue cible (certification, durée d'apprentissage du français, diplôme obtenu en France, cursus lié à la langue-culture française) n'a pas donné de résultat significatif dans notre étude, à l'exception de la variable liée à la durée d'apprentissage du français (P= 0.051). Une revue des principales études concernant ces variables a montré deux tendances générales concernant le niveau d'expertise, dont la deuxième qui corrobore nos résultats. La première tendance concerne des études montrant que le niveau de maitrise linguistique avait une influence directe sur le développement de la compétence pragmatique (Trosborg 1987 ; Yorio 1989 dans Bardovi-Harlig & Bastos 2011). La deuxième tendance réfère à des études ne montrant à l'inverse aucun effet direct de l'expertise sur la compétence pragmatique (Matsumura 2003 ; Schmitt, Dörnyei, Adolphs et Durow 2004). Ces études montrent en revanche un effet indirect, car l'expertise dans la LC peut en soi favoriser les occasions d'interagir avec des LN et permettre ainsi au LNN de développer sa palette lexicopragmatique.

Roever (2005) a montré que l'expertise influençait grandement la compétence pragmatique en langue cible lors de la production d'actes de langage fréquemment enseignés, mais ne semblait pas influencer significativement la reconnaissance en contexte des routines pragmatiques. Les résultats de l'étude de Roever sont d'autant plus intéressants que nous avons pu remarquer dans notre étude que les principales difficultés des LNN portaient principalement sur les phraséologismes peu standards et peu enseignés (par exemple, *c'est le pied !* ou *c'est chaud !*). Taguchi (2005, 2008) observe que le niveau d'expertise en L2 a une influence sur la capacité du LNN à inférer le sens d'une expression en fonction du contexte, mais pas sur la capacité de production de ces expressions. Bardovi-Harlig (2008, 2009) corrobore les résultats de Taguchi (2005, 2008) en constatant que l'expertise en L2 avait un impact favorable sur la compétence pragmatique, mais que le développement de la compréhension était plus significatif que celui de la compétence de production. À l'instar de ces études antérieures, nous remarquons cette différence de l'aisance en compréhension par

rapport à la production chez nos informateurs. Cela est notamment observable par la comparaison des taux de compréhension au taux de réussite, mais également grâce à l'indice phraséologique évoqué plus haut. Plus récemment, Bardovi-Harlig & Bastos (2011) montrent que le lien entre niveau d'expertise en langue cible et fréquence d'usage des expressions conventionnelles était significatif ($p < 0.001$) et observent une augmentation du nombre de formes phraséologiques à mesure que le niveau d'expertise augmente.

4.3. Effet de l'intensité d'exposition à la langue cible sur la compétence lexicopragmatique

L'étude de la variable liée à l'intensité d'exposition à la langue cible, c'est-à-dire la fréquence à laquelle le LNN est amené à communiquer avec un LN n'a pas donné de résultat significatif ($P = 0.958$, non significatif). Toutefois, nous avons pu observer lors de notre passage en revue des études antérieures sur cette variable que la plupart des résultats retrouvés ne corroboraient pas nos résultats et tendaient vers une forte corrélation entre fréquence d'interaction avec des LN de la LC en milieu écologique et développement de l'emploi et de la compréhension des routines pragmatiques[159]. Certaines études observent même que l'intensité semble jouer un rôle plus décisif que la longueur de l'exposition (Bardovi-Harlig & Bastos, 2011).

Dewaele et Regan (2001) explorent le lien entre développement du lexique pragmatique et exposition à différents types d'*input* parmi lesquels la fréquence d'interaction avec les LN, la fréquence de lecture dans la langue cible ou la durée d'apprentissage formel de la langue cible durant le séjour à l'étranger. L'étude montre que seule la fréquence d'interaction avait un effet significatif sur le développement de la compétence en langue cible. Kinginger et Blattner (2008) observent qu'une immersion soutenue et diversifiée en interlocuteurs et en situation de communication avait un impact significatif sur la maitrise des routines pragmatiques. Dans leur étude, Bardovi-Harlig & Bastos (2011) observent que l'intensité d'interaction à une forte incidence sur la compétence de reconnaissance des expressions conventionnelle.

---

[159] Une grande partie des études recensée mentionnaient le plus souvent la notion de *pragmatic routines* comme marques de l'oral ou énoncés ritualisés/conventionnels, mais peu d'entre-elles se sont intéressées aux expressions idiomatiques de l'oral. Il n'existe pas à notre connaissance d'étude portant spécifiquement sur l'acquisition des phraséologismes pragmatiques de la conversation quotidienne.

Encore une fois, ces études sont relativement peu diversifiées et se concentrent essentiellement sur le cas des étudiants en échange dans le pays de la LC. Il nous a semblé que dans ce cas-là, les étudiants qui sont fortement motivés dans leur processus d'acquisition et engagé dans un parcours d'apprentissage encadré. Ce n'est pas le cas pour tous les informateurs dans nos données où nous avons des locuteurs qui ne sont pas tous étudiants et n'ont pas le même niveau d'implication dans l'acquisition du français. Cette différence d'implication peut selon nous affecter la fréquence et la diversité avec laquelle les LNN vont interagir avec les LN de la LC. En ce sens, notre étude prend tout son intérêt puisqu'elle permet d'élargir le champ d'analyse à des profils des locuteurs diversifiés, notamment des profils dont les besoins en communication sont principalement liés à la conversation quotidienne et d'autre à une activité essentiellement académique ou professionnelle.

4.4. Effet de la diversité d'exposition sur la compétence lexicopragmatique

L'étude des variables liées à la diversité d'exposition à la langue cible (contexte de communication et type de locuteur) n'a pas donné de résultat significatif, que ce soit pour la variable du type de contexte d'interaction (P= 0.714, non significatif) ou la variable du type d'interlocuteur (P=0.064, non significatif). Ces résultats ne sont pas corroborés par les études antérieures qui montrent pour la plupart un effet significatif de cette variable sur la maitrise de la compétence pragmatique. Le rôle du type de LN avec lequel le LNN interagit est déterminant pour l'apprentissage, ainsi que l'importance de créer et de multiplier des opportunités de converser avec différents types d'interlocuteurs (Adolphs & Durow 2004 ; Dörnyei et al. 2004 ; Kasper and Rose 2002 ; Shea 1994). Kinginger & Farrell (2004) vont ainsi observer que les LNN qui s'engagent dans plus d'interaction avec des LN d'âge et de contextes sociaux variés jouissent d'une progression de leur compétence pragmatique plus accrue. Enfin, Taguchi (2008) montre que le temps d'exposition hors murs de la classe avec notamment une exposition à l'interaction avec des LN en langue cible combinée avec la consommation de médias culturels en langue cible avait un impact significatif sur la compétence pragmatique en compréhension.

Un des reproches que l'on peut faire aux différentes études observées jusqu'à présent est qu'elles interprètent souvent le développement de la compétence pragmatique sous l'angle d'un ou deux paramètres tout au plus. Cette focalisation peut donner l'impression que le

développement de cette compétence est associé à un ensemble des variables simple, prenant le risque d'occulter des phénomènes plus profonds et moins visibles. Or il nous a semblé qu'au contraire, ce développement est dû à la congruence et la complémentarité de divers facteurs à divers niveaux de l'acquisition. C'est l'articulation entre ces différents facteurs et non pas le fait de favoriser un facteur de manière isolée qui joue un rôle crucial dans l'acquisition des unités lexicopragmatiques. Nous avions remarqué par exemple dans l'étude diagnostique que l'informateur japonophone du corpus de Tsukuba avait un état de développement de la compétence lexicopragmatique supérieur au niveau obtenu lors de sa certification linguistique. Cela était notamment dû à son engagement fort dans la conversation et à sa motivation à expérimenter de nouveau acquis linguistiques en LC.

5. Remarques finales

Comme nous l'avons mentionné auparavant, la plupart des études statistiques recensées précédemment se limitent essentiellement à l'étude du cas des étudiants en échange universitaire, mais peu d'études semblent se focaliser sur le développement de la compétence chez des personnes ayant vécu ou vivant depuis plus de 4 ans dans le pays cible hors du cadre de l'échange scolaire. En outre, notre étude statistique des résultats du test lexicopragmatique montre que le développement de la compétence lexicopragmatique en L2 n'est pas linéaire et est le fruit de l'influence de nombreux paramètres (Taguchi 2015). Les conditions propices au développement de la compétence pragmatique ne semblent pas encore clairement établies, et de nombreux paramètres complexes à mesurer restent à tester (Kasper & Rose 2002 : 232). La longueur du séjour est une variable complexe : bien que facile à mesurer, elle est complexe dans la mesure où la qualité de ce séjour et l'expérience vécue diffèrent pour chaque locuteur. Ces différences d'expérience durant ce séjour ont une influence sur la qualité du développement de la compétence pragmatique (Dörnyei *et al.* 2004, Kinginger et Blattner 2004). De même, les variables de l'intensité et la diversité de l'input semblent être des pistes d'analyse quantitative intéressantes. Toutefois elles nécessitent, semble-t-il, d'être confronté à une analyse systématique et qualitative pour chaque informateur auprès desquels il apparait nécessaire de collecter des données sociobiographiques minutieuses recensant le type de locuteur, de situation d'interaction et la fréquence de ces interactions, choses que nous n'avons pas pu faire dans le cadre de notre recherche en raison de notre protocole de recueil à distance.

Une majorité des études citées précédemment montrent par ailleurs qu'un niveau de compétence élevé en langue permet en général une meilleure acquisition des structures pragmatiques, mais ne garantit pas des emplois similaires aux natifs. Nous retrouvons ici le problème évoqué dans la partie méthodologie de cette deuxième étape de notre étude, à savoir l'illusion comparative (*Comparative fallacy,* Bley-Vroman 1983). En effet, un certain nombre d'études partent du présupposé qu'une compétence développée est nécessairement une copie de la pratique du LN. Or, ce critère d'évaluation de la production du LNN nous apparait peu pertinent pour évaluer la compétence lexicopragmatique. L'objectif principal du locuteur n'est pas de s'exprimer comme les interlocuteurs natifs, mais de pouvoir s'exprimer de manière fluide, nuancée sans pour autant subir une tension cognitive particulière. Il apparait donc judicieux de se garder de définir systématiquement le niveau de développement de la compétence pragmatique en fonction de la fréquence et de la nature des emplois similaires aux LN, mais plutôt en fonction des objectifs pragmatiques atteints et de la pertinence du matériel linguistique mobilisé pour réaliser ces objectifs pragmatiques. En outre, une grande partie de ces études se focalisent sur la production, mais moins sur la compréhension qui est pourtant une compétence complémentaire de la production. Dans la perspective de notre étude, nous tenterons d'intégrer ce facteur de compréhension au travers de plusieurs dispositifs, dont les questions d'autoévaluation, les activités 4 (QCM lexical) et 5 (QCM sémantique) et la troisième partie de cette étude organisée autour des entretiens post-tests.

Par ailleurs, il nous semble au vu de ces résultats quantitatifs et statistiques que le développement de la compétence lexicopragmatique est le fruit de variables moins visibles et plus complexes. Ces variables n'agissent pas de manière isolée, mais comme plusieurs paramètres qui une fois articulés permettent aux locuteurs de réunir les conditions nécessaires au bon développement de leur compétence lexicopragmatique. Parmi ces variables, la fréquence, la qualité et la diversité d'exposition à la langue française couplée à un enseignement guidé et une motivation professionnelle ou didactique à l'apprentissage et au perfectionnement des phraséologismes pragmatiques sont ressortis dans cette étude statistique comme des facteurs facilitateurs de l'acquisition de la compétence lexicopragmatique.

Il faut néanmoins se garder d'une interprétation hâtive de ces résultats. En effet, les statistiques ne permettent pas d'affirmer qu'il y a un réel lien de causalité entre les variables observées et les résultats au test. Elles permettent en revanche de montrer que les résultats obtenus pour ces différentes variables ne sont pas dus au hasard. Les statistiques donnent une estimation de la significativité potentielle des résultats observés par rapport à une ou plusieurs variables. Elles permettent donc d'orienter ou de valider des hypothèses préalables. L'usage de mesures statistiques et quantitatives apparait comme un outil d'exploration des données tout à fait utile, mais ne semble selon nous pas être suffisant pour rendre compte des spécificités de la compétence lexicopragmatique des LNN observés dans cette étude. Pour avoir un éclairage pertinent des indicateurs obtenus par le traitement statistique et quantitatif, il apparait indispensable de confronter ces résultats à des données qualitatives nous permettant d'observer les spécificités de la compétence des informateurs au travers de leurs productions en contexte (Taguchi 2018). Ce faisant, nous allons dans le prochain chapitre étudier dans une perspective qualitative les réponses proposées dans le test afin de détailler les spécificités d'emploi et les difficultés rencontrées par les LNN dans l'emploi des PhPex de notre test.

Chapitre 12 - Analyse qualitative et comparative des données

Dans cette partie, nous analysons de manière individuelle chaque item proposé dans le test afin de caractériser le plus finement possible les spécificités d'emploi des PhPex relatives à la variété non native des LNN qui ont participé à notre étude. Rappelons les questions qui seront observées ici :

- Les LNN de niveau avancé produisent-ils moins de séquences préfabriquées que les locuteurs non natifs ? Rencontrent-ils plus de difficulté en production ou en compréhension ?

- Observe-t-on chez les LNN une dissymétrie entre la compétence de compréhension et celle de production des PhPex ?

- Ces difficultés se retrouvent à quel niveau de la compétence lexicopragmatique (pragmalinguistique, sociopragmatique et métapragmatique) ?

- Quelle est l'influence de la LM sur le choix du lexique, des stratégies de compréhension et de production ? quelle influence sur l'attitude communicationnelle transmise dans la production ?

- Quelles sont les spécificités lexicales et linguistiques permettant de caractériser l'emploi des PhPex chez les LNN étudiés ?

Concernant les trois premières activités de production, nous proposerons une analyse des spécificités lexicales retrouvées dans les résultats proposés par les informateurs de notre test sous forme de tableaux synthétiques. Les deux dernières activités de compréhension seront accompagnées des valeurs calculées pour le taux de compréhension correspondant à chaque item. Pour chaque expression observée, nous proposons dans un premier temps une définition prenant en compte leur spécificité d'emploi. Dans un second temps, nous observons les résultats des LNN que nous tenterons de contraster avec ceux des LN. Nous tenterons également de mettre en parallèle de nos observations certains des éléments statistiques analysés plus haut. Enfin nous proposerons une synthèse ainsi qu'une discussion des résultats. Nous rappelons qu'un tableau de synthèse de la structure du test est disponible dans la partie méthodologique de cette deuxième phase de notre étude (Chapitre 10, section

2, p. 219) et qu'un exemplaire du test est disponible dans l'annexe 8. Nous invitons le lecteur à se munir de ces compléments d'information pour étayer leur lecture de l'analyse.

1. Analyse linguistique de l'activité 1 (complétion)

1.1. Cadre de l'activité 1

L'activité 1 est une activité dite de *complétion* dans laquelle il est demandé aux informateurs de produire la suite d'un dialogue tronqué à un endroit prédéfini en employant une expression de préférence préfabriquée. Cette activité s'inspire du *discourse completion task*[160] (DCT) que nous avons discuté dans le cadre théorique de ce travail de recherche (Chap. 7, section 1). Le but de ce protocole exploratoire est double. Dans un premier temps, il s'agit de solliciter la compétence de compréhension de l'informateur afin qu'il puisse repérer les spécificités de la situation de communication (compétence sociopragmatique). Dans un second temps, nous testons sa compétence de production en lui demandant de continuer le dialogue lacunaire proposé dans l'activité à l'aide d'expressions préfabriquées (compétence pragmalinguistique).

Les deux dialogues proposés ont été conçus de manière à susciter une réaction expressive chez l'informateur. Dans le premier cas, nous avons mis en scène une discussion entre deux amis dans laquelle l'un des interactants confie à son interlocuteur qu'il vit un enfer au travail à cause de l'une de ses collègues. Le but ici est de faire exprimer son indignation et son exaspération à l'informateur. Afin d'initier la réponse, nous avons commencé le tour de parole à compléter avec *C'est…* afin d'aider l'informateur à trouver une réponse. L'emploi de ce type de « démarreur » a été suggéré lors d'une réunion de préparation du protocole expérimental[161] au cours de laquelle, nous nous sommes rendu compte que le fait de laisser un trou sans indication risquait de compliquer grandement la réponse à l'informateur. Nous avons donc imaginé cette stratégie afin de démarrer la production plus facilement chez les répondants. Dans le deuxième cas, nous avons proposé un cadre où une interlocutrice raconte à son ami qu'elle a failli être agressée par deux inconnus la veille en rentrant du travail tard le soir et qu'elle leur a échappé *in extrémis*. Le but dans ce deuxième dialogue est de faire

---

[160] Pour une discussion plus approfondie de ce mode de recueil de la production des apprenants voir : Kasper (2008 [1999]) ; Roever (2015) ; Sweeney & Hua (2016).
[161] Nous remercions ici Agnès Tutin (LIDILEM) pour cette suggestion.

exprimer à l'informateur un sentiment de surprise ou d'indignation et de colère envers les agresseurs. De même que pour le premier dialogue à compléter, nous avons introduit un démarreur en utilisant le PhPex *sérieux ?* qui marque la surprise négative dans le cas présent.

Notre hypothèse pour cette activité est que l'emploi des PhPex répond à des *schémas conversationnels* spécifiques. Ces schémas constituent un assemblage de paramètres contextuels associés de manière récurrente constituant ainsi une configuration conversationnelle culturellement stéréotypée dans laquelle il est possible d'employer de manière prédictible un inventaire d'expressions préfabriquées plus ou moins stable. Des exemples saillants de ces schémas conversationnels seraient par exemple l'expression de l'empathie, les flatteries ou les félicitations. Dans ces trois cas, la situation de communication en elle-même n'appelle pas l'utilisation quasi obligatoire d'une expression en particulier comme ce serait le cas avec les salutations (ou de manière générale les *pragmatèmes,* cf. Mel'čuk 1995). En revanche, ces situations activent l'emploi de toute une série d'expressions pouvant remplir le besoin de communication (thème abordé, action en cours, relation entre les participants, but ou lieu de la conversation). Nous avons donc essayé ici de voir si le LNN parvenait à produire une unité lexicale préfabriquée (ou plusieurs) en adéquation avec le schéma conversationnel proposé.

1.2. Spécificités lexicales des réponses observées dans l'activité 1

			Variabilité Lexicale	Variabilité registre		Indice phraséologique	Acceptabilité réponse (taux de réussite)		Formes les plus fréquentes[162]	Taux de compréhension
LNN	D1	Item 1	54%	**Courant**	34%	71%	**Acceptable**	83%	1. C'est chiant ! (3) 2. C'est dingue ! (3) 3. C'est énervant ! (2)	91%
				Familier	54%		**Inacceptable**	11%		
				Très familier	9%		**Sans réponse**	6%		
	D2	Item 1	45%	**Courant**	28%	57%	**Acceptable**	69%	1. C'est fou ! (2) 2. C'est pas possible !(2) 3. C'est terrible ! (2)	89%
				Familier	69%		**Inacceptable**	22%		
				Très familier	0%		**Sans réponse**	9%		
LN	D1	Item 1	66%	**Courant**	14%	94%	**Acceptable**	100%	1. C'est chiant ! (14) 2. C'est toujours la même chose ! (5) 3. Ça me saoule ! (3)	
				Familier	71%		**Inacceptable**	0%		
				Très familier	15%		**Sans réponse**	0%		
	D2	Item 1	66%	**Courant**	14%	86%	**Acceptable**	97%	1. C'est chaud ! (7) 2. Ça craint ! (4) 3. (C'est) n'importe quoi ! (3)	
				Familier	74%		**Inacceptable**	3%		
				Très familier	12%		**Sans réponse**	0%		

[162] Par ordre décroissant. Nombre d'occurrence entre parenthèses.

Tableau 15 : Spécificités lexicales des réponses analysées dans l'activité 1[163]

Le tableau 15 compose de 6 entrées indiquant chacune une donnée sur les spécificités lexicales des résultats enregistrés chez les LN et LNN. Nous rappelons que les méthodes de calcul et les hypothèses qui ont conduit à l'élaboration de ces mesures ont été développées dans la section 3.3 du chapitre 10 dédié à la méthodologie de cette deuxième phase de notre étude. Ces tableaux ont pour objectif de mettre en relief les spécificités lexicales des résultats des LNN tout en permettant de les contraster avec les emplois retrouvés chez les LN. Nous rappelons également que l'objectif de cette mise en contraste ne sous-tend pas que la variété native constitue le standard sur lequel nous allons fonder notre analyse.

La variabilité lexicale observée dans cette première activité affiche une différence de 15% en moyenne. Cette différence indique une plus grande diversité des formes employées chez les LN, mais ne constitue selon nous pas une différence assez grande pour avancer l'hypothèse d'une compétence lexicopragmatique insuffisante chez les LNN. On observe au niveau du registre une plus grande propension des LNN à employer le registre courant en comparaison des LN et un usage plus modéré du registre familier et très familier. L'indice phraséologique indique aussi une différence significative entre le nombre de séquences phraséologiques proposées par les LN et celles proposées par les LNN. L'indice phraséologique des LNN semble toutefois élevé et atteste d'un usage fréquent des phraséologismes dans cette activité. Le taux d'acceptabilité des réponses LNN, bien qu'inférieur à celui des LN, demeure élevé et semble montrer globalement une bonne compréhension des schémas de communication proposés. Enfin, le taux de compréhension calculé à partir des questions d'autoévaluation indique une compréhension de 90% en moyenne, ce qui vient corroborer l'hypothèse ici d'une certaine aisance des LNN dans cette activité de production libre.

---

[163] D = Dialogue

1.3. Analyse des réponses de l'activité 1[164]

1.3.1. Analyse comparative

Nous observons dans les deux cohortes une très forte majorité d'emplois suivant les deux patrons syntaxiques suivants : [c'est + ADJ / N] (*c'est chaud !*, *c'est galère !*) et [ça + V] (*ça craint !*). Les termes fréquemment associés à ces deux patrons syntaxiques sont à polarité négative et rentrent dans le champ lexical des termes d'indignation ou de surprise négative. Ces propositions semblent correspondre aux schémas conversationnels proposés en mettant à la fois en avant le caractère négatif de la situation et la dimension expressive avec l'indignation et l'inconfort de la situation qui ressortent fortement. De plus, les réponses proposées chez les LN et les LNN sont pour la plupart des PhPex plutôt fréquents dans la conversation quotidienne, ce qui correspond selon nous à un degré de compétence en production qui semble déjà bien stabilisé pour une bonne partie des LNN. On remarque chez les LN un emploi plus systématique des formes familières, voire dans quelques cas, de formes employant un registre très familier (*c'est chiant ! c'est casse-couilles ! putain c'est pas possible enfin !*). En outre, nous remarquons une prépondérance de l'emploi des formes exprimant l'indignation et la colère (ex. : *ils se prennent pour qui !, c'est n'importe quoi !*) chez les LN par rapport aux formes exprimant uniquement la surprise, plus présente chez les LNN (*c'est incroyable !*, *c'est vrai ?*).

À la différence des LN, les résultats des LNN montrent une répartition plus homogène entre ces deux catégories d'affect, ainsi que des réactions moins directes ou porteuses d'invectives. Les réactions émotionnelles exprimées par les LNN ont tendance à être plus introspectives et centrées sur le sentiment négatif ressenti par le locuteur plutôt que sur l'expression de la colère et du ressentiment envers le déclencheur de la réaction émotionnelle. On peut faire l'hypothèse ici qu'il s'agit d'une stratégie communicative proche de la LM des LNN. Par ailleurs, des expressions familières très fréquentes à l'oral *c'est chaud !, ça craint !, c'est abusé !* et *ça fout la rage !* ont une fréquence plus importante parmi les expressions relevées chez les LN, mais n'apparaissent pas dans les réponses relevées chez les locuteurs japonais.

---

[164] Le détail des réponses enregistrées pour les deux cohortes est consultable en annexe 9 et 10.

Nous avons également relevé quelques réponses chez les LNN à caractère référentiel, car non expressives et dépourvues de réelle valeur actionnelle (ex. : *elle se comporte toujours comme si c'est elle qui a raison* ou *c'est totalement égoïste ce qu'elle fait*). Dans ces différents cas, il s'agit plutôt d'un énoncé explicatif de la situation que d'une séquence pragmatique ayant une fonction expressive. En revanche, bien que ces expressions ne correspondent pas à des phraséologismes, ces réponses sont naturelles et tout à fait acceptables, mais comportent moins d'information pragmatique et discursive. On observe aussi que certaines des séquences lexicopragmatiques proposées sont connectées à un complément qui a pour fonction de spécifier les raisons de la réaction expressive. Ce procédé se retrouve à la fois chez les LNN (ex. : *c'est chiant de faire la tête pour un rien*) et chez les LN (ex. : *c'est vraiment frustrant qu'elle ne comprenne pas mon point de vue*). Les PhPex sont fréquemment expansés de la sorte dans cette activité de production libre qui laisse un plus grand choix dans le lexique à employer et une plus grande liberté dans la structure syntaxique de la réponse.

Nous observons chez les LN la forte récurrence d'énoncés périphériques accompagnant les PhPex employés, énoncés qui constituent à première vue des étayages de la réaction expressive précédemment exprimée. Toutefois, ces énoncés sont indépendants de la structure du PhPex dont ils précisent en général le contexte d'emploi (ex. : *j'ai pensé à aller la voir, mais au final je pense que c'est à elle de venir me voir, c'est un dialogue de sourds, on n'arrive pas à se comprendre c'est plutôt frustrant*). Ce type d'énoncé accompagnateur est fréquent lors de la conversation et permet de détailler la situation ou de renforcer l'effet illocutoire du PhPex. Les LNN ont employé ce type d'énoncé peu de fois en comparaison des LN. L'ajout de cet énoncé périphérique rend la réponse plus naturelle et interactive. Il y a, semble-t-il, ici un savoir-faire langagier qui touche plutôt à la dimension procédurale (norme d'emploi, attitude, stratégie et éthos communicatif) qu'à la dimension formelle (spécificités morphosyntaxiques, grammaticales et sémantiques) de la compétence lexicopragmatique et qui semble caractéristique d'une compétence lexicopragmatique avancée.

Nous notons par ailleurs que les LN emploient fréquemment des procédés d'intensification afin de renforcer le degré d'expressivité de leur production. Parmi ces procédés, nous remarquons l'emploi d'adverbes intensifieurs (*vraiment* / *tellement*) et l'usage du registre très familier qui renforcent la dimension réactive de la réponse en abolissant la distance

relationnelle. À l'inverse, nous avons remarqué chez les LNN l'emploi de l'expression *c'est un peu compliqué* qui repose une stratégie de minimisation de l'intensité expressive. Cette stratégie de minimisation sous-entend que la situation est tellement tendue que le locuteur ne souhaite pas s'étendre sur le sujet, qu'il souhaite en minimiser l'importance. Il y a donc ici une stratégie de modulation de l'intensité qui n'a pas été retrouvée dans les résultats des LN. À la manière de l'emploi des adverbes, nous notons également chez les LN l'emploi d'éléments satellites compacts accompagnant l'emploi du PhPex et permettant de renforcer sa dimension expressive. Il s'agit notamment de termes tels que *putain* (ex. : *putain c'est chaud*) ou que *quand même* (ex. : *c'est chaud quand même*). Ce phénomène ne se retrouve pas dans les réponses des LNN.

Dans le dialogue 2, nous avons pu remarquer chez les LNN de nombreuses réponses empathiques exprimant le réconfort ou la peur partagée pour la situation vécue par l'interlocuteur (ex. : *ça va ? Tu n'as pas eu de problème).* Cette dimension de réconfort se traduit par des stratégies discursives assez complexes visant à s'assurer de l'état mental ou physique de l'interlocuteur ou exprimer de l'empathie à l'égard de l'interlocuteur. Afin de mettre en œuvre cette stratégie communicative d'empathie, les LNN mobilisent des structures affirmatives visant à exprimer une forte empathie concernant le traumatisme vécu et la peur ressentie (ex. : *tu as dû avoir la trouille, tu devais avoir très peur*) et visant parfois à exprimer un certain soulagement que l'interlocuteur soit en bonne santé (ex. : *Heureusement* + que / *tu l'as échappé belle*). Nous faisons ici l'hypothèse d'un possible transfert de l'éthos communicatif japonais. En effet, en japonais, il est fréquent d'exprimer son empathie envers quelqu'un qui a vécu un moment difficile. Cela fait notamment partie d'une stratégie de convergence émotionnelle. Enfin, on observe dans cette tendance à l'empathie l'usage à plusieurs reprises d'expressions visant à témoigner sa peur vis-à-vis des évènements qui sont racontés (ex. : *ça fait trop peur, ça me fait peur, j'ai trop peur pour toi*). Nous faisons également l'hypothèse ici d'un transfert expressif emprunté à la langue maternelle. En effet, il est assez fréquent que lorsqu'un locuteur japonais est directement ou indirectement confronté à une situation très surprenante, il réagit en exprimant une certaine peur ou appréhension. Ces deux hypothèses du transfert nous semblent d'autant plus plausibles que les résultats retrouvés chez les locuteurs natifs tendent à un sous-emploi des formes de réconfort ou d'empathie/peur au profit de l'usage prépondérant des formes permettant d'exprimer l'indignation.

1.3.2. Analyse des difficultés

Concernant les réponses jugées inadéquates, nous avons constaté chez les LNN des difficultés de plusieurs ordres. Le premier type de difficulté correspond à des emplois non pertinents du point de vue pragmatique. Dans ce type d'erreur, le locuteur produit des réponses qui sont soit contradictoires, soit inadaptées sur le plan expressif et/ou discursif à la situation de communication présentée. Par exemple, la proposition de réponse *c'est nul* n'a pas été acceptée par les juges, car elle a une valeur évaluative, mais n'exprime pas la colère ou l'indignation. Lorsqu'elle a une valeur expressive, cette expression est généralement employée pour exprimer une déception banale. De même, la proposition de réponse *c'est un truc de merde* ne semble pas constituer non plus un emploi adéquat dans le contexte de conversation proposé. Elle ne traduit pas spécifiquement l'indignation, ni la colère éprouvée par le locuteur, mais plutôt une évaluation à polarité négative. On peut formuler l'hypothèse ici d'une influence lexicale de la langue maternelle. En effet, le locuteur japonais emploie assez fréquemment dans ce contexte des formules telles que *saïaku* (*le pire*)[165] qui expriment au départ un jugement évaluatif, mais qui avec l'usage répété dans la conversation quotidienne, a acquis une fonction expressive d'indignation.

Le deuxième type de difficulté correspond à un emploi erroné du point de vue lexical ou combinatoire. Il s'agit notamment de réponses partielles ou lacunaires dans lesquelles il manque une unité lexicale, de combinatoires bancales sur l'axe syntagmatique ou de l'insertion d'une unité qui engendrent l'actualisation d'un sens et d'une fonction inadaptée au schéma de communication proposé. Nous pouvons prendre l'exemple de la réponse *c'est absolument pas possible !* dans laquelle l'insertion de l'adverbe d'intensification *absolument* dans l'expression *c'est pas possible !* neutralise le sens expressif de surprise ou d'indignation pour lui redonner son sens premier de marquage l'impossibilité ou du refus. Le LNN à l'origine de cette réponse a probablement voulu renforcer le côté expressif du PhPex, car il a bien compris que le locuteur était excédé dans le schéma conversationnel proposé. Sa compétence métapragmatique ne lui a cependant pas permis de repérer que la combinatoire générée actualisait un sens et une fonction incompatible avec la situation proposée.

---

[165] Nous remercions Tomoko Higashi (LIDILEM) qui nous a suggérer cette piste lors de la relecture de cette thèse.

Le troisième type de difficulté porte sur la compétence stratégique et notamment la capacité d'ajustement de l'intensité expressive et de l'intention discursive véhiculée. Un premier exemple de ce type de difficulté est la proposition de réponse *c'est dégueulasse !*. La dimension expressive première de cette expression n'est pas d'exprimer l'indignation, mais le rejet violent en réaction à une chose ou un évènement particulièrement révoltant. Cette fonction d'indignation très forte rend cet emploi peu approprié dans le contexte du premier dialogue de l'activité 1. Le fait que cette expression dotée d'une forte charge expressive soit employée dans un schéma conversationnel dans lequel le locuteur se plaint de sa collègue qui l'énerve donne une impression de non-naturel, car trop intensif. À l'inverse, la proposition *c'est dommage* semble avoir une intensité expressive trop faible pour être employée dans le contexte de l'exaspération. Ce PhPex est généralement employé pour exprimer une déception, sans pour autant qu'elle soit liée à de la colère. La nuance que peut revêtir cette expression est celle aussi du regret ou de l'acte inachevé ou manqué. Or, il nous semble que dans le cas du dialogue 1, tout laisse à penser que le locuteur ne regrette pas, mais au contraire est plein de ressentiment envers sa collègue.

1.4. Synthèse intermédiaire de l'activité 1

L'analyse des données relevées pour cette première activité a montré un taux de réussite élevé dans les deux cohortes. Ce haut taux de réussite chez les LNN semble indiquer une compétence lexicopragmatique plutôt stable pour ce type d'activité de complétion. Cela se remarque d'autant plus au niveau des réponses enregistrées dans les deux cohortes qui ne présentent pas de grandes disparités. De même, les expressions employées par les LNN étaient pour la plupart adaptées et bien intégrées aux schémas conversationnels proposés, ce qui semble indiquer là aussi une certaine maitrise de l'inférence des différents paramètres paralinguistiques impliqués par les situations proposées. Cette capacité d'interprétation élevée se confirme avec des taux de compréhension de 89% et 91% pour le dialogue 1 et 2. La dimension expressive de l'indignation ou de la surprise semble également bien maitrisée. Nous avons pu également remarquer l'émergence d'une forte tendance aux marques d'empathie et de connivence chez les locuteurs japonais alors que chez les locuteurs natifs, c'est la dimension d'indignation qui a été exprimée en priorité. Il nous a donc semblé que les éthos communicatifs et les marques expressives semblaient ici différer entre les LN et les LNN, et que cette variation pouvait constituer une trace de l'influence de la manière de communiquer en langue maternelle.

En outre, nous avons remarqué que les LNN modulaient très peu l'intensité expressive de leur réponse à l'aide d'éléments de modalisation comme ceux repérés chez les natifs. Cela est selon nous dû au fait que les LNN ne semblent pas encore disposer d'une palette lexicopragmatique suffisamment riche pour accéder aux outils permettant de moduler la nuance ou l'intensité expressive du dire. Au niveau de la combinatoire des PhPex cette fois-ci, nous avons noté que l'insertion des adverbes intensifieurs ne semblait pas complètement acquise chez les LNN alors que les LN en ont un usage plus systématique. Enfin, l'usage d'éléments satellites de renforcement et de phrases périphériques d'étayage a été moins observé chez les LNN que chez les LN.

2. Analyse linguistique de l'activité 2 (reformulation)

2.1. Cadre de l'activité 2

L'activité 2 est une activité dite de reformulation. Il est demandé aux informateurs de produire une paraphrase explicative permettant de rendre compte de l'expression soulignée tout en conservant les mêmes propriétés sémantiques et pragmatiques initiales. Par exemple, l'expression *ça me prend la tête* pourra être paraphrasée par *je suis très agacé par cette situation*. Cette activité combine à la fois une tâche de complétion par substitution et une tâche d'élicitation de l'illocutoire (Partie théorique, Chap. 7, section 1). Le but de ce second protocole exploratoire est d'observer l'articulation entre la compréhension et la production. Ce protocole a été élaboré afin que l'informateur mobilise sa compétence pragmalinguistique pour interpréter le sens et la fonction de l'expression proposée pour ensuite proposer une reformulation adéquate au contexte (compétence sociopragmatique). Il ne s'agit donc ni de gloser précisément l'expression cible (ce qui serait le travail du linguiste ou du lexicographe) ni de proposer un équivalent fonctionnel (ce qui sera demandé dans l'activité 3). Les dialogues proposés sont des interactions quotidiennes et informelles tantôt entre amis, tantôt entre membres d'une fratrie. Les schémas conversationnels qui ont été utilisés sont diversifiés et mettent en scène des emplois expressifs à polarité positive ou à polarité négative. Les expressions choisies pour cette activité ont été sélectionnées en fonction de leur forte idiomaticité et de la faible prédictibilité de leur emploi. Nous avons aussi intégré à cette activité un distracteur afin de nous assurer que les réponses au test proposées par les LNN ne sont pas le fruit du hasard et sont bien dues à leur compétence à différencier un

emploi phraséologique d'un emploi littéral. Rappelons que l'emploi des distracteurs est aussi un bon moyen de détecter les effets de *speeding*, notamment chez les LN.

En prémices de toute analyse, nous mentionnons ici une constatation importante concernant l'adéquation des réponses avec la consigne formulée pour l'activité 2. En effet, que ce soit chez les LN ou les LNN, nous avons pu remarquer des réponses qui s'apparentent plus à des propositions d'équivalent qu'à des reformulations. On peut ici émettre l'hypothèse d'une complexité liée à la consigne et notamment au fait de demander une reformulation du sens et de la fonction aux informateurs. Suite à plusieurs discussions menées lors de l'analyse des données, nous nous sommes aperçu que l'acte de reformulation est complexe à mettre en place et aurait nécessité plus de temps de réponse pour permettre aux répondants de faire le travail demandé, ainsi qu'une consigne plus explicite et détaillée. Nous avons malgré tout décidé de conserver ces réponses et de les analyser en tant que telles, car ces résultats présentent des éléments fort utiles pour l'analyse des spécificités de la compétence lexicopragmatique en production des LNN observés.

2.2. Spécificités lexicales des réponses observées dans l'activité 2

			Variabilité Lexicale	Variabilité registre		Indice phraséologique	Acceptabilité réponse (taux de réussite)		Formes les plus fréquentes	Taux de compréhension
LNN	D1	C' est le pied !	37%	Courant	43%	74%	Acceptable	77%	1. C'est super ! (11) 2. C'est génial ! (3) 3. C'est le top ! (3)	91%
				Familier	57%		Inacceptable	20%		
				Très familier	0%		Sans réponse	3%		
	D2	C' est pas possible (distracteur)	31%	Courant	86%	49%	Acceptable	83%	1. Je ne peux pas (13) 2. Je ne suis pas disponible (11) 3. Je suis occupé demain (2)	97%
				Familier	14%		Inacceptable	17%		
				Très familier	0%		Sans réponse	0%		
	D3	Faut pas pousser !	57%	Courant	63%	8%	Acceptable	34%	1. Il ne faut pas exagérer ! (2) 2. C'est beaucoup trop ! (2)	81%
				Familier	37%		Inacceptable	60%		
				Très familier	0%		Sans réponse	6%		
		C' est abusé !	74%	Courant	50%	45%	Acceptable	60%	1. Ce n'est pas possible !(4) 2. C'est trop demander ! (3) 3. C'est exagéré ! (2)	81%
				Familier	44%		Inacceptable	37%		
				Très familier	6%		Sans réponse	3%		
	D4	C' est pas possible !	63%	Courant	20%	62%	Acceptable	49%	1. C'est incroyable ! (11)	90%
				Familier	74%		Inacceptable	51%		
				Très familier	6%		Sans réponse	0%		

		J' y suis pour rien !	20%	**Courant** **Familier** **Très familier**	63% 37% 0%	91%	**Acceptable** **Inacceptable** **Sans réponse**	89% 8% 3%	1. Ce n'est pas de ma faute ! (23) 2. Je n'ai rien fait ! (6) 3. Ce n'est pas à cause de moi ! (2)	91%
LN	**D1**	**C' est le pied !**	46%	**Courant** **Familier** **Très familier**	6% 94% 0%	97%	**Acceptable** **Inacceptable** **Sans réponse**	97% 3% 0%	1. C'est génial ! (10) 2. C'est trop bien ! (7) 3. C'est super ! (6)	
	D2	**C' est pas possible (dis tracteur)**	46%	**Courant** **Familier** **Très familier**	46% 54% 0%	71%	**Acceptable** **Inacceptable** **Sans réponse**	100 % 0% 0%	1. C'est mort ! (3) 2. Je peux pas ! (15) 3. Y'a pas moyen (2)	
	D3	**Faut pas pousser !**	57%	**Courant** **Familier** **Très familier**	14% 83% 3%	80%	**Acceptable** **Inacceptable** **Sans réponse**	94% 6% 0%	1. C'est abusé ! (6) 2. Faut pas abuser ! (6) 3. Faut pas exagérer ! (4)	
		C' est abusé !	51%	**Courant** **Familier** **Très familier**	17% 80% 3%	75%	**Acceptable** **Inacceptable** **Sans réponse**	94% 0% 6%	1. C'est trop ! (7) 2. C'est n'importe quoi ! (5) 3. C'est exagéré ! (2)	
	D4	**C' est pas possible !**	80%	**Courant** **Familier** **Très familier**	6% 83% 11%	95%	**Acceptable** **Inacceptable** **Sans réponse**	86% 14% 0%	1. C'est pas vrai ! (4) 2. Sérieux ? / sérieusement ? (3)	
		J' y suis pour rien !	20%	**Courant** **Familier** **Très familier**	37% 63% 0%	100%	**Acceptable** **Inacceptable** **Sans réponse**	97% 0% 3%	1. C'est pas ma faute ! (27) 2. J'ai rien fait ! (3)	

Tableau 16 : Spécificités lexicales des réponses analysées dans l'activité 2

Concernant la variabilité lexicale observée dans cette seconde activité, nous pouvons observer en moyenne une différence entre les LN et les LNN de 15%. Là aussi, cette différence semble indiquer une plus grande diversité des formes chez les LN, mais ne constitue pas nécessairement selon nous le signe d'une compétence lexicopragmatique insuffisante chez les LNN. Cette différence peut être en revanche un indicateur d'un répertoire lexical spécifique à la conversation orale quotidienne moins diversifié que les LN. Par ailleurs, nous avons constaté des valeurs inhabituelles dans la variabilité lexicale sur 3 items : une égalité à l'item 1 (*faut pas pousser !)* du dialogue 3, un taux plus élevé pour les LNN à l'item 2 (*c'est abusé !*) du dialogue 3 et à nouveau une égalité à l'item 2 (*j'y suis pour rien !*) du dialogue 4. Concernant les cas d'égalité, une hypothèse serait que dans les deux cas, il est relativement complexe de reformuler de plusieurs manières différentes les deux expressions proposées en raison du champ d'emploi restreint dont elles font l'objet. Concernant le taux plus élevé chez les LNN, nous pensons à deux hypothèses : 1) les possibilités d'interprétation de l'expression *c'est abusé !* sont plus large et permettent un

panel de réponses plus grand, 2) il y a un consensus au niveau de la reformulation possible plus grand chez les LN (moins de variabilité, car plus de réponses similaires) que chez les LNN (plus de variabilité, car plus de réponses différentes). La variabilité du registre montre globalement un usage plus important du registre courant chez les LNN que chez les LN. On constate que le dialogue 2 a donné lieu à un fort taux de réponse utilisant le registre courant chez les LN. Cela peut être dû à l'expression *c'est pas possible* qui était employée dans son sens d'incapacité ou d'impossibilité. Ce faisant, le caractère peu expressif de l'expression à reformuler a donné lieu à un plus grand nombre de reformulations standards. Par ailleurs, l'acte de refus étant généralement un acte de langage pouvant possiblement engendrer une tension entre les interactants, il aura tendance à être formulé de façon à ne pas être trop direct et trop familier.

L'indice de phraséologisation montre un taux d'emploi plus important de structures préfabriquées chez les LN que chez les LNN. Toutefois, cet indicateur sera observé avec précautions compte tenu du fait qu'en principe, il était attendu des reformulations et non pas des équivalents phraséologisés. Les équivalences proposées étant nombreuses, nous avons choisi de mentionner cet indice également. Le score le plus bas enregistré pour cette activité est de 8% et concerne l'expression *faut pas pousser !.* L'idiomaticité de l'expression, sa faible prédictibilité et le fait qu'aucune traduction[166] n'existe dans la langue maternelle des LNN semblent être une explication possible à ce taux très faible. On constate un taux de phraséologisation s'élevant à 91% chez les LNN pour l'expression *j'y suis pour rien !.* La spécialisation de l'expression au contexte très prédictible de la justification et du dédouanement a sans doute eu un effet sur l'homogénéisation des réponses. Le taux d'acceptabilité des réponses est globalement élevé chez les LNN pour cette activité. Le taux le plus faible enregistré est de 39% de réussite et vient confirmer nos hypothèses concernant l'expression *faut pas pousser !* dont le sens et la fonction semblent complexe à inférer à partir du contexte. Enfin le taux de compréhension affiche un score de 88% en moyenne, ce qui semble indiquer une bonne compréhension globale des situations de communication observées.

---

[166] Cette constatation a été confirmée par Tomoko Higashi (LIDILEM). Elle ajoute par ailleurs que la structure *il faut pas* utilisé en réaction à un événement ou un jugement négatif n'existe pas en japonais.

2.3. Analyse des réponses de l'activité 2

2.3.1. Analyse comparative

Dialogue 1

Le premier dialogue montre une conversation avec l'utilisation du PhPex *c'est le pied !*. Cette expression est ici employée pour exprimer un sentiment agréable de satisfaction qui peut être physique ou psychologique. À l'instar de la première activité, nous observons dans les deux cohortes une forte productivité du patron syntaxique [*c'est* + ADJ / N]. Les expressions proposées correspondent dans la plupart des cas analysés à une axiologie positive pour le premier item et une axiologie négative pour les items suivants. De même, une grande partie des réponses proposées par les LNN semblent répondre aux schémas interactionnels proposés sur plan pragmatique et expressif. Nous remarquons dans ce premier dialogue l'emploi chez les LNN de quelques adjectifs superlatifs tels que *top*, *extra* ou *super* utilisés ici pour souligner l'axiologie positive forte de la réaction expressive. De même, nous observons l'emploi d'adverbes d'intensification (*très, trop, vraiment*) qui semble témoigner de la capacité des LNN à moduler l'intensité expressive pour ce type de schéma de conversation à polarité positive. Nous observons le même type d'emploi chez les LN d'adjectifs superlatifs (*top, super, génial, bien, énorme*) ou d'adverbes d'intensité (*vraiment, trop, très*). L'emploi de ces termes est plus fréquent chez les LN que chez les LNN et nous avons pu remarquer uniquement chez les LN que ceux-ci avaient tendance à moduler l'intensité expressive de leur propos en employant des *adverbes intensifieurs dupliqués*, c'est-à-dire deux adverbes intensifs combinés ensemble permettant ainsi de renforcer grandement l'intensité de l'expression. C'est notamment le cas d'expressions telles que *c'est* ***vraiment super*** *bien* ou *c'est* ***vraiment trop*** *bien*. Nous observons par ailleurs chez les LNN quelques formes familières (*c'est dingue ! c'est un truc de dingue ! c'est le top !*), mais cela reste assez marginal en proportion des emplois familiers des LN (par ex : *y'a pas mieux ! j'adore ! j'ai atteint le septième ciel ! c'est le feu !*, etc.). Ces propositions sont dotées d'une forte idiomaticité qui les rendent opaques sur le plan sémantique et peu prédictibles sur les plans pragmatiques et fonctionnels.

Dialogue 2

Le second dialogue montre une discussion informelle avec l'utilisation de la séquence *c'est pas possible* dans son acception de refus ou de l'incapacité à réaliser une action. Il s'agit

donc d'une acception non expressive et compositionnelle de la séquence lexicale. Cette séquence constitue un *distracteur* (cf. Chap. 10, section 2). De manière très surprenante, nous remarquons que la majeure partie des réponses enregistrées pour ce dialogue comporte cette fois-ci un grand nombre de reformulations conformes à la consigne formulée pour l'activité 2. La plupart de ces reformulations prennent la forme d'une structure phrastique à la forme négative qui permet de rendre compte de la dimension de refus exprimée dans le dialogue (*je ne suis pas disponible, je suis occupée, je ne peux pas, etc.*). En outre, nous observons l'emploi chez les LNN d'une forme plus familière *ça va être chaud* qui constitue cependant un cas isolé parmi le nombre important de formes standards retrouvées dans les reformulations proposées. Nous observons également quelques formes idiomatiques telles que *je ne peux pas me libérer* et *ça ne m'arrange pas* qui sont fréquemment employées dans ce type de schéma interactionnel. Une autre différence entre les emplois enregistrés chez les LN et les LNN est celle de l'emploi des formes standards telles que la double négation qui sont beaucoup plus présentes chez les LNN[167] que chez les LN. En effet, le relevé des résultats des LN nous montre une grande diversité des structures de refus employées avec une prépondérance de l'emploi de la négation simple et de réductions morphologiques propres à la langue orale quotidienne. Le registre qui prédomine dans les réponses des LN est le registre familier avec un certain nombre de PhPex caractéristique de la conversation quotidienne (*ouais non demain y'a pas moyen, ça va être tendu, c'est chaud !,* etc.). En outre, nous observons que les expressions de refus proposées par les LN se déclinent en trois motifs qui correspondent à trois stratégies de refus différentes le *motif de l'indisponibilité* (*je suis occupée, je suis prise, j'ai un empêchement*), *motif de l'incapacité* (*Je pourrai pas, je peux pas, je vais pas pouvoir*) et *motif de la faisabilité* (*ça va être chaud, ça va être trop galère, c'est mort !*), ce qui n'est pas le cas dans les réponses des LNN.

Dialogue 3

Le troisième dialogue questionne l'usage des PhPex *faut pas pousser !* et *c'est abusé !*. Le PhPex *faut pas pousser !* est employé par le locuteur en réaction à un évènement, un dire ou un comportement qui dépasse les limites de ce qui est conventionnellement acceptable, qui est exagéré et hors de la sphère du raisonnable. La réaction implique en principe une

---

[167] A ce stade de l'analyse, et après réflexion, nous nous demandons si cette sous-représentation des formes élidées ne serait pas due au format écrit des réponses qui aurait inconsciemment favorisé la production de formes standards plutôt que familières chez les LNN.

dimension expressive d'indignation et d'agacement. Le deuxième PhPex proposé dans le dialogue est l'expression *c'est abusé !* et est employé en réaction à un évènement, un dire ou un commentaire étant excessif ou tirant parti de la situation de manière à affecter négativement le locuteur. La réaction expressive implique généralement l'indignation et parfois le sentiment d'impuissance lorsque le locuteur est débordé par les évènements. L'analyse des taux de réussite tend à montrer que les LNN ont éprouvé de grandes difficultés pour interpréter puis reformuler les PhPex proposés, notamment l'expression *faut pas pousser !* qui semble constituer une expression dont la fonction est particulièrement imprédictible. Globalement, les LNN ont su bien cerner la dimension négative du contexte ayant déclenché l'usage des deux expressions. On retrouve dans les réponses proposées par les LNN tout un champ sémantique autour des notions d'excédent ou d'exagération (*beaucoup, trop*, *exagérer*, *autant*, *abusif*) qui montre bien que le contexte et l'illocutoire produit par les expressions du dialogue sont compris. Nous remarquons également qu'une petite partie des LNN a fait ouvertement référence au sentiment d'exaspération en employant les expressions *c'est chiant, c'est pénible* ou *c'est pas possible*. Quelques autres LNN semblent même avoir employé des expressions exprimant le caractère insensé de la situation (*c'est ridicule, c'est hallucinant*). Toutefois, il ressort de l'étude qualitative des réponses à ce dialogue que globalement, les autres formes proposées par les LNN ne suffisaient pas à pleinement restituer le sens, l'expressivité et la fonction des deux expressions de base.

Nous avons également observé un manque de précision au niveau fonctionnel et expressif dans les réponses des LNN. En effet, ces réponses tendent à ne pas rendre complètement compte des différentes modalités d'usage des expressions de base, mais de certains paramètres ou états des choses liés à la situation d'énonciation en elle-même. Par exemple, certaines des réponses mettent en avant la dimension du travail à outrance (*il ne faut pas nous forcer trop, faut pas trop demander, il vaut mieux de ne pas trop faire travailler*) ou sur les conditions de travail précaires (*on ne peut pas travailler vite, on ne peut pas finir*). D'autres reformulations insistent sur les conséquences directes de la situation dans laquelle se trouvent les locuteurs (*ça me donne le stress, on est fatigué*), sur le caractère non approprié du comportement de l'enseignant (*le prof n'est pas gentil, ils sont méchants*) ou sur l'incompréhension liée à la situation (*notre situation n'est pas comprise, ça ne peut pas être justifié*). En outre, certaines de ces formulations sont parfois approximatives, ce qui tend à générer des ambigüités pour la compréhension des réponses. Par exemple, pour la réponse *il faut pas trop forcer*, on peut se demander si le locuteur fait allusion au sens 'ne pas trop

investir de l'énergie dans le travail demandé' ou bien au sens de 'il ne faut pas que les enseignants nous forcent à faire ce travail dans ces conditions très contraignantes' ? Il y a donc certes une bonne compréhension, mais toutefois des difficultés à exprimer avec précision le sens le PhPex reformulé ou à proposer un équivalent.

Notre analyse des résultats des LN a montré que la modulation de l'expressivité se fait au travers de divers dispositifs linguistiques qui sont peu, voire pas employé dans les résultats des LNN. Ainsi, nous remarquons que les LN emploient spontanément plus de tournures emphatiques que les LNN avec notamment l'usage d'adverbes intensifs (*franchement, beaucoup, vraiment, beaucoup, trop*), d'adverbes intensifs dupliqués (*beaucoup trop*) et de termes satellites ayant une fonction discursive expressive qui permet de renforcer l'expressivité du PhPex de base ([PhPex + *quoi*], ex. : *c'est abusé quoi !* / [PhPex + *sérieux*], ex. : *c'est abusé sérieux !* / [PhPex + *non plus*], ex. : *faut pas abuser non plus !*). Ces derniers termes satellites sont extrêmement peu employés par les LNN en français. Nous observons par ailleurs la présence de structures fortement idiomatiques et spécifiques au registre familier (par ex. : *c'est abusé ! faut se calmer ! ça craint ! ils forcent !*) qui là aussi ne se retrouvent pas dans les productions des LNN. L'emploi de ces formes familières, voire parfois vulgaires (*faut pas nous prendre pour des cons non plus !*) permet de renforcer le côté expressif de la prise de parole du locuteur. En outre, au niveau combinatoire, nous remarquons que les réponses des LN formulés autour des verbes *abuser* et *exagérer* semblent suivre le même schéma de construction : soit le schéma [*faut pas* + V] (ex. : *faut pas abuser !, faut pas exagérer !*), le schéma [c'est + ADJ] (ex. : *c'est abusé !, c'est exagéré !*) ou le schéma [Pro + V_{ind}] (ex. : *ils abusent !, ils exagèrent !*). D'autres expressions suivent la matrice phraséologique construite sur le verbe *falloir* [*faut pas* + V_{inf}] ou [*faut* + V_{inf}] (*faut arrêter !, faut se calmer !, faut pas déconner !, faut pas pousser !*). Comme évoqué précédemment, ce type de structures n'existant pas en japonais, elles sont quasi absentes des réponses des LNN.

Dialogue 4

Le quatrième dialogue teste la compréhension des expressions *c'est pas possible !* et *j'y suis pour rien !*. L'expression *c'est pas possible !* est employé par le locuteur pour exprimer sa surprise, son étonnement ou sa consternation face à un évènement, un état des choses ou un comportement inattendu ou invraisemblable. L'expression peut être également employée de manière ironique ou pour exprimer la colère. Son emploi permet au locuteur de désapprouver

ou de manifester son incompréhension, voire même d'exprimer l'incrédulité. Ce faisant, il nie la possibilité même de ce qu'il constate, ce qui est une manière pour lui d'exprimer le fait que ce à quoi il réagit est inconcevable[168]. Le PhPex *J'y suis pour rien !* est employé par le locuteur pour se dégager de la responsabilité d'une action dont les conséquences sont négatives pour son ou ses interlocuteurs. L'emploi de l'expression peut aussi exprimer la colère ou l'angoisse liée à une accusation dont le locuteur souhaite se dédouaner. Sur le plan pragmatique, l'expression permet d'atténuer le degré d'implication du locuteur d'un acte aux conséquences négatives, et lui permet de prendre de la distance énonciative vis-à-vis de son interlocuteur en niant sa responsabilité dans les évènements en cours[169]. Les deux expressions sont à axiologie négative, opaques sur le plan sémantique et peu prédictibles sur les plans fonctionnel et discursif. Notons que l'emploi de *c'est pas possible !* est ici en lien avec le distracteur du dialogue 2 et permet de faire contraster dans cette activité le sens littéral et le sens phraséologique.

Les réponses proposées dans les deux cohortes pour l'expression *c'est pas possible !* comportent un certain nombre de PhPex permettant d'exprimer l'exaspération et la surprise à axiologie négative tels que *t'es pas sérieuse ?, j'y crois pas ! ou c'est pas vrai ça !*. Ces expressions fondent leur illocutoire sur un procédé aléthique de négation de la réalité visant à renforcer le sentiment d'indignation et de rejet de la situation préjudiciable subie par le locuteur. La productivité de ces emplois aléthiques n'est pas la même dans les deux cohortes, on remarque un emploi plus fréquent chez les LN de cette stratégie discursive. D'autres réponses mettent l'accent sur l'agacement ressenti par le locuteur au vu du manque de précaution de sa sœur (*ça m'énerve, c'est insupportable, tu fais des conneries, tu peux faire plus attention*). On remarque que les réponses des LNN enregistrent une plus grande fréquence d'emplois standards (*c'est insupportable, tu aurais dû faire attention, mais qu'est-ce que tu fais ?*), avec peu d'utilisation des formes très familières fréquemment observées chez les LN (*putain sérieux !, putain fais gaffe !, j'hallucine !*, etc.). Concernant les reformulations de l'expression *c'est pas possible !* nous observons chez les LNN plusieurs réponses autour du motif de l'incompréhension (*je ne comprends pas comment tu as pu le*

---

[168] Pour plus de détail sur l'expression, nous renvoyons au travail lexicographique de Krzyżanowska, Grossmann et Kwapisz-Osadnik (2021 : 271).

[169] Idem p. 392

faire, je ne comprends pas ton comportement, comment ça se fait ?, je ne comprends pas pourquoi tu es tombée ici, etc.), motif inexistant dans les réponses des LN.

Par ailleurs, l'observation des données des LN nous montre une grande variété de formes et de stratégies discursives avec pour certaines expressions non seulement de l'indignation, mais aussi du reproche (*fais gaffe, c'est pas possible de faire attention ?, tu le fais exprès !, incroyable d'être comme ça !*). Sur le plan morphosyntaxique, nous observons principalement que les omissions de la double négation et les formes elliptiques sont plus fréquentes dans les productions des LN que celles des LNN[170]. Sur le plan combinatoire, nous observons uniquement chez les LN l'usage d'un procédé d'emphase expressive à l'aide de l'adjonction de constituants périphériques antéposés ou postposés au PhPex (***non, mais*** *c'est fou !,* ***mais*** *c'est terrible* ***ça*** *!,* ***putain*** *fais gaffe,* etc.)[171]. L'emploi de ces satellites en antéposition exprime une valeur d'opposition et/ou de contradiction et permet de renforcer l'étonnement exprimé par le locuteur. L'emploi de l'interjection très familière *putain* en antéposition permet de mettre en valeur la dimension expressive à polarité négative de l'expression. Comme évoqué précédemment, l'emploi de ces constituants périphérique se retrouve peu dans les résultats des LNN.

2.3.2. Analyse des difficultés

La première difficulté observée dans les résultats des LNN de cette activité concerne l'ajustement du registre de langue par rapport au registre attendu dans la situation de communication proposée. En effet, le registre employé dans les propositions de réponse au test des LNN ne semble pas tout à fait naturel dans les contextes de communication familière proposés en comparaison des réponses enregistrées chez les LN. Ainsi, on remarque par exemple la forte productivité de réponses telles que *c'est très bien !, c'est magnifique, c'est sublime* ou *c'est parfait* pour reformuler le PhPex *c'est le pied !*, expressions qui semblent trop standards pour une conversation entre amis. Cette différence d'ajustement du registre de langue entre LN et LNN se remarque d'autant plus dans le choix lexical effectué par les

---

[170] Nous avions d'ailleurs observé le même phénomène dans le chapitre précédent concernant l'analyse diagnostique de la production orale de locuteurs japonophones de niveau A2/B1.

[171] Krzyżanowska, Grossmann et Kwapisz-Osadnik (2021) remarquent également la présence des constituants périphérique dans bon nombre de formules expressives observées dans leur travail lexicographique. Ils noteront que ces constituants peuvent permettre de désambigüiser le sens de la formule expressive lorsqu'elle est polyfonctionnelle.

LN où l'on constate l'occurrence d'expressions telles que *y'a pas mieux !, c'est le feu !* ou *c'était énorme !* qui sont plus adaptées à une communication quotidienne. Cette difficulté d'ajustement du registre se retrouve également dans l'usage de la double négation ou des formes complètes. Par exemple, la forme *ce n'est pas de ma faute* est très productive chez les LNN alors qu'elle est beaucoup moins employée chez les LN qui auront tendance à employer principalement la forme réduite (négation simple, pas de préposition *de*) *c'est pas ma faute*.

Le deuxième type de difficulté observée concerne les choix combinatoires et leurs effets sur l'actualisation, ou non, de la fonction pragmatique visée en contexte. En effet, comme nous l'avions observé dans l'analyse de la première activité, les choix opérés sur la combinatoire lors de la sélection des éléments lexicaux qui constituent le PhPex cible peuvent grandement influencer la fonction et l'intensité de l'expression produite. Ainsi, nous remarquons à nouveau dans cette activité des difficultés combinatoires notamment sur le choix lexical opéré, à propos des adjectifs à axiologie positive sélectionnés pour intégrer la matrice phraséologique [*c'est* + ADJ / N]. Par exemple, dans le premier dialogue de cette activité, le choix des adjectifs employés par les LNN pour cette matrice ne rend pas complètement compte de la fonction pragmatique ou expressive impliquée par l'emploi de *c'est le pied !*. Il s'agit notamment de l'emploi d'expressions telles que *c'est bien !, c'est magnifique !* ou *c'est superbe !*. Nous pouvons par exemple observer que *c'est bien !* présente un problème au niveau de l'intensité expressive et dans sa portée pragmatique. En effet, l'expression *c'est bien !* renvoie à ce qui est convenable ou acceptable et ce faisant, à un jugement appréciatif projeté sur un évènement ou un objet. Il s'agit ici d'une appréciation de la qualité intrinsèque du déclencheur avec peu de substrat émotionnel. Il y a donc une incompatibilité avec le sens de *c'est le pied !* qui implique une réaction émotionnelle enthousiaste. Toutefois, l'insertion d'un adverbe intensifieur, par exemple *vraiment, trop* ou *super* aurait permis d'ajouter cette dimension expressive enthousiaste manquante[172]. Un autre exemple qui a retenu notre attention est la réponse proposée dans le dialogue 3 (indignation des deux étudiants) *c'est dégoutant* employée pour paraphraser le PhPex *c'est abusé !*. Cet emploi inhabituel nous semble être une confusion avec l'expression *être dégouté* dans son acception de dégout moral ou de déception (ex. : *je suis vraiment dégouté ! y'avait plus une seule place assise dans le TER !*). Dans la forme proposée par le LNN, l'acception qui semble primer est celle

[172] Cette insertion a d'ailleurs été fréquemment observée dans les réponses des LN.

liée au dégoût comme réaction physiologique. Il n'existe pas à notre connaissance d'emploi de la forme *c'est dégoutant* pour exprimer le dégout moral, la déception ou le désaccord comme le ferait *être dégouté*. Cette difficulté observée au niveau de la combinatoire a donc eu pour effet d'actualiser un sens non pertinent pour l'illocutoire visé.

La troisième difficulté concerne l'ajustement de la visée pragmatique et de l'illocutoire qui en découle. Si l'on reprend les cas cités concernant le dialogue 1 dans le paragraphe précédent (*c'est bien !, c'est magnifique !, c'est superbe !* et *c'est parfait*), nous avons pu observer une certaine incompatibilité de ces propositions de réponse pour restituer l'illocutoire de *c'est le pied !*. En effet, le schéma conversationnel proposé (réaction positive suite à une séance en institut de massage) implique une relation physique entre le locuteur et le déclencheur de la réaction expressive. Ce faisant, l'expérience du massage sollicite tout un ensemble de sensations internes à travers le canal physiologique et la perception sensorielle qui nécessite finalement la rencontre du corps avec le déclencheur. L'objectif de l'emploi du PhPex *c'est le pied !* est donc principalement d'exprimer une réaction émotionnelle par le biais physiologique et de susciter une réaction positive chez l'interlocuteur vis-à-vis d'une expérience vécue et ressentie. Dans le cas des expressions jugées non pertinentes, il s'agit d'évaluatifs qui ne peuvent à notre sens être associés à la perception corporelle (odorat, toucher, goût), mais à des voies perceptives indirectes et externes telles que la vue ou l'ouïe. Concernant les reformulations de l'expression *c'est pas possible !* nous observons chez les LNN plusieurs réponses autour du motif de l'incompréhension (*je ne comprends pas comment tu as pu le faire*, *je ne comprends pas ton comportement*, *comment ça se fait ?, je ne comprends pas pourquoi tu es tombée ici*, etc.), motif inexistant dans les réponses des LN[173]. Ce sens d'incompréhension nous semble peu compatible avec la reformulation de l'emploi du PhPex *c'est pas possible !*. En effet, il parait curieux que le locuteur manifeste son incompréhension ou demande à son interlocuteur de justifier ce qu'il s'est passé alors qu'il connait manifestement les tenants et les aboutissants de la situation (il en est lui-même témoin) puisque cette même situation est le déclencheur de sa réaction expressive.

---

[173] Il nous a semblé ici que cette visée pragmatique était un transfert de la LM. Toutefois, en l'absence de données contrastives, nous ne pouvons confirmer cette hypothèse.

La quatrième difficulté d'emploi remarquée dans les réponses des LNN concerne l'ajustement du contenu linguistique à la stratégie de communication attendue. Nous avons ainsi pu observer dans le deuxième dialogue de cette activité l'usage de réponses telles que *je ne peux pas y aller avec toi, cela ne me va pas* ou *j'ai pas de temps*. Ces réponses semblent être d'un point de vue fonctionnel en adéquation avec une situation de refus, mais présentent quelques problèmes d'ajustement sur le plan de la stratégie de communication liée à la formulation et à la construction de la justification. Dans le cas de la première expression (*je ne peux pas y aller avec toi*), cette impression d'emploi inhabituel vient du fait que la formulation est un peu ambigüe et laisse sous-entendre un refus directement lié à la présence de l'interlocuteur et non pas à une indisponibilité. Dans les deux autres cas présentés plus haut (*cela ne me va pas, j'ai pas de temps*), l'anomalie vient de la formulation trop abrupte du refus. En effet, le schéma conversationnel proposé implique deux amis qui discutent ensemble. De ce fait, il semble nécessaire de maintenir une relation interpersonnelle harmonieuse. Or, ces propositions de réponses semblent à la fois très catégoriques et ne rentrent pas dans le schéma pragmatique observable du dialogue 2. En observant ce dialogue, on se rend compte que la locutrice décline diligemment l'invitation de son interlocuteur en raison de son indisponibilité, mais propose tout de même de planifier la rencontre un autre jour. Le refus observable dans le dialogue n'est donc pas catégorique et implique la volonté du locuteur de concrétiser la rencontre dans d'autres conditions.

2.4. Synthèse intermédiaire de l'activité 2

L'analyse des données a permis de mettre en évidence plusieurs différences entre les productions des LN et celles des LNN. En effet, bien que globalement adaptées au schéma conversationnel proposé, les réponses des LNN étaient moins diversifiées à la fois au niveau des formes employées, de leur registre, mais aussi des stratégies communicatives exprimées. Nous avons également observé que l'emploi de formes plus spécifiques au registre familier, voire très familier, était beaucoup moins représenté chez les LNN que chez les LN. De plus, les dispositifs de modulations de l'intensité expressive (éléments périphériques, utilisation d'intensifieurs[174], termes vulgaires) étaient plus variés chez les LN que chez les LNN dont

[174] Concernant l'emploi des intensifieurs, nous avons pu constater que les LNN avait employé dans le premier dialogue à contexte positif significativement plus d'adverbe intensifieurs que dans toutes les données observées

on a pu observer que la maitrise de certaines nuances inhérentes aux PhPex proposés ne semblait pas encore acquise. Cette compétence lexicopragmatique des LNN qui semble dans les résultats observés déjà bien stabilisée ne semble pas permettre encore l'usage plus fin et nuancé de marques de modalisation comme nous avons pu l'observer chez les LN. Enfin, nous avons pu observer un décalage évocateur entre les taux de réussite des LNN (65% sur l'activité) qui sont plutôt moyens et leur taux de compréhension (88% sur l'activité) qui s'avère élevé. Cela pourrait être le signe d'une compétence en compréhension avancée, mais en revanche d'une palette lexicopragmatique et d'une compétence pragmalinguistique encore insuffisante. Nous rappelons toutefois que ces résultats sont à nuancer du fait que les réponses enregistrées ne répondaient pas toutes à la consigne formulée.

3. Analyse linguistique de l'activité 3 (équivalence fonctionnelle)

3.1. Cadre de l'activité 3

L'activité 3 est une activité d'équivalence dans laquelle il est demandé aux informateurs de produire un équivalent fonctionnel ayant les mêmes propriétés lexicopragmatiques que l'expression soulignée dans le dialogue proposé. Par exemple, pour le PhPex *quelle plaie !*, il s'agit de proposer un équivalent[175] tel que *quelle galère !*. Pour les LNN, il était aussi demandé de proposer un équivalent fonctionnel dans leur langue maternelle. Cette activité se fonde sur le principe d'un protocole de complétion par substitution combiné à une compréhension de la fonction illocutoire (Cadre théorique, Chap. 7, section 1). Toutefois, à la différence de l'activité 2, il n'y a pas d'élicitation explicite du sens ou de la fonction. L'objectif de ce troisième protocole exploratoire est de tester la compétence de compréhension et de production. En effet, l'informateur mobilise dans un premier temps sa capacité de compréhension du contexte (compétence sociopragmatique) et de l'expression à substituer (compétence pragmalinguistique) pour ensuite mobiliser sa compétence de production et proposer un équivalent fonctionnel au PhPex désigné dans le dialogue. À la différence de l'activité de reformulation, on attend ici la production d'une séquence

---

jusqu'à présent. On peut alors se demander s'il n'y aurait pas une plus grande facilité à utiliser ce mécanisme combinatoire dans les expressions utilisées en contexte positifs que dans celles utilisées en contexte négatif.

[175] Bien évidemment, il n'existe aucun « équivalent parfait » à telle ou telle expression, nous prendrons donc en compte les expressions qui remplissent à minima la fonction expressive et pragmatique de l'expression initiale.

préfabriquée à fonction expressive pouvant s'intégrer au schéma conversationnel proposé. Les dialogues proposés constituent à l'instar des précédentes activités des interactions orales quotidiennes et informelles, dans tous les cas entre amis. De même, les schémas conversationnels qui ont été utilisés sont diversifiés et mettent en scène des emplois expressifs. Comme dans les activités précédentes, le contexte d'interaction est précisé à la fois dans une petite didascalie ajoutée avant le dialogue et dans les informations extralinguistiques inférables dans le dialogue. Les informateurs ont aussi accès au paramètre intonatif et expressif dans un extrait sonore qui est fourni pour chaque dialogue.

3.2. Spécificités lexicales des réponses observées dans l'activité 3

			Variabilité Lexicale	Variabilité registre		Indice phraséologique	Acceptabilité réponse (taux de réussite)		Formes les plus fréquentes (FR puis JP)	Taux de compré.
LNN	D1	J' y crois pas !	FR : 37%	Courant	FR : 27%	FR : 88%	Acceptable	FR : 80%	1. C'est pas possible ! (15) 2. C'est incroyable ! (9) 3. Ça ne se fait pas ! (2)	94%
					JP : 10%			JP : 89%		
				Familier	FR : 70%		Inacceptable	FR : 17%		
			JP : 34%		JP : 90%	JP : 88%		JP : 8%	1. shinjirarenaï (14) 2. arienaï (13) 3. maji(de/ka) (5)	
				Très familier	FR : 3%		Sans réponse	FR : 3%		
					JP : 0%			JP : 3%		
	D2	C' est pas vrai !	FR : 42%	Courant	FR : 29%	FR : 88%	Acceptable	FR : 77%	1. C'est pas possible ! (6) 2. Sérieux ! (6) 3. Je n'y crois pas ! (3)	83%
					JP : 3%			JP : 89%		
				Familier	FR : 71%		Inacceptable	FR : 23%		
			JP : 14%		JP : 97%	JP : 100%		JP : 11%	1. uso(deshô) (13) 2. maji(de) (16) 3. hontô(ni) ? (8)	
				Très familier	FR : 0%		Sans réponse	FR : 0%		
					JP : 0%			JP : 0%		
		C' est chaud !	FR : 42%	Courant	FR : 17%	FR : 43%	Acceptable	FR : 37%	1. Ça craint ! (4) 2. C'est dur ! (3) 3. Oh la vache ! (2)	83%
					JP : 9%			JP : 71%		
				Familier	FR : 83%		Inacceptable	FR : 51%		
			JP : 28%		JP : 91%	JP : 57%		JP : 13%	1. yabaï (11) 2. taïhen da (10) 3. hidoï (2)	
				Très familier	FR : 0%		Sans réponse	FR : 12%		
					JP : 0%			JP : 15%		
	D3	C' est pas gagné !	FR : 60%	Courant	FR : 63%	FR : 26%	Acceptable	FR : 43%	1. Ça va être difficile ! (3) 2. Ce n'est pas sûr. (3)	91%
					JP : 11%			JP : 57%		
				Familier	FR : 37%		Inacceptable	FR : 57%		
			JP : 68%		JP : 89%	JP : 31%		JP : 43%	1. muri(da/deshô) (13) 2. dô kana (2)	
				Très familier	FR : 0%		Sans réponse	FR : 0%		
					JP : 0%			JP : 0%		
		Tu plaisantes ?	FR : 42%	Courant	FR : 17%	FR : 80%	Acceptable	FR : 74%	1. (Tu es) sérieux ? (5) 2. Tu rigoles ? (5) 3. C'est une blague ? (3)	93%
					JP : 6%			JP : 82%		
				Familier	FR : 80%		Inacceptable	FR : 23%		
			JP : 34%		JP : 94%	JP : 86%		JP : 18%	1. uso (deshô/darô) ? (9) 2. jôdan (deshô/darô) ? (17) 3. fuzaketeru ? (2)	
				Très familier	FR : 0%		Sans réponse	FR : 3%		
					JP : 0%			JP : 0%		

LN	**D1**	**J' y crois pas !**	51%	**Courant**	0%	100%	**Acceptable**	97%	1. Sérieux ? (14) 2. C'est pas vrai ! (6) 3. Mais non ! (5)
				Familier	100%		**Inacceptable**	3%	
				Très familier	0%		**Sans réponse**	0%	
	D2	**C' est pas vrai !**	57%	**Courant**	3%	94%	**Acceptable**	100%	1. Sérieux ? (12) 2. Mais non ! (7) 3. J'y crois pas ! (3)
				Familier	91%		**Inacceptable**	0%	
				Très familier	6%		**Sans réponse**	0%	
		C' est chaud !	49%	**Courant**	11%	74%	**Acceptable**	63%	1. Ça craint ! (8) 2. Dur ! (4) 3. C'est terrible ! (3)
				Familier	86%		**Inacceptable**	31%	
				Très familier	3%		**Sans réponse**	6%	
	D3	**C' est pas gagné !**	77%	**Courant**	12%	94%	**Acceptable**	94%	1. Ça va être compliqué. (5) 2. C'est pas sûr. (3) 3. C'est pas dit (2)
				Familier	88%		**Inacceptable**	6%	
				Très familier	0%		**Sans réponse**	0%	
		Tu plaisantes ?	34%	**Courant**	3%	100%	**Acceptable**	100%	1. C'est une blague ? (11) 2. (Tu es) Sérieux ? (18) 3. Tu déconnes ? (4)
				Familier	97%		**Inacceptable**	0%	
				Très familier	0%		**Sans réponse**	0%	

Tableau 17 : Spécificités lexicales des réponses analysées dans l'activité 3

Une des différences de ce tableau récapitulatif par rapport aux autres observées jusqu'à présent est qu'il prend aussi en compte les données produites en japonais dans le calcul des différents indices. Nous rappelons ici qu'en raison du fait que nous ne sommes pas locuteur natif du japonais, nous avons fait appel à 3 locuteurs natifs dans le cadre d'un accord interjuge pour valider l'identification de la variabilité du registre et de l'indice phraséologique (cf. Chap. 10, section 3).

La variabilité lexicale de l'activité 3 affiche en moyenne une différence entre les LN et les LNN de 15% pour les réponses en français. Nous observons ainsi une diversité des formes plus grande chez les LN que chez les LNN, mais cette tendance demeure toutefois relativement modeste. Si l'on regarde à présent la variabilité lexicale en français et en japonais dans le groupe des LNN, on ne remarque pas de différences significatives, à l'exception du deuxième dialogue. Le dialogue 2 se compose des expressions *c'est pas vrai !* (item 1) et *c'est chaud !* (item 2). L'item 1 affiche un taux de variabilité de 42% en français contre 14% en japonais pour un taux de réussite de 77% en français et de 89% en japonais. Les taux de réussite étant élevés, la grande différence de variabilité entre le français et japonais semble être le signe d'un consensus fort concernant les réponses données en japonais. En revanche, il semblerait que les réponses proposées en français soient plus

diversifiées. L'item 2 affiche un taux de variabilité de 42% en français contre 28% en japonais pour un taux de réussite de 37% en français et de 71% en japonais. Le taux de réussite en français est ici plus modeste (37%), ce qui peut indiquer que la diversité lexicale observée peut aussi être un signe de lacune lexicopragmatique en français. Le même phénomène semble se produire dans l'item 1 du dialogue 3 (*c'est pas gagné !*) pour lequel les données semblent indiquer un fort taux de variabilité pour un taux moyen de réussite. En parallèle, on constate également un très haut taux de variabilité chez les LN pour cet item, ce qui nous amène à formuler l'hypothèse d'une complexité inhérente à la formulation d'un équivalent pour *c'est pas gagné !*.

La variabilité du registre relevée dans cette activité montre d'une part une plus grande propension à l'usage du registre familier chez les LNN qui reste toutefois en moyenne 15% en dessous des ratios enregistrés pour les LN. D'autre part, nous observons que les réponses en japonais relèvent principalement du registre familier avec une différence assez significative avec les taux observés dans les réponses des LNN en français. En outre, le ratio des réponses relevant du registre familier observé pour les réponses en japonais est relativement proche de celui observé chez les LN. Cela nous amène à nous demander si les LNN ont une compréhension de l'ajustement du registre plus avancée que ce que nous avons pu constater jusqu'à présent sur les résultats en français. Cela pourrait étayer la possibilité d'une difficulté qui se trouve principalement en production et au niveau lexical en français, plutôt qu'en compréhension.

L'indice phraséologique des LNN est relativement élevé, que ce soit en japonais ou en français, mais demeure tout de même en dessous de l'indice des LN. On remarque deux items pour lesquels ces indices de phraséologisation sont très faibles, ce qui pourrait indiquer une palette lexicopragmatique en français et en japonais qui ne permet pas fournir d'équivalents. Le premier cas concerne l'item 2 du dialogue 2 *c'est chaud !* et le second cas concerne l'item 1 du dialogue 3 *c'est pas gagné !*. Dans les deux cas, nous avons une expression fortement idiomatique et non prédictible, ce qui pourrait expliquer le fait que les réponses proposées soient peu phraséologisées, car un certain nombre des LNN ne semble pas disposer d'équivalents fonctionnels à ces expressions dans leur répertoire lexical ou ne discerne pas assez bien le sens ou la fonction dont il est question pour proposer un équivalent viable.

3.3. Analyse des réponses de l'activité 3

3.3.1. Analyse comparative

La modalité contrastive de cette activité constitue une approche d'exploration de la compétence lexicopragmatique tout à fait intéressante. Elle a notamment permis de mettre en évidence d'une part certains des mécanismes de compréhension des LNN, ainsi que les stratégies pragmatiques employées pour répondre aux situations de communications proposées. Dans la lignée des activités précédentes, nous remarquons dans les deux cohortes une forte productivité du patron syntaxique [c'est + ADJ / N]. Cette forte productivité ne semble pas fortuite et mériterait d'être vérifiée sur d'autres corpus d'interaction quotidienne orale et/ou écrite (médiée par les réseaux sociaux). Les résultats analysés montrent une compréhension avancée de la situation de communication avec des taux de réussite et des taux de compréhension globalement haut chez les LNN, à l'exception des deux items précédemment mentionnés dans les spécificités lexicales que nous développerons plus bas dans l'analyse. Les expressions proposées par les LNN correspondent dans la plupart des cas analysés à une axiologie négative et semblent se conformer au contenu pragmatique attendu dans les schémas conversationnels proposés.

Dialogue 1

L'expression à reformuler est le PhPex *j'y crois pas !* (Krzyżanowska, Grossmann et Kwapisz-Osadnik 2021 : 386). Il est employé par le locuteur en tant que réaction émotionnelle face à une situation, une action ou un phénomène qui ont une influence négative sur le locuteur ou qui suscitent une surprise à polarité négative [176]. Le phraséologisme est employé le plus souvent pour exprimer son indignation et le fait de désapprouver l'élément déclencheur de la réaction expressive. Dans certains cas plus rares, l'expression peut être employée de manière à exprimer la satisfaction et la surprise à polarité positive (ex. : *J'y crois pas ! tu as réussi à réparer ma voiture ! t'es vraiment le meilleur !*). L'emploi de cette expression s'observe principalement dans un registre familier, mais peut dans une certaine mesure convenir dans un cadre un peu plus formel pourvu qu'il ait une certaine proximité relationnelle entre les interactants.

---

[176] Pour plus de détail, cf. Krzyżanowska, Grossmann et Kwapisz-Osadnik (2021 : 386)

De manière générale, les résultats enregistrés montrent que les LNN semblent avoir une compréhension plus ou moins fine du schéma conversationnel proposé. La totalité des réponses proposées semble s'orienter vers une polarité négative sur le plan sémantique, et la majeure partie de ces réponses constituent des phraséologismes à fonction expressive qui expriment bien l'étonnement et l'indignation face à la situation que dénonce le locuteur dans le dialogue. Nous remarquons que ce sont en majorité les expressions d'indignation *c'est incroyable !* et *c'est (ce n'est) pas possible* qui ont été majoritairement sélectionnées dans les réponses des LNN. Tout comme l'expression *j'y crois pas !*, nous retrouvons ici dans bon nombre des propositions le sens de remise en cause de la réalité reposant sur une modalité aléthique employée pour mettre en avant le caractère irréel et absurde de la situation rencontrée (ex. : *c'est pas possible !, c'est pas vrai, c'est incroyable !*, etc.). Nous observons également dans les résultats des LNN le motif de l'absurdité qui permet de remettre en cause le bien-fondé de l'évènement ayant déclenché la réaction expressive chez les informateurs (*c'est n'importe quoi !, c'est abusé !, ça ne se fait pas !*).

On retrouve chez les LN quelques réponses similaires à celles retrouvées chez les LNN, avec toutefois une plus grande diversité des formes et l'usage du registre familier plus fréquent. Les propositions d'équivalence relevées pour cette cohorte se fondent également sur un procédé aléthique de déni de la réalité à laquelle le locuteur est confronté pour exprimer une forte réaction émotionnelle. Ce procédé aléthique repose lui-même sur des stratégies discursives diverses parmi lesquelles le motif de la plaisanterie (*c'est une blague ?, tu plaisantes ?*), le déni de la réalité (*j'hallucine !, non, mais je rêve !, j'en reviens pas !, etc.*), le déni du propos de l'interlocuteur (*c'est pas vrai ?, sérieusement ?, pas croyable !*) et l'usage de la négation interjective (*mais non !*). Ce dernier procédé discursif est très productif chez les LN (beaucoup moins chez les LNN) et se manifeste par l'emploi du constituant périphérique interjectif *non,* constituant souvent précédé de la conjonction, *mais.* La conjonction, *mais* peut être employée à l'oral comme connecteur pour réagir à un propos ou prendre la parole (parfois même rebondir durant le tour de parole du locuteur) et introduit souvent un contenu expressif lorsque la prise de parole est réactive à un évènement causant une réaction émotionnelle chez le locuteur. L'emploi de la négation *non* permet d'exprimer une forte surprise qui repose sur un procédé discursif consistant à nier la réalité à laquelle le locuteur se confronte. En outre, nous observons aussi dans quelques réponses des LNN l'utilisation de ces constituants périphériques employés pour renforcer la dimension expressive ou interlocutive du PhPex. Il s'agit notamment des expressions *c'est fou* ***quoi*** *!*

et ***non*** *c'est pas vrai !*. Ces satellites agissent comme des particules discursives intersubjectives pour la première expression et marquant la négation interjective pour la seconde.

Concernant les équivalents proposés en japonais, la première chose à remarquer est que l'on retrouve les mêmes procédés aléthiques que dans les réponses en français des LNN et celles des LN. Cela est d'autant plus intéressant que ces procédés fonctionnent sur les mêmes motifs lexicaux qu'en français, c'est-à-dire le motif de l'impossibilité (*arienai = c'est pas possible !*), de la croyance (*shinjirarenai = j'y crois pas !*) et du doute (*maji de = sérieux ?, uso deshô = c'est une blague ?*). De même, on retrouve comme en français l'usage de constituants périphériques sous la forme de l'emploi de sous la forme de la particule finale à valeur d'affirmation du ressentiment *ne* (ex. : *arienai* ***ne*** *= c'est vraiment incroyable !*) ou de la particule finale à valeur expressive *yo* (ex. : *shinjirarenai* ***yo*** *= je peux pas le croire !*). De manière générale, on observe l'usage important de structures préfabriquées (indice phraséologique élevé => 88%) qui sont particulièrement bien adaptées au dialogue proposé.

Dialogue 2

Les deux expressions à reformuler sont les PhPex *c'est pas vrai !* et *c'est chaud !*. Le PhPex *c'est pas vrai !* est employé par le locuteur en tant que réaction d'étonnement face à un évènement ou un comportement auquel il est confronté[177]. Cet étonnement est souvent mêlé à de l'indignation, de la joie, ou de l'admiration en fonction de la situation dans laquelle va s'actualiser l'expression. On observe parfois des emplois exprimant également une forme de découragement et d'abandon de la part du locuteur. Cette expression s'emploie généralement dans un registre familier, mais n'exclut pas le registre un peu plus formel en fonction du degré de proximité entre le locuteur et son interlocuteur. Le PhPex *c'est chaud !* est employé par le locuteur afin d'exprimer une évaluation négative face à un évènement. Dans le cas du dialogue dont il est question, l'expression peut exprimer une certaine empathie à l'égard d'un évènement difficile ou jugé critique par le locuteur, mais qui ne le concerne pas directement. L'une des difficultés d'emploi de *c'est chaud !* réside dans sa forte polyfonctionnalité. En effet, le sens du PhPex peut s'actualiser de différentes manières en

[177] Pour plus de détail, cf. Krzyżanowska, Grossmann et Kwapisz-Osadnik (2021 : 282)

fonction du contexte[178]. Par exemple, on pourra tantôt observer le sens habituel de température élevée (*ah ! c'est chaud ! fais attention !*), tantôt le sens de danger (*wouah, C'est chaud ! j'ai cru qu'on allait y passer*), voire le sens de complexité (*c'est chaud ! je ne capte rien à cet exercice de math !*). Cette expression s'emploie exclusivement dans un registre familier avec de locuteurs étant plus ou moins proches sur le plan relationnel. En outre, il semblerait que l'emploi de ce PhPex soit plus fréquent chez une tranche de locuteur jeune que chez une tranche de locuteur un peu plus mature (ce qui n'exclut pas l'emploi chez la tranche mature). Cette complexité d'emploi liée à cette polyfonctionnalité à semble-t-il eu un effet autant sur les résultats des LN que des LNN.

Les équivalents proposés par les LNN pour *c'est pas vrai !* et *c'est chaud !* mettent en œuvre plusieurs types de visées illocutoires qui se regroupent en trois catégories :

- Les expressions se fondant sur un procédé aléthique (*sérieux ? , j'y crois pas !, c'est vrai ?*, etc.).

- Les expressions expriment une surprise à polarité négative (*ça me choque !*, *quoi ??*, *mais non !*, *(oh) la vache !*, etc.).

- Les expressions exprimant l'empathie (*quel malheur !*, *c'est horrible !*, *c'est dur !*).

Nous remarquons également dans les données relevées chez les LNN l'emploi de constituants périphériques leur permettant de renforcer l'illocutoire des expressions utilisées. Il s'agit notamment des interjections *oh* et *ah*, ainsi que l'interjection à valeur de négation *non.* L'adjonction d'interjections aux PhPex est un procédé relativement fréquent à l'oral et se retrouve fréquemment dans les productions des natifs. Les interjections étant par essence des termes à fonction expressive, leur ajout permet de renforcer l'expressivité et l'intention discursive véhiculée par le PhPex lors de l'emploi. En outre, la présence d'expressions suivant les patrons syntaxiques [c'est + ADJ / N] élidées au niveau du démonstratif (*c'est sérieux => sérieux ?, c'est incroyable => incroyable !*) constitue selon nous une trace de la systématisation chez certains des LNN de la construction des formes raccourcies. En effet,

---

[178] L'adjectif *chaud* employé en tant qu'attribut peut aussi renvoyer dans le sociolecte des jeunes / jeunes adultes au fait d'être motivé ou déterminé (« il est chaud pour le match de demain. » / « je suis pas très chaud pour rentrer ce week-end. »), ou alors renvoyer à une certaine lascivité (« il est chaud le gars là-bas, il passe son temps à courir après les filles au bar »).

l'emploi de ces formes tronquées est assez fréquent à l'oral familier. Elle nécessite une compréhension de la combinatoire interne et de la fonction de l'expression (compétence pragmalinguistique) pour pouvoir raccourcir l'expression sans en altérer son illocutoire ou sa force expressive.

Les équivalents proposés par les LN montrent une grande diversité des formes qui pour la plupart s'inscrivent dans un registre oral et familier. Comme nous nous y attendions, les LN ont une plus nette tendance à employer des séquences fortement idiomatiques (ex. : *oouah … dur…, quel coup dur !, ça craint !, quand même !*). Concernant les propositions d'équivalent pour *c'est pas vrai !,* nous avons pu observer la forte productivité du *non* interjectif*,* ou de la séquence, *mais non !.* Ces deux emplois ont été observés tantôt en autonomie, tantôt antéposés à un autre PhPex (ex. : ***non**, c'est une blague ?*). Dans ce dernier cas, nous pensons que l'emploi en antéposition à une fonction d'intensification expressive qui accentue le degré de surprise ou d'affect du PhPex qui suit. D'autres propositions d'équivalent reposent sur une stratégie de remise en cause du propos de l'interlocuteur avec une forte productivité des expressions basée sur l'adjectif sérieux (ex. : *non !! t'es pas sérieuse ?, non ! sérieux ?, wow sérieux ?*, etc.). On observe également des expressions marquant le doute lié à la grande surprise du locuteur (ex. : *je te crois pas, non ! vraiment ?*, etc.) ou basées sur le déni de la réalité (*C'est vrai ?, pour de vrai ?, mais non, c'est vrai ?, j'y crois pas !*, etc.). Nous remarquons l'emploi de la séquence *ah ouais ?*, emploi très intéressant, car reposant cette fois-ci sur une demande de ratification. Autre phénomène intéressant, nous avons pu observer dans les propositions d'équivalent relevées une expression d'empathie *oh merde ! le pauvre !* qui mélange à la fois un terme vulgaire qui a pour but ici de renforcer l'intensité expressive à un terme hypocoristique (*le pauvre*) permettant de mettre en valeur la dimension empathique.

Les propositions d'équivalence en japonais montrent une fois de plus que les LNN ont globalement saisi le type de réponse attendu dans le cadre du schéma interactionnel proposé avec une forte proportion d'emplois préfabriqués. Dans le cadre du premier item (*c'est pas vrai !*), les équivalents proposés se fondent sur le même procédé aléthique que le PhPex proposé dans le dialogue (ex. : *e, uso ! = quoi, c'est une blague ! / uso deshô !* = c'est une plaisanterie, pas vrai ? / *maji de ? = sérieux ? / maji ka yo ?* = est-ce que c'est sérieux ?, etc.). De même, dans les équivalents proposés pour le deuxième item (*c'est chaud !*), on remarque également une correspondance avec l'évaluation négative empathique (ex. : *sore*

wa yabaï ! = *ça craint !*, *sore wa taihen da ne* = *c'est dur !, n'est-ce pas*, etc.). On observe à nouveau dans ces quelques exemples l'usage de constituants périphériques facultatifs employés par le locuteur pour renforcer l'expressivité de son expression (particule finale *yo*, interjection placée en tête d'énoncé *e*). En outre, la construction démonstrative *sore wa* antéposée à *yabaï* ou *taihen* est considérée ici comme une interjection emphatique mettant en avant le propos du locuteur (Makino & Tsutsui 2013 : 434) et permet de renforcer la réaction expressive du locuteur. Il s'agit d'un procédé de renforcement de l'expression employée similaire aux constituants périphériques employés par les LN.

Dialogue 3

Les deux expressions à reformuler dans le cadre de ce dialogue sont les PhPex *c'est pas gagné !*[179] et *tu plaisantes ?*[180]. Le PhPex *c'est pas gagné !* est employé par le locuteur en tant que réaction évaluative permettant d'indiquer à l'interlocuteur que ce dont il est question a de très faibles chances de se concrétiser. Cette réaction dubitative se mêle bien souvent à de l'indignation et/ou de la déception en réaction à la désillusion provoquée par la non-réalisation possible de l'évènement déclencheur de la réaction expressive. Sur le plan discursif, l'illocutoire de ce PhPex repose sur une modalité épistémique visant exprimer le degré de certitude du locuteur sur un état des choses ou lui permettant de moduler le degré de certitude de son assertion. Le PhPex *tu plaisantes ?* est employé par le locuteur en réaction à un évènement déclencheur suscitant chez ce locuteur une forte surprise mêlée à de l'indignation ou de la réprobation. L'expression est parfois employée pour exprimer l'incrédulité ou le mépris. Mentionnons que cette expression est le plus souvent employée en réaction à un propos plutôt qu'à un évènement. Sur le plan discursif, l'expression se base sur une interrogation à valeur rhétorique, directement orientée vers l'interlocuteur et permettant d'exprimer une modalité aléthique.

Nous remarquons dans le cas de *C'est pas gagné !* que les réponses proposées par les LNN semblent indiquer une certaine difficulté à comprendre le sens ou l'emploi de cette expression. La séquence en elle-même présente une forte idiomaticité et ses fonctions pragmatiques apparaissent peu prédictibles. De plus, nous nous sommes aperçu lors de

---

[179] Pour plus de détail, cf. Krzyżanowska, Grossmann et Kwapisz-Osadnik (2021 : 224)
[180] Idem p. 527

discussions ultérieures avec des natifs japonais qu'il n'existait en japonais *à priori* pas d'équivalents traductionnels ou fonctionnels dont le sens et la structure lexicale soient proches de *c'est pas gagné !*. Cela semble confirmé par l'indice phraséologique très faible relevé pour cette expression (26% en FR, 31% en JP). Cette constatation explique d'une part la variété importante d'expressions que nous avons relevées (cf. variabilité lexicale forte : 60% FR, 68% JP), d'autre part, le caractère « paraphrasé » de nombre de réponses en français. En revanche, l'expression *tu plaisantes ?* a semble-t-il posé moins de problèmes aux LNN.

Concernant les équivalents produits en français pour l'expression *c'est pas gagné !,* nous avons pu dégager les trois paradigmes suivants dans les réponses des LNN:

- Les expressions construites autour de la notion de difficulté (ex. : *ça va être difficile !, ça va être dur, c'est chaud*, etc.)

- Les expressions à modulation du degré de certitude (ex. : *on est pas sûr, c'est improbable, c'est peu probable qu'il vienne, c'est presque mort*, etc.)

- Les expressions à modulation de l'assertion (ex. : *je ne pense pas, je crois pas*)

À l'instar des propositions d'équivalent en français, les propositions de réponses en japonais sont foisonnantes et peuvent être divisées en trois classes similaires aux classes évoquées précédemment :

- Les expressions construites autour de la notion de difficulté (*kibishii me da ne = la situation est complexe, zannen nagara muzukashii sô = malheureusement, ça à l'air compliqué*, etc.).

- Les expressions à modulation du degré de certitude (*mada wakannaï* = on ne sait pas encore*, hakkiri shinaï = ce n'est pas certain*, *kare wa muri -ppoi = on dirait que ça ne va pas être possible en ce qui le concerne*, etc.)

- Les expressions à modulation de l'assertion (*muri da to omou = je pense que ça va pas être possible, kibishii to omou = ça va être complexe je pense*, etc.).

Ces propositions de reformulation en japonais sont tout à fait intéressantes, car l'on retrouve des éléments linguistiques habituellement employés en français qui prennent la forme de constituants périphériques. Par exemple, nous observons :

- Des formes spécifiques à l'oral et relevant principalement du registre oral (verbe *savoir* à la forme négative *wakannaï*, adverbe conjectural *-ppoi*, particule finale de questionnement autoréflexif *ka na*).

- L'emploi de modulateurs assertifs verbaux (cf. Tuchais 2015, 2020 et la construction verbale *to omou*).

- L'emploi de particule énonciative permettant de renforcer l'expressivité (particule finale *yo*), L'emploi de marques modales permettant de moduler le degré de certitude (l'auxiliaire de conjecture *darô,* adverbe marquant le degré faible de certitude *sô*).

Contrairement aux résultats observés chez les LNN, nous observons ici que les LN ne montrent aucune difficulté particulière pour proposer des équivalents fonctionnels à l'expression *c'est pas gagné !*. On observe tout d'abord plusieurs expressions construites autour de l'adjectif *sûr* : *c'est loin d'être sûr, rien n'est sûr, c'est pas sûr*. Nous observons ensuite des propositions à modalité épistémique utilisant des verbes de pensée pour moduler une assertion dont on n'a pas la certitude (*je crois pas, je pense pas, à mon avis non*, etc.). À l'instar des résultats observés chez les LNN, nous avons remarqué dans cet inventaire très riche la présence d'expressions dont l'illocutoire se fonde sur le motif de la difficulté (*c'est chaud !, ça va être compliqué !, c'est tendu*, etc.). En outre, nous avons remarqué l'usage des expressions *c'est pas dit* et *c'est mal parti* qui semblent fonder leur sens sur l'idée d'un processus inachevé ou non réalisé qui par analogie évoque une faible probabilité du déroulement de l'action. Enfin, nous avons pu remarquer lors de notre analyse un certain nombre d'expressions familières à forte idiomaticité, sémantiquement opaques et/ou à forte charge culturelle (*ça craint, ça m'étonnerait, bon courage !, on n'est pas sortis de l'auberge, faut pas trop y compter, tu peux toujours courir, tu peux toujours rêver*).

Concernant les équivalents proposés par les LNN pour l'expression *tu plaisantes ?* deux paradigmes se dégagent : les expressions construites autour de la notion de *plaisanterie* (ex. : *non, tu rigoles ?, c'est une blague ?, tu déconnes*, etc.) et les expressions de remise en cause

de la parole de l'interlocuteur. Ce second paradigme se scinde en deux sous-ensembles, les expressions de remise en cause de la vérité (ex. : *non, c'est pas vrai !, c'est vrai ?, c'est pas possible*, etc.) et les expressions de remise en cause du sérieux de l'interlocuteur (ex. : *sérieusement ?, tu es sérieux ?*, etc.). Ces expressions se fondent sur la modalité aléthique pour exprimer une réaction émotionnelle telle que le locuteur va remettre en cause la véracité ou le sérieux de ce qu'il entend.

Les propositions d'équivalent en japonais pour *tu plaisantes ?* sont beaucoup moins diversifiées que pour le PhPex précédent. Nous observons dans les résultats un consensus au niveau des formes proposées qui s'avèrent pour la plupart très adaptées au contexte de communication proposé (faible variabilité lexicale [34%] , indice phraséologique élevé : [86%], taux de réussite élevé [82%]). D'autre part, il existe en japonais un équivalent structurellement quasi identique à *tu plaisantes ?*, l'expression *jôdan deshô / darô = ça doit être une blague*. À l'instar des résultats analysés en français, l'illocutoire des réponses proposées en japonais (*uso deshô / darô* ?= *c'est une blague ?*, *fuzaketeru = tu te fiches de moi ?*) se fonde là aussi sur une modalité aléthique mise au service de la mise en valeur de la réaction émotionnelle chez le locuteur.

Les propositions d'équivalence relevées chez les LN pour *Tu plaisantes ?* se caractérisent principalement par le recours à la modulation de la modalité aléthique, tout comme les LNN avec qui ils partagent bon nombre de propositions d'équivalent pour cet item. Nous avons tout d'abord pu observer tout un ensemble d'expressions tournant autour de la notion de *sérieux* (*sérieusement ?, non, sérieux ?, t'es pas sérieux !, non, mais sérieux ?, sérieux ?? il abuse…*, etc.). Notons également la présence abondante de constituants périphériques antéposés (*non, mais*). Comme évoqué précédemment, ces satellites contribuent à renforcer l'illocutoire de l'expression autour de laquelle ils gravitent. Tout comme les LNN, un certain nombre des propositions relevées étaient liées au motif de la plaisanterie (*c'est une blague ?, tu rigoles j'espère ?, tu te moques de moi là ?, tu déconnes ?*, etc.). Nous observons également la présence du terme *arrête* que l'on retrouve dans des expressions telles que *arrête de me charrier* ou *arrête ! tu me fais marcher !*, et fonctionne de manière autonome ou en tant que constituant périphérique. Enfin, nous mentionnons ici l'emploi des deux formes orales et populaires *tu te fous de moi* et *sans blague ?* qui constituent des formes fréquemment employées à l'oral informel et peu fréquentes dans les résultats des LNN (une seule occurrence de *sans blague ?*).

3.3.2. Analyse des difficultés observées dans l'activité 3

La première difficulté repérée dans les propositions de réponse des LNN concerne l'ajustement du registre. En effet, même si nous avons pu remarquer dans notre analyse une forte prépondérance des emplois relevant du registre familier, nous avons observé des réponses qui d'une part semblaient trop proches du standard pour le dialogue proposé. Nous relevons par exemple l'expression *c'est insensé !* employée en substitut de l'item 1 du dialogue 1 (*j'y crois pas !*). En outre, on s'aperçoit que sur le plan morphologique, l'usage de la double négation et des formes complètes est beaucoup plus productif chez les LNN que chez les LN. Nous observons par exemple l'emploi prépondérant de la forme *c'est pas possible !* chez les LN alors que les LNN privilégieront la forme *ce n'est pas possible !*. De même, nous avons retrouvé une prépondérance de la forme *j'y crois pas !* chez les LN alors que les LNN avaient plus tendance à employer la forme complète *je ne peux pas le croire*. L'abaissement du registre suppose un effacement de la distance relationnelle et interpersonnelle entre le locuteur et son interlocuteur (Krzyżanowska, Grossmann et Kwapisz-Osadnik 2021 : 27). Cet abaissement contribue selon nous à renforcer l'expressivité et l'illocutoire de l'expression. C'est notamment cet abaissement insuffisant du registre qui expliquerait en partie le fait que les formes standards proposées par les LNN peuvent paraitre peu naturelles dans certains dialogues.

La deuxième difficulté observée concerne l'ajustement de la fonction pragmatique et de l'expressivité. Nous avons pu remarquer dans nos données l'usage d'expression en français ou en japonais dont l'expressivité ne semblait pas correspondre à ce qui était attendu dans le dialogue. Par exemple, dans l'item 1 du dialogue 3, nous remarquons concernant l'expression *c'est pas gagné !* une certaine inadéquation du degré de certitude exprimé par certaines réponses. Le PhPex *c'est pas gagné !* sous-entend de faibles chances de concrétisation d'un évènement sans pour autant l'affirmer catégoriquement. Cette nuance laisse la possibilité, certes faible, à l'évènement dont il est question de se produire (ici la venue de l'ami pour le déménagement), d'où la modalité épistémique identifiable dans l'expression permettant d'exprimer un certain degré de certitude ou non. De même, le locuteur employant ce PhPex exprime le fait que lui-même ne sait pas si son ami viendra. Or, il nous a semblé que des propositions comme *c'est impossible*, *il peut pas* ou *il ne viendra pas* exprimaient non pas une modulation de la modalité épistémique, mais la certitude de la

non-venue de l'ami. Cette contradiction avec l'illocutoire attendu dans l'item 1 du dialogue 3 est à l'origine de l'inadéquation relevée.

La troisième difficulté observée concerne la combinatoire des expressions proposées et notamment la fonction pragmatique que cette combinatoire actualise dans le dialogue (compétence pragmalinguistique). Ainsi, un premier exemple de cette difficulté est l'emploi de *c'est étonnant !* comme substitut de l'expression *c'est pas vrai !*. En effet, *c'est étonnant !* marque la surprise neutre et parfois le fait d'être légèrement déconcerté ou décontenancé par un imprévu. L'emploi de cette expression peut aussi être ironique et exprimer le fait que l'on n'est en réalité pas vraiment surpris par un évènement anticipé au préalable. Or, le sens de *c'est pas vrai !* employé dans l'item 1 dans le dialogue 2 (collègue de classe accidenté) implique à la fois de l'étonnement à polarité négative avec la manifestation d'une certaine empathie. En outre, *c'est étonnant !* ne fonctionne pas sur une modalité aléthique, mais sur une modalité appréciative, ce qui diffère donc de l'expression de départ sur le plan pragmatique. Un autre exemple est la proposition d'équivalence *tu dis la blague !* proposée pour substituer l'expression *tu plaisantes ?*. Cette proposition comporte des anomalies à la fois sur le plan lexical, pragmatique et sémantique. Sur le plan lexical, il nous semble que la séquence *dire la blague* apparait moins naturelle que *dire une blague*, cette dernière séquence étant elle-même moins naturelle que la collocation *raconter une blague* qui permet de mettre en valeur toute la mise en scène et le jeu narratif inhérent à la blague. Outre ces problèmes de choix lexical, cette expression libre sur le plan combinatoire n'exprime pas un fort antagonisme ou de l'indignation comme ce serait le cas avec *tu plaisantes ?*, mais semble être un énoncé exclamatif qui semble plus se rapprocher d'une phrase déclarative dans sa formulation. Sur le plan pragmatique, on n'observe ni expressivité ni dimension discursive particulière. Nous pensons donc que le locuteur avait une hypothèse partielle de l'équivalent fonctionnel qui aurait pu convenir, mais ne disposait pas de la palette lexicopragmatique nécessaire pour formuler cette hypothèse. Un dernier exemple intéressant est la confusion observable sur la séquence *certainement pas !* proposée pour remplacer *c'est pas gagné !* dans le dialogue 3 (ami qui fait faux bond lors d'un déménagement). Il nous a semblé que l'informateur voulait exprimer le fait que l'ami ne viendra certainement pas en employant cette séquence. Toutefois, employée dans cette forme et de manière expressive, l'expression *certainement pas !* actualise une valeur de refus ou de rejet. Il aurait fallu préciser quelques éléments syntaxiques (ex. : *Il ne viendra certainement pas !*) pour actualiser un sens et une fonction hypothétique plus cohérents avec le dialogue proposé.

La quatrième difficulté repérée concerne l'ajustement de la stratégie de communication. Dans l'item 1 du dialogue 1 (*j'y crois pas !*), nous avons pu observer la proposition d'équivalence *c'est insupportable !* qui n'exprime pas la stupéfaction, mais un jugement de valeur subjectif reposant sur le caractère insupportable de la situation. De plus, l'emploi de cette forme laisse sous-entendre que le locuteur est lui-même victime de l'influence néfaste de cette situation, ce qui n'est pas le cas dans le schéma conversationnel proposé. En effet, dans le dialogue 1, le locuteur réagit avec surprise et indignation face à la situation difficile que vit l'une de ses connaissances. Une supposition est que l'informateur a voulu employer cette expression sous la forme *ça doit être insupportable* afin d'exprimer de l'empathie envers l'amie du locuteur qui vit une situation peu enviable, mais ne maitrisant pas l'utilisation de la conjecture dans cette séquence, cet informateur pourrait avoir formulé son énoncé de manière affirmative.

La cinquième difficulté constatée concerne les contresens liés à des erreurs au niveau formel. En effet, nous avons pu observer dans les résultats tout un ensemble d'expressions présentant des anomalies sur le plan grammatical ou sémantique qui pour une grande partie étaient liées à une non-compréhension du PhPex proposé ou à la proposition de réponses issues d'une traduction directe depuis les réponses proposées en japonais. Un exemple de ce type de difficulté est la proposition de reformulation *je n'en pense pas* proposée dans l'item 1 du dialogue 1 comme équivalent de l'expression *j'y crois pas !.* Cette proposition semble indiquer que l'informateur ne semble pas avoir bien compris le sens phraséologique ni la fonction expressive du PhPex. En effet, la proposition de réponse *je n'en pense pas* est employée ici pour rendre compte du sens littéral de l'expression *j'y crois pas !* à savoir l'incrédulité ou le fait de ne pas être d'accord. Cela se remarque d'autant plus dans la proposition d'équivalent en japonais de cet informateur (*boku wa sô omowanai keredo*). Cet énoncé traduit littéralement veut dire 'je ne pense pas cela cependant'. De même, l'usage de la modalité déclarative de cette réponse ne semble pas être en adéquation avec la dimension expressive de l'expression *j'y crois pas !.* Une autre hypothèse qui a émergé lors de l'analyse est que le locuteur a pu procédé par traduction en retranscrivant mot à mot dans un premier temps *j'y crois pas !* en japonais puis en proposant une expression en français proche de la traduction en japonais sur le plan lexical (soit : *j'y crois pas ! -> boku wa sô omowanai keredo -> Je n'en pense pas*). Il est intéressant de noter que l'informateur (MN33FJP062020, cf. Annexe 6) est de niveau C2 et est titulaire d'un doctorat en sciences du langage. Il a

appris le français pendant plus de 10 ans et a vécu en France pendant plus de 8 ans. Malgré son profil de locuteur expert dans la langue cible, nous avons constaté que les difficultés lexicopragmatiques demeurent visibles lors du test. Cette constatation constitue un argument supplémentaire en faveur de notre hypothèse que l'exposition prolongée ou le haut niveau d'expertise en LC ne sont pas les seuls facteurs de développement de la compétence lexicopragmatique.

3.4. Conclusion intermédiaire de l'activité 3

À l'instar des LN, nous avons pu observer chez les LNN un large panel d'expressions et de stratégies pragmatiques déployées pour exprimer l'illocutoire attendu. Les expressions employées par les LN semblaient toutefois plus diversifiées et se composaient de quelques expressions très idiomatiques que nous n'avons pas retrouvées chez les LNN. Il est tout de même important de noter que les LNN ont fait preuve d'une grande créativité phraséologique et ont montré qu'ils avaient un niveau de maitrise avancé des schémas conversationnels proposés. Cependant, il nous a semblé que les LNN éprouvaient certaines difficultés à ajuster le registre et le niveau d'expressivité de leur production pour s'adapter pleinement aux spécificités des dialogues proposés. En effet, la palette lexicale dont disposent les LNN est certes diversifiée, mais reste tout de même limitée pour ce qui est du lexique spécifique à l'interaction quotidienne. Il en va de même pour l'expression d'affects complexes qui passent avant tout par une compréhension approfondie des dynamiques socioculturelles et interactionnelles qui sous-tendent la conversation quotidienne.

En outre, notre analyse des difficultés a permis de mettre au jour un certain nombre d'erreurs récurrentes dans les activités de production de ce test. Il s'agit notamment d'erreurs liées à la compétence combinatoire, avec souvent des erreurs lexicopragmatiques liées à des associations lexicales peu pertinentes ou qui actualisent dans le schéma conversationnel proposé une fonction pragmatique non pertinente. Nous avons pu également observer que les LNN avaient une tendance à tantôt proposer des énoncés déclaratifs et compositionnels, tantôt traduire directement la proposition d'équivalent du japonais vers le français ou vice-versa. Ces tendances constituent des indices des stratégies déployées lorsque les LNN ne disposent pas de la connaissance linguistique ou des ressources lexicales nécessaires pour

répondre de manière adéquate. En observant les scores de réussite, et notamment les appariements possibles entre les scores de réussite en français et en japonais par locuteur, nous avons pu obtenir la modélisation suivante :

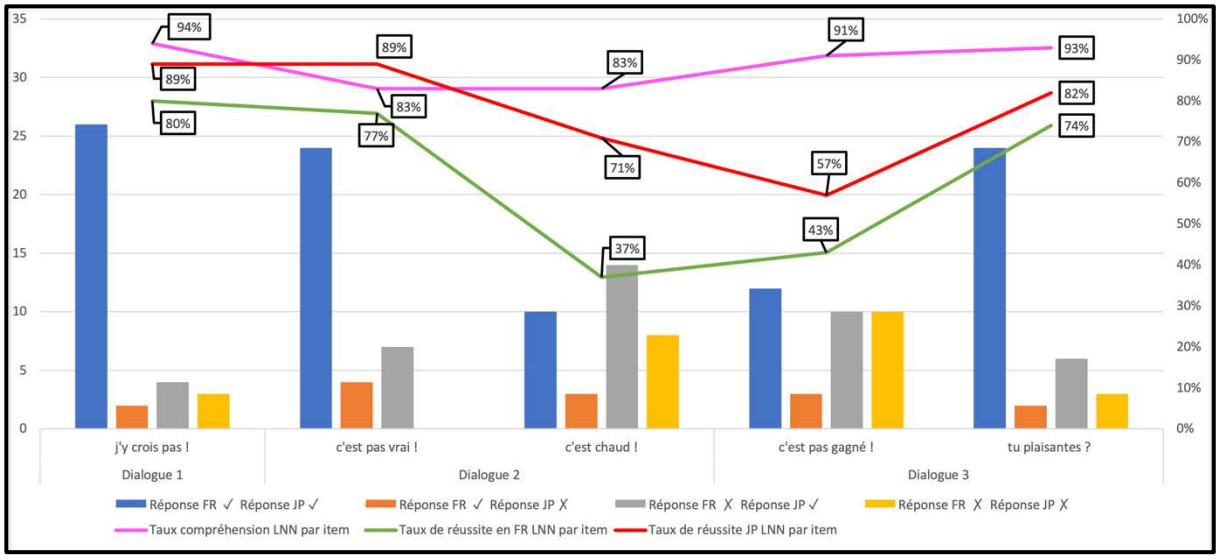


Figure 54 : Répartition des LNN en fonction du score de réussite obtenu dans l'activité 3 sur les propositions d'équivalence FR / JP avec mise en parallèle du taux de compréhension

Chaque colonne du diagramme représente une catégorie fondée sur l'appariement entre l'acceptabilité de la réponse en français et en japonais. La courbe rose correspond au taux de compréhension des LNN enregistré pour chaque item. La courbe verte correspond au taux de réussite des LNN pour les réponses en français et pour chaque item. La courbe rouge correspond au taux de réussite pour les réponses en japonais par item. Dans les deux premiers items (*j'y crois pas !* et *c'est pas vrai !*) et le dernier item de l'activité 3 (tu plaisantes ?), on observe une faible dispersion des taux de réussite entre les différents appariements. En outre, les valeurs les plus élevées s'observent dans la colonne bleue qui correspond aux informateurs ayant proposé une réponse adéquate en français et en japonais. En observant les courbes, on s'aperçoit que les taux de compréhension et de réussite en français et en japonais pour ces trois items sont également élevés (<75%). Ce type de modélisation avec une faible dispersion entre les différents appariements et des valeurs élevées dans les courbes semble être caractéristique des PhPex où les LNN ont rencontré peu de difficultés que ce soit en compréhension et en production.

En revanche, les deux PhPex restants (*c'est chaud !* et *c'est pas gagné !*) montrent une tendance inverse. En effet, on observe pour ces deux PhPex une forte dispersion entre les

différents appariements dont plus de la moitié de l'effectif total se trouve dans les appariements négatifs. En parallèle, on observe des valeurs relativement basses pour les courbes des taux de réussite et une différence très visible entre le taux de compréhension et le taux de réussite. Deux hypothèses qui ont émergé dans l'analyse qualitative sont alors confirmées :

1. Les LNN ont une compréhension élevée du contexte du dialogue et du type d'expression attendu (taux de compréhension haut), mais ne possèdent pas la palette lexicopragmatique suffisante pour répondre de manière adéquate (taux de réussite en baisse et dispersion des appariements forte).

2. Il ne semble pas exister d'équivalent fonctionnel en japonais, ce qui complique la tâche d'interprétation et de substitution en français (score faible des réponses en japonais)

L'hypothèse 2 a d'abord été confirmée par les statistiques lexicales qui montraient que pour ces deux items en particulier, on observait un indice phraséologique faible que ce soit en français ou en japonais, corrélé à une variabilité lexicale élevée qui indique ici un fort taux de réponses non-préfabriquées. L'analyse qualitative a ensuite montré les difficultés rencontrées par les LNN pour produire des équivalents fonctionnels adéquats.

4. Analyse linguistique de l'activité 4 (QCM lexical)

4.1. Cadre de l'activité 4

L'activité 4 est une activité de complétion à choix multiple où il était demandé aux informateurs de compléter un dialogue en choisissant parmi un inventaire de quatre PhPex l'expression idoine au schéma conversationnel proposé. Il s'agit alors d'analyser les spécificités du contexte de communication du dialogue afin de pouvoir déterminer laquelle des quatre propositions est la plus appropriée. Cette activité s'inspire des questionnaires à choix multiple d'élicitation de l'illocutoire et des tests de complétion (Cadre théorique, Chap. 7, section 1). L'objectif de cette quatrième activité est de tester la compétence de compréhension des PhPex proposés (compétence pragmalinguistique), mais aussi du contexte dans lequel s'insèrent ces expressions (compétence sociopragmatique). Ce

protocole exploratoire permet donc d'observer à la fois les phénomènes de macrocompréhension (contexte, relations entre les participants, thème de la conversation, enjeux, etc.) et de microcompréhension (expressions employées, prosodie, modalisation, visée pragmatique, visée discursive, etc.). Il est important de noter que cette activité nécessite cependant un effort de concentration plus important que les autres activités puisqu'elle implique pour l'informateur de devoir réaliser une tâche coûteuse sur le plan cognitif : a) discriminer l'expression idoine parmi les autres propositions tout en b) identifiant le sens et la fonction de chaque proposition puis en c) sélectionnant celle qui présente un haut degré d'adéquation avec le schéma interactionnel proposé. À cela s'ajoute d) une opération de traduction et de reconstitution sémantique et lexicale pour les LNN. À l'instar des activités précédentes, nous avons proposé des dialogues mettant en scène des interactions orales quotidiennes diverses, avec à chaque fois des didascalies insérées en début de dialogue pour situer le contexte du dialogue et la relation entre les participants, ainsi que le thème de l'interaction. Les informateurs avaient également accès à la prosodie et au contexte de l'interaction à l'aide d'un enregistrement audio fourni pour chaque dialogue. Parmi les suggestions de PhPex, nous avons volontairement inséré des distracteurs afin d'éviter les réponses aléatoires et de voir si les LNN parvenaient à différencier le sens littéral du sens phraséologique.

4.2. Analyse des réponses de l'activité 4

4.2.1. Dialogue 1

Les quatre suggestions de réponse étaient les PhPex *c'est génial !, c'est le pied !, c'est le bouquet !* et *c'est une horreur !*. La première expression est employée par le locuteur afin d'exprimer une grande satisfaction ou son admiration face à un évènement ou une action auquel il est confronté dans la situation d'énonciation. *C'est le pied !* que nous avons décrit dans la section dédiée à la première activité est employée par le locuteur pour exprimer sa satisfaction ou le sentiment positif ressenti après l'expérience d'une situation agréable. L'expression *c'est le bouquet !* permet au locuteur d'exprimer son indignation et sa désillusion dans une situation qui tend à s'envenimer. Ce PhPex permet également d'exprimer la plainte et le reproche. Enfin, l'expression *c'est une horreur !* permet d'exprimer un jugement appréciatif très négatif en réaction à une situation ou un évènement

à polarité négative. La réponse attendue pour ce dialogue était *c'est le bouquet !*[181]. En effet, cette expression est employée suite à une situation préjudiciable pour le locuteur qui dure depuis quelque temps et qui a tendance à se compliquer. Ce faisant, la locutrice qui est excédée par le mutisme croissant du garçon qu'elle a rencontré va utiliser cette expression pour se plaindre de cette situation qui l'indigne. L'expression *c'est une horreur !* qui aurait pu constituer un candidat possible ne semble pas fonctionner ici, car ce PhPex s'emploie plus généralement dans des contextes où le locuteur se retrouve confronté à une situation négative dès le départ, et non pas confronté à un évènement dont l'ampleur a progressé par étape jusqu'à atteindre un niveau préjudiciable comme ce serait le cas avec *c'est le bouquet !*. Les deux expressions restantes sont incompatibles, car elles expriment en principe une réaction émotionnelle plus ou moins intense en réponse à un évènement positif.

Concernant maintenant les résultats, nous pouvons observer que les LN ont obtenu un score de réussite de 57% alors que les LNN ont obtenu un score de 17 %. (avec un taux de compréhension de 67%). La première chose qui semble attirer notre attention est le score très faible relevé, tant pour les LN que pour les LNN. Les deux cohortes semblent s'être heurtées à un problème lors de la passation du test. Voici un diagramme récapitulatif du pourcentage de réponse par expression :

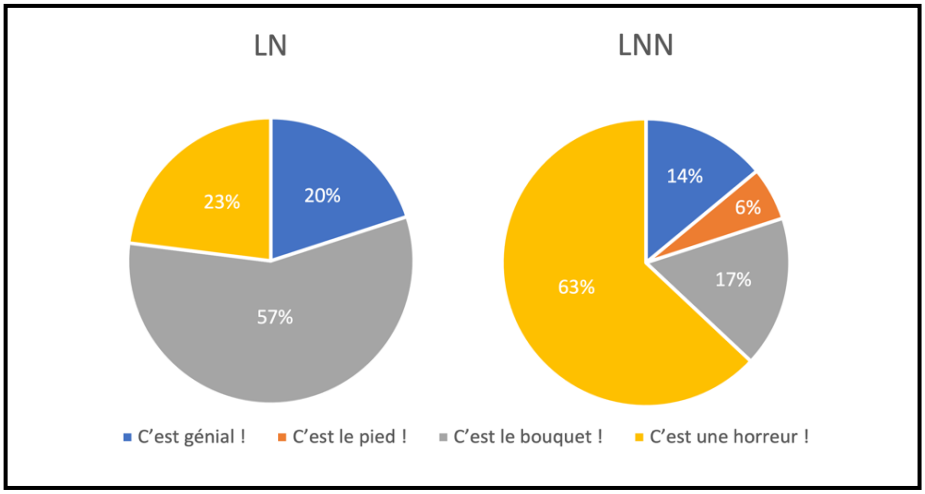


Figure 55 : Distribution des choix de réponse des LN et des LNN entre les 4 items proposés dans le dialogue 1[182]

[181] Lors d'une relecture de ce manuscrit, Agnès Tutin (LIDILEM) a suggéré que *c'est le bouquet !* était une expression en perte de vitesse, notamment chez les locuteurs jeunes. Cette faible représentation dans l'usage actuel peut être à l'origine d'une méconnaissance de cette expression chez les LNN.
[182] N = 35 informateurs dans les deux cohortes.

Nous pouvons remarquer dans ce graphique que l'expression attendue lors de la conception de cette activité n'est pas celle qui a été majoritairement sélectionnée, tant chez les LN que chez les LNN. Concernant les choix privilégiés par les LN, nous observons que *c'est le bouquet* a été sélectionné par 57% des informateurs de la cohorte. En revanche, le reste des locuteurs se partagent entre *c'est génial !* (20%) et *c'est une horreur !* (23%). En observant les résultats des LNN, on s'aperçoit là aussi que *c'est le bouquet !* n'est pas la réponse la plus représentée (17%). Une majorité des réponses s'est orientée vers le PhPex *C'est une horreur !* (63%) et le reste des informateurs s'est réparti entre *c'est génial !* (14%) et *c'est le pied !* (6%). Ces résultats très surprenants, notamment en ce qui concerne les réponses des LN, nous ont amenés à remarquer plusieurs choses. D'une part, l'expression *c'est une horreur !* ne semblait pas si improbable dans le contexte proposé bien que n'étant pas la réponse la plus appropriée. D'autre part, nous nous sommes aperçu que les locuteurs ayant choisi le PhPex *c'est génial !* ont en réalité pensé à un emploi que nous n'avions pas anticipé lors de la conception de ce test, à savoir l'emploi ironique de ce PhPex. Cela a été confirmé à plusieurs reprises lors des entretiens post test qui a également montré que *c'est le bouquet !* était généralement écarté d'office par les LNN en raison de son opacité sémantique. Concernant le choix massif de *c'est une horreur !* par les LNN, il nous a semblé que la première raison justifiant le choix est que l'expression de base ne présente pas une grande opacité sémantique. Cette expression ne comportant aucune ambigüité au niveau du lexique, nous émettons ici l'hypothèse que les LNN, procédant par élimination, ont choisi l'expression négative la plus accessible. Nous pouvons également observé que le PhPex *c'est le pied !* n'a été sélectionné par aucun des LN alors que 6% des LNN ont choisi cette réponse erronée. Mentionnons que l'accord interjuge a montré que le PhPex *c'est le bouquet !* semblait être l'expression la plus appropriée avec 5 choix sur 6 en faveur de cette dernière. Nous avons donc pris le parti de compter comme inadéquat les autres réponses, mais les résultats ainsi obtenus pour ce dialogue restent à « prendre avec des pincettes », et en connaissance de cause étant donné le biais méthodologique mentionné plus haut. Ce biais semble difficilement évitable du fait de la complexité des expressions étudiées et de leur potentialité d'interprétation.

4.2.2. Dialogue 2

Les quatre suggestions de réponse étaient les PhPex : *c'est génial !, c'est pas* gagné *!, ça tombe bien !, c'est une horreur !.* La première expression est employée par le locuteur afin d'exprimer une grande satisfaction ou son admiration. La seconde expression (cf. activité 3) permet au locuteur d'exprimer son incrédulité et sa déception face à un évènement dont ignore s'il va prendre la tournure escomptée. L'expression *ça tombe bien !* permet au locuteur d'exprimer une surprise à polarité positive et une certaine satisfaction face à une action, un évènement ou un état des choses qui se déroule comme prévu ou dont il en retire un bénéfice au moment le plus opportun. La quatrième expression *c'est une horreur !* a été traitée plus haut et permet au locuteur d'exprimer son indignation reposant sur une appréciation négative portant sur un élément de la situation d'énonciation.

La réponse attendue pour ce dialogue était *c'est une horreur !.* Le schéma conversationnel proposé implique un contexte d'interaction ou le locuteur réagit face à une situation injuste et qui semble avoir des retombées négatives sur ses conditions de travail. Ce faisant il exprime une appréciation négative du préjudice vécu afin d'exprimer son indignation et le fait qu'il désapprouve l'attitude des fumeurs dans son entourage professionnel. L'expression *c'est pas gagné !* ne semble pas convenir ici, car elle nécessite que le locuteur ait des attentes particulières qui ne semblent pas se réaliser tout en exprimant son incrédulité face aux chances que l'évènement attendu se produise. Or, le locuteur ici ne nourrit pas d'attente particulière envers les fumeurs, mais semble juste vouloir montrer son indignation en exprimant un jugement négatif à l'égard de leur comportement. Les deux expressions restantes permettent d'exprimer une appréciation positive et ne constituent donc pas des choix viables dans le contexte proposé.

Les scores de réussite à ce dialogue sont de 51% pour les LNN et de 89% pour les LN (avec un score de compréhension de 76%). Une fois de plus, les LNN semblent avoir éprouvé des difficultés à d'une part comprendre pleinement la situation de communication proposée, d'autre part à sélectionner le PhPex le plus approprié au contexte visé. Les LN quant à eux ont un score qui ne semble pas indiquer de grandes difficultés, mais on constate tout de même quelques confusions avec les autres expressions proposées. Observons les résultats un peu plus en détail :

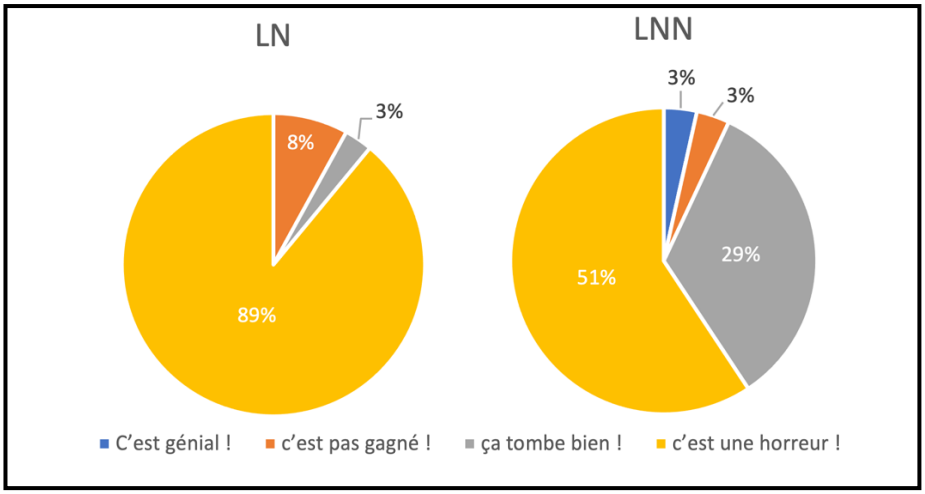


Figure 56 : Distribution des choix de réponse des LN et des LNN entre les 4 items proposés dans le dialogue 2

Une grande majorité des choix enregistrés se sont portés sur l'expression *c'est une horreur !*. Nous constatons cependant que certains des informateurs ont opté pour d'autres choix de réponse jugés moins pertinents. Une fois de plus, nous pensons que cette légère dispersion des réponses est non seulement due à la relative complexité de la tâche demandée, mais aussi au fait que *c'est pas gagné !* peut être viable si l'on considère que le locuteur a des attentes particulières et espère qu'un jour, les fumeurs auront un comportement différent, mais reste dubitatif quant à la réalisation de cette attente. Toutefois, si l'on observe l'intégralité du dialogue, il n'est mentionné nulle part ce cas de figure et la configuration de la conversation suggère ici plutôt une réaction appréciative immédiate et l'expression de l'indignation. Pour ce qui est des résultats des LNN, les ratios indiquent ici une forte dispersion avec seulement la moitié de la cohorte ayant choisi la réponse attendue. De manière surprenante, de nombreux informateurs ont opté pour *ça tombe bien !*. Après réflexion, il nous a semblé qu'un emploi ironique ne pouvait être envisageable ici puisque la situation n'implique pas que le locuteur bénéficie en bien ou en mal d'un évènement dont la mise en œuvre aurait pu être jugée comme opportune ou non. Notre intuition a été validée lors de l'accord interjuge au cours duquel cette expression a été jugée inadéquate à l'unanimité. Le troisième choix inadéquat ayant obtenu le plus de réponses est *c'est pas gagné !* et n'est pas acceptable pour les raisons susmentionnées. Enfin, la réponse restante *c'est génial !,* même dans son acception ironique ne semble pas être viable dans le schéma conversationnel proposé.

4.2.3. Dialogue 3

Les quatre réponses proposées étaient : *tu crains !, tu parles !, t'inquiète !, tu l'as dit !.* L'expression *tu crains !* est employée par le locuteur pour exprimer son désaccord et une appréciation négative à l'encontre de son interlocuteur qui a réalisé une action ou tenu un propos que le locuteur ne cautionne pas ou qui n'est pas approprié. La seconde expression permet au locuteur de ce dialogue d'exprimer son incrédulité et son désaccord en réaction au propos de son interlocuteur. La troisième expression possède deux emplois différents. Le premier emploi constitue une réaction d'empathie visant à réconforter et soutenir son interlocuteur face à une difficulté ou en proie au doute. La deuxième acception renvoie à un terme de validation permettant de répondre de manière positive à l'interlocuteur. Enfin, la quatrième expression *tu l'as dit !* est généralement employée par le locuteur pour exprimer un accord fort et enthousiaste vis-à-vis du propos de l'interlocuteur qu'il cautionne complètement.

La réponse attendue pour ce dialogue était le PhPex *tu parles !.* En effet, dans le schéma conversationnel proposé, le locuteur exprime clairement son désaccord et son doute face au propos de son ami. Il montre ainsi que les propositions du maire n'ont jamais été mises en œuvre, ce qui suscite chez lui un fort sentiment de reproche pour ce maire. L'expression *tu crains !* ne peut pas fonctionner ici étant donné qu'elle exprime une réprobation forte envers l'interlocuteur. Or, dans le cas présent, il s'agit de la déclaration du maire qui est désapprouvée et non pas l'interlocuteur qui n'a fait que rapporter cette déclaration. Les deux autres expressions ne semblent pas pertinentes de par leur polarité positive qui ne convient pas à la situation de communication présentée. Par ailleurs, il n'est question ici ni de rassurer l'interlocuteur ni de ratifier son propos, mais au contraire de s'indigner et d'exprimer son désaccord.

Le score de réussite pour ce dernier dialogue est de 54% pour les LNN (avec un score de compréhension de 71%) et de 94% pour les LN. À l'instar des précédents dialogues, les LNN enregistrent un score assez faible alors que les LN enregistrent un score qui demeure élevé (exception faite du premier, dialogue ou le score n'est pas haut, mais reste tout de même largement en deçà du score des LNN). Observons à présent les ratios par réponse pour les deux cohortes :

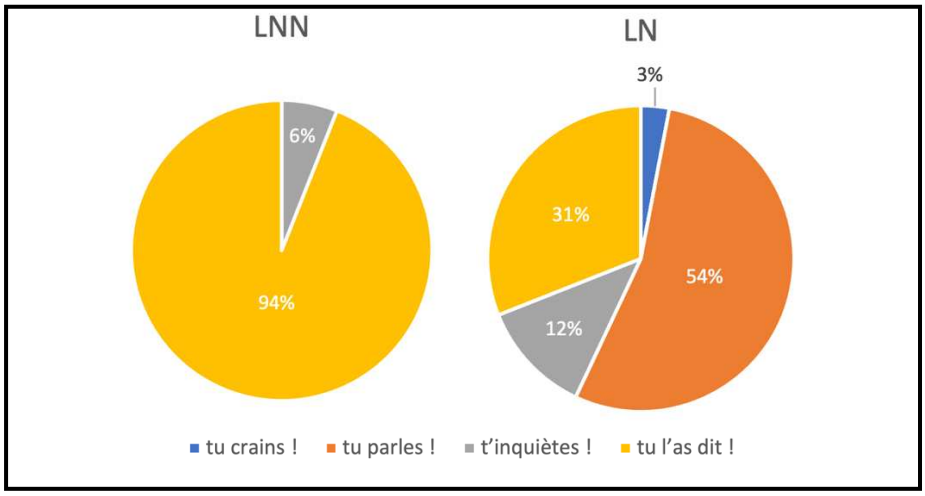


Figure 57 : Distribution des choix de réponse des LN et des LNN entre les 4 items proposés dans le dialogue 3

Nous pouvons observer que le taux de réussite des réponses des LN reste homogène avec seulement 6% des réponses inadéquates enregistrées. En revanche, les résultats des LNN montrent une forte dispersion entre les réponses. Un peu plus de la moitié ont proposé la réponse qui était attendue pour 1/3 des répondants ayant répondu *tu l'as dit !* dont l'emploi semble ici constituer un contresens et 12% pour *t'inquiète !*. Nous remarquons en outre qu'à la fois les LN et LNN semblent avoir choisi la réponse inadéquate *t'inquiète.* Cela est probablement lié au deuxième sens de l'expression qui permet de valider le propos de l'interlocuteur. En y réfléchissant un peu plus, on s'aperçoit que la confusion peut avoir porté sur un sens ironique visant non pas rassurer ou à acquiescer, mais à sous-entendre que le pire reste à venir et qu'il serait au contraire plus judicieux de s'inquiéter.

4.3. Synthèse intermédiaire de l'activité 4

Les résultats observés pour l'activité 4 ont été très surprenants du fait de l'hétérogénéité des réponses remarquées à la fois chez les LLN, mais aussi chez les LN. Alors qu'il nous a semblé que ce format présentait la plus grande facilité de passation lors de l'élaboration de notre test, nous nous sommes finalement aperçus en analysant ces résultats que cette activité comportait plusieurs difficultés de taille.

La première difficulté est la charge cognitive induite (surtout pour les LNN) par cette activité qui a été confirmée par l'étude de Wang (2021). En effet, le fait que le taux de réponses inadéquates sur le premier dialogue soit à la fois élevé chez les LN et les LNN nous a fait

prendre conscience que cette activité demandait à la fois de comprendre le contexte ainsi que ses spécificités et contraintes, de connaitre chaque expression ou à défaut, de tenter d'en inférer le sens et enfin de comparer chaque expression au contexte pour sélectionner l'expression idoine. Dans le cas des non-natifs, il nous a semblé que ceux-ci procédaient par élimination ce qui rallongeait d'autant plus l'effort cognitif nécessaire pour finir cette activité.

La seconde difficulté vient du fait que les PhPex utilisés dans le test sont parfois très polyfonctionnels. En effet, certains emplois ironiques nous sont passés inaperçus lors de la conception de notre test. C'était notamment le cas pour les expressions *c'est génial !* et *t'inquiète !*. Dans ces différents cas, nous nous sommes aperçus que les informateurs avaient su déceler cet emploi alors que nous avions en tête l'emploi pragmatique habituel lors de l'élaboration du test. Cela constitue en revanche des traces intéressantes de la compétence métapragmatique des informateurs qui ont su juger de l'acceptabilité d'une remotivation du sens des PhPex *c'est génial !* et *t'inquiète !* tout en évaluant la viabilité de ces nouveaux emplois en contexte. Une amélioration possible si ce test venait à être répliqué serait de prononcer les quatre expressions en exprimant clairement dans l'enregistrement audio l'intonation caractéristique de chaque emploi pour désambigüiser les expressions polyfonctionnelles.

La dernière difficulté vient du fait que les quatre expressions proposées à chaque dialogue présentaient quelques proximités d'emploi, ce qui a pu compliquer le choix de la réponse adéquate. Ces ressemblances fonctionnelles ont dû poser d'autant plus de problèmes aux LNN que les expressions utilisées dans cette activité sont relativement opaques et ne font pas partie pour la plupart des expressions idiomatiques enseignées fréquemment en classe de FLE. Il aurait été souhaitable selon nous de proposer 4 expressions avec peu de similitudes pour diminuer la charge cognitive et permettre aux LNN de procéder plus facilement par élimination en fonction de leur perception (même minimale) du degré d'adéquation entre l'expression et le contexte proposé. Voici à présent un schéma récapitulatif des différents cas de figure observés dans les réponses des LNN formés à partir de l'appariement du taux de réussite avec l'autoévaluation pour chaque item. Ces

appariements sont présentés avec les courbes du score de réussite et du taux de compréhension[183] :

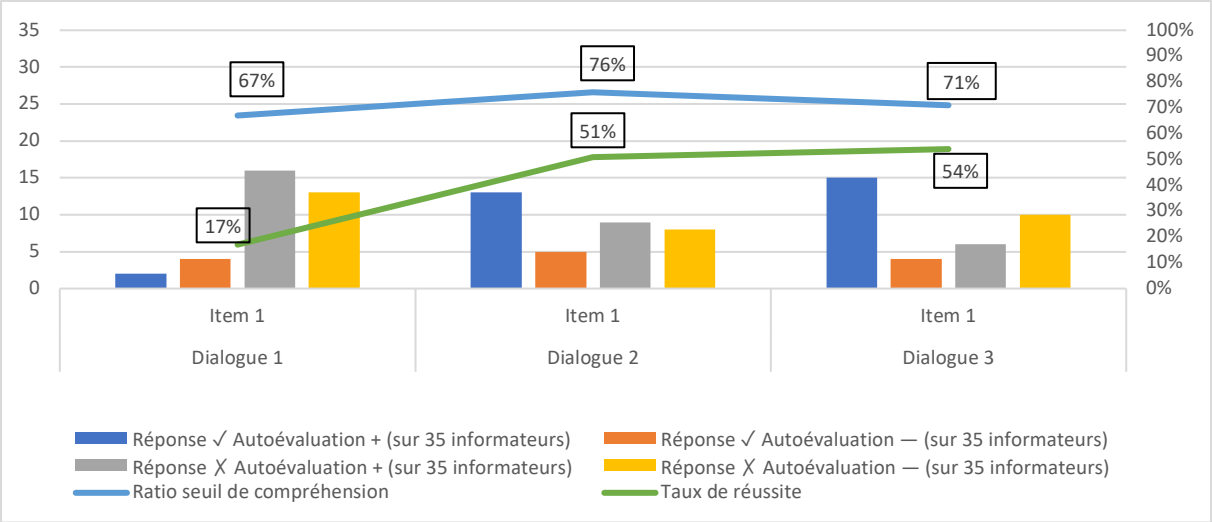


Figure 58 : Répartition des LNN en fonction de l'appariement de l'exactitude de la réponse et du type d'autoévaluation pour chaque dialogue avec mise en parallèle du taux de compréhension et du score général

On remarque une certaine prépondérance dans les trois dialogues de deux cas de figure : 1) le cas d'une réponse inadéquate avec une autoévaluation qui suggère une compréhension avancée du dialogue et des expressions proposées (en gris) et 2) le cas d'une réponse inadéquate avec une autoévaluation qui suggère l'incompréhension du dialogue et des expressions proposées (en jaune). La distance entre les deux courbes montre que les LNN semblent comprendre le contexte et avoir une intuition du lexique approprié, mais rencontre cependant des difficultés à sélectionner le PhPex le plus pertinent parmi ceux proposés. On remarque également une grande dispersion des effectifs entre les différents appariements, ce qui reflète la dispersion des réponses observée lors du commentaire des résultats proposé plus haut. Par ailleurs, le premier dialogue est celui pour lequel les LNN semblent avoir rencontré le plus de difficultés, comme on peut l'observer sur les taux de réussite très bas. Les deux autres dialogues ont en revanche des indicateurs plus haut, mais montrent toutefois que moins de la moitié de l'effectif total est parvenu à sélectionner la réponse tout en ayant un taux de compréhension élevé.

[183] Nous rappelons que le taux de réussite a été calculé à partir de l'accord interjuge réalisé lors de l'évaluation des réponses et que le taux de compréhension a été élaboré sur la base des questions d'autoévaluation proposée pour chaque item aux LNN. Pour plus de détails, consulter le Chap. 10, section 3.

5. Analyse linguistique de l'activité 5 (QCM sémantique)

5.1. Cadre de l'activité 5

L'activité 5 est une activité de reconnaissance à choix multiples où il est demandé aux informateurs pour chaque expression soulignée dans le dialogue proposé de choisir parmi un inventaire de trois ou quatre propositions de sens celle qui correspond selon eux au sens le plus approprié. Cette dernière activité s'inspire des protocoles de type questionnaire à choix multiple permettant de tester la compréhension de l'illocutoire véhiculé (Cadre théorique, Chap. 7, section 1). Il s'agit dans cette activité d'analyser les spécificités des PhPex proposés, leur sens et le contexte dans lequel ils sont employés pour ensuite sélectionner la proposition de sens la plus appropriée. L'objectif de cette cinquième activité est de tester la compétence de compréhension des PhPex. Cette activité fait appel à la fois à la compétence de compréhension du contexte (compétence sociopragmatique), mais aussi de l'expression proposée (compétence pragmalinguistique). Dans la dynamique des activités précédentes, nous avons proposé des dialogues mettant en avant des interactions orales quotidiennes spontanées. De même, nous avons ici aussi pris soin de préciser le contexte du dialogue au moyen de didascalies et d'indices contextuels disséminés dans le dialogue. Enfin, tout comme dans les activités précédentes, les répondants avaient accès à un enregistrement audio fourni pour chaque dialogue. Parmi les propositions de sens, nous avons volontairement inséré des distracteurs afin d'éviter les réponses aléatoires et afin de voir si les LNN parvenaient à différencier le sens littéral du sens phraséologique (compétence pragmalinguistique et métapragmatique).

5.2. Analyse des réponses de l'activité 5

5.2.1. Dialogue 1

Les deux expressions identifiées dans le dialogue sont *t'es sérieux ?* et *c'est pas possible.* Le PhPex *t'es sérieux ?* est employé par le locuteur afin d'exprimer sa surprise négative et son exaspération face à un propos ou une action réalisée par l'interlocuteur qui va déclencher chez le locuteur une réaction émotionnelle. L'expression fonctionne de la même manière que *tu plaisantes ?* en jouant sur la remise en cause du propos ou d'une action de l'interlocuteur pour exprimer le fait que ce propos ou cette action suscite une telle

indignation ou surprise que le locuteur demande confirmation à l'interlocuteur. La deuxième séquence est un distracteur ayant un sens littéral, c'est-à-dire marquant l'impossibilité de réaliser une action. Ci-dessous, les propositions de sens données dans l'activité pour chaque expression (réponse adéquate en gras) :

T'es sérieux ?	Sens 1	Eric demande à Pierre si c'est une personne sérieuse.
	Sens 2	**Eric est exaspéré par ce que Pierre vient de dire.**
	Sens 3	Eric demande à Pierre si ce qu'il vient de dire est vrai.
C'est pas possible.	**Sens 1**	**Eric pense que rendre le dossier dans ces délais est impossible.**
	Sens 2	Eric est exaspéré par ces délais injustes.
	Sens 3	Eric est surpris par ces délais injustes.

Tableau 18 : Choix de réponse proposés pour les deux expressions testées dans le dialogue 1

Dans ce tableau, nous pouvons observer les différents sens proposés aux répondants de notre test. Il est important de noter que nous avons travaillé ces intitulés de manière à ce que la difficulté soit bien dosée, et ce faisant nous avons essayé de proposer des intitulés peu ambigus, mais présentant toutefois une certaine proximité, et ce de manière à tester la finesse du niveau de compréhension chez les informateurs. Concernant les résultats à ce dialogue, les LLN ont obtenu 63% (score de compréhension : 91%) pour la première expression et 60% (score de compréhension : 94%) de réussite pour la seconde contre 60% pour la première expression et 69% pour la seconde chez les LN :

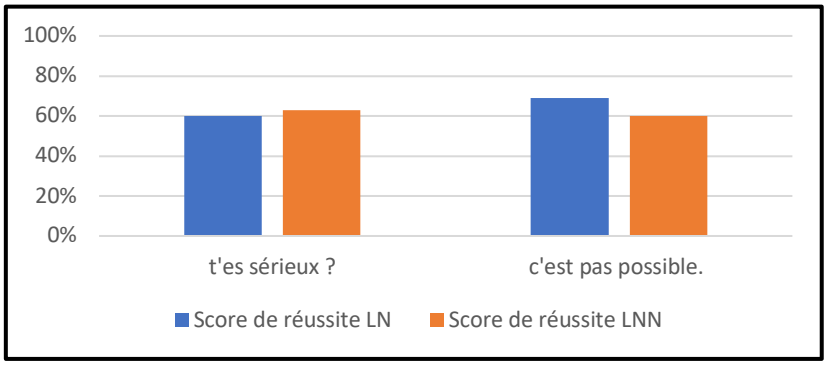


Figure 59 : Score de réussite par cohorte pour chaque item dans le dialogue 1

Ce score est tout à fait surprenant puisque le degré de réussite est quasi identique dans les deux cohortes. À l'instar de l'activité précédente, il semble donc il y a avoir un biais externe aux résultats relevés pour cette question. Le premier biais qui nous vient à l'esprit serait une

possible ambigüité dans les propositions de sens ou le dialogue en lui-même. Toutefois, après réflexion croisée avec d'autres locuteurs natifs, il semblerait que les intitulés ne présentent pas d'ambigüité particulière, de même que le dialogue. Reste alors un deuxième biais qui pourrait être la conséquence de la fatigue liée à la longueur du test qui a engendré un effet de précipitation dans les réponses au test. Observons le taux de réponse par proposition de sens :

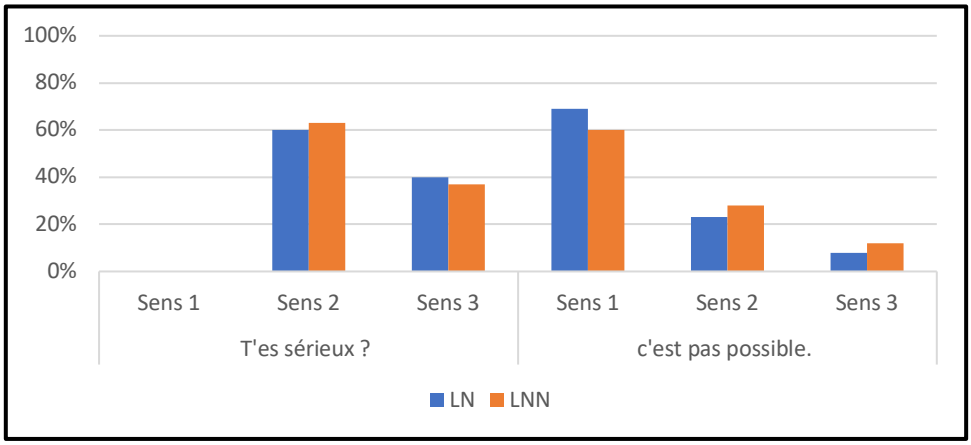


Figure 60 : Répartition des choix de réponse par cohorte dans le dialogue 1

Nous remarquons que les proportions de réponses concernant chacune des propositions de sens semblent plus ou moins similaires entre les deux cohortes. Concernant la première expression, nous remarquons que les LN et les LNN ont hésité sur les mêmes propositions, à savoir la seconde et la troisième proposition. Il nous semble ici qu'il y a une confusion entre le procédé sémantique sur lequel se fonde l'illocutoire de l'expression et son sens. Effectivement, l'emploi de *t'es sérieux ?* fonctionne sur un procédé de remise en cause du propos de l'interlocuteur par le locuteur qui feint de ne pas croire son interlocuteur, mais le sens et la fonction de ce PhPex expriment en réalité l'exaspération et la forte surprise à polarité négative face à une nouvelle qui le contrarie. La fonction première de l'expression n'est donc pas de demander si l'interlocuteur dit la vérité ou qu'il est sérieux dans ces propos, mais d'exprimer une surprise et une contrariété telle que le locuteur en vient à questionner métaphoriquement le bien-fondé du propos de son interlocuteur. Si l'objectif principal du locuteur était effectivement de demander à son interlocuteur s'il dit la vérité, nous pensons que ce locuteur aurait employé une expression plus explicite demandant directement à l'interlocuteur de confirmer son dire.

Pour ce qui est de la seconde expression qui constituait un distracteur dans ce dialogue, la confusion a principalement porté sur la première et la deuxième proposition de sens. Là aussi, il nous semble qu'il y a eu un phénomène de surinterprétation du sens et de la fonction de la séquence proposée. Le dialogue proposé avait été spécialement conçu pour que la prosodie de *c'est pas possible* soit assertive, avec peu voire pas d'expressivité, et ce de manière à éviter toute ambigüité avec le sens de stupéfaction et d'indignation du PhPex observé dans les autres activités. Ce faisant, l'écoute de l'audio montre très clairement que *c'est pas possible* est employé ici dans le sens d'incapacité à réaliser une action. Néanmoins, les informateurs ont pour certains interprété un sens d'indignation du fait de la situation d'incapacité des deux amis à rendre le travail demandé à temps en raison des délais intenables imposés par l'enseignant. C'est effectivement ici l'une des conséquences possibles de la situation dans laquelle se trouve les deux amis, mais ce n'est pas le sens exprimé par *c'est pas possible* qui se limite au marquage de l'incapacité. L'accord interjuge a montré peu de discordances quant au choix du sens idoine. On observe une moyenne de 5 juges sur 6 validant la réponse attendue pour ce premier dialogue.

5.2.2. Dialogue 2

Les expressions proposées dans ce dernier dialogue sont : *J'y crois pas !, les bras m'en tombent ! c'est une honte !* et *c'est le pied.* Cette dernière séquence constitue un distracteur mis en œuvre en écho au PhPex *c'est le pied !* employé plus tôt dans le premier dialogue de l'activité 2 (reformulation).

Le PhPex *j'y crois pas !* que nous avons développé plus haut dans l'activité 3 (équivalence) est employé par le locuteur afin d'exprimer son indignation et sa réprobation vis-à-vis d'un évènement qui suscite chez lui une forte réaction émotionnelle. La seconde expression *Les bras m'en tombent !* est employée par le locuteur pour exprimer sa forte stupéfaction en réaction à un évènement ou un propos inopiné qui suscite son indignation, qu'il désapprouve ou qu'il juge condamnable. L'expression *c'est une honte !* est employée de manière réactive par le locuteur face à un évènement qui suscite son indignation. L'emploi de cette expression est particulièrement prégnant dans des situations où le locuteur se trouve face à une injustice ou un évènement qui semble entrer en contradiction avec les règles conventionnellement instituées dans son groupe social. Enfin, *c'est le pied* est une séquence compositionnelle

désignant de manière référentielle le pied de l'un des locuteurs du dialogue. Voici un tableau reprenant les propositions de sens pour chaque expression (réponse attendue en gras) :

J'y crois pas !	Sens 1	Éric ne croit pas ce que Paul lui raconte.
	Sens 2	**Éric est exaspéré et étonné par ce que Paul lui raconte.**
	Sens 3	Éric est d'accord avec ce que Paul lui raconte.
Les bras m'en tombent !	Sens 1	Eric est motivé.
	Sens 2	**Éric est surpris et exprime son incompréhension.**
	Sens 3	Les bras d'Éric sont tombés.
C'est une honte !	**Sens 1**	**Éric est en colère et n'est pas d'accord avec la situation.**
	Sens 2	Éric a honte de lui-même.
	Sens 3	Éric est étonné par la situation.
C'est le pied.	Sens 1	Éric pense que les démangeaisons sont agréables.
	Sens 2	**Éric parle de ses pieds.**
	Sens 3	Éric est à pied pour rentrer chez lui.

Tableau 19 : Choix de réponse proposés pour les expressions testées dans le dialogue 2

Tout comme pour le dialogue précédent, nous avons pris soin de construire les items proposés de manière à ce que la difficulté soit bien dosée et que ces propositions de sens ne présentent pas d'ambigüités susceptibles de complexifier la tâche d'analyse et de discrimination des informateurs. Voici à présent un diagramme reprenant le score de réussite globale pour chaque réponse et par cohorte :

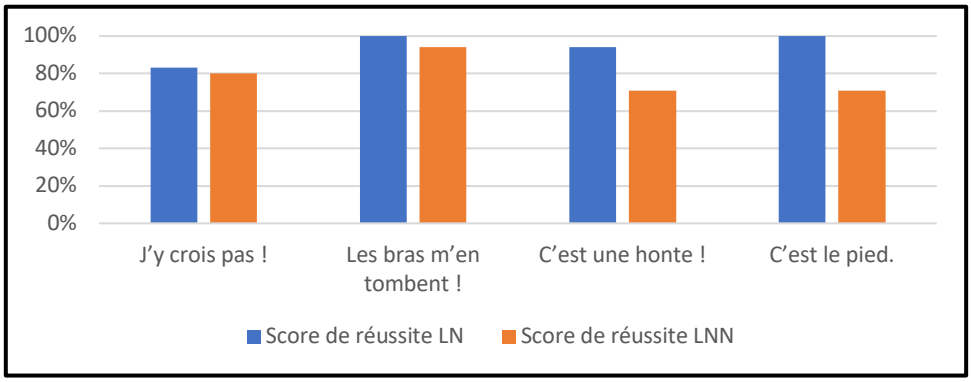


Figure 61 : Score de réussite par cohorte pour chaque item dans le dialogue 2

Ce tableau montre que le taux de réussite à ce dialogue est plutôt élevé avec un taux supérieur à 70% dans les deux cohortes et en moyenne un taux de compréhension pour les LNN de 87%. Il semble donc que les informateurs n'aient pas rencontré de difficultés particulières quant à la compréhension du contexte ou à l'identification du sens des expressions à définir. Regardons la répartition des choix de réponse par cohorte :

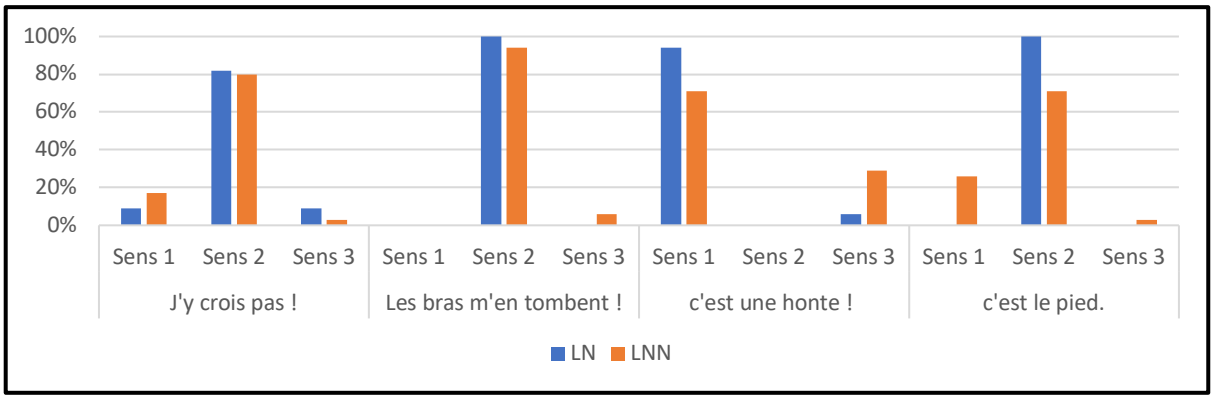


Figure 62 : Répartition des choix de réponse par cohorte dans le dialogue 2

La répartition observée des réponses pour chaque cohorte nous montre plusieurs choses intéressantes. Nous remarquons en premier lieu que les propositions de sens de la première expression ont suscité des choix plutôt disparates avec 9% des réponses des LN attribuées au sens 1 et au sens 3 alors que la réponse attendue était la proposition 2. En effet, dans le dialogue étudié, *j'y crois pas !* était employé comme structure d'indignation et de stupéfaction à polarité négative. Les deux autres propositions de sens impliquaient tantôt l'incrédulité tantôt l'accord. Le problème d'interprétation qui s'est posé ici selon nous vient d'un effet de précipitation chez les LN ou de non-maitrise de l'expression visée chez les LNN. En effet, il nous a semblé que le premier sens concernant l'incrédulité du locuteur ne peut pas fonctionner ici, car le problème des travaux est un problème connu et accepté des deux personnes et c'est justement pour marquer son étonnement et son indignation face au fait que ces travaux interminables nuisent toujours plus à la circulation des usagers que le locuteur emploie ce PhPex. Le troisième sens quant à lui n'est pas non plus viable, car il sous-entend l'approbation du locuteur quant à la situation actuelle. Toutefois, il nous semble qu'il y a peut-être eu une surinterprétation de cette approbation. Ce faisant les informateurs ont probablement imaginé que le locuteur était d'accord sur le fait que la situation était préjudiciable. Le problème est que cette interprétation ne porte alors plus sur l'expression, mais sur les locuteurs, ce qui sort du cadre de la consigne de l'activité.

La deuxième expression *les bras m'en tombent* n'a pas semblé avoir prêté à confusion, les deux cohortes ayant un taux très élevé de réussite. Nous remarquons toutefois que 6% des LN ont sélectionné l'interprétation littérale du PhPex, ce qui est dû selon nous à la méconnaissance de l'expression ou à une traduction « au pied de la lettre ». Une recherche

rapide sur le métacorpus Orfeo[184] a montré une faible productivité de l'expression (1 occurrence) ce qui peut être le signe d'une expression un peu vieillie, ou tout du moins en perte de vitesse à l'oral[185].

C'est une honte ! a semble-t-il posé plus de problèmes aux LNN avec une confusion retrouvée sur le sens d'étonnement. Cependant, le propre de ce PhPex n'est pas d'exprimer la stupéfaction, mais d'exprimer l'indignation tout en portant un jugement appréciatif à polarité négative sur un élément de la situation ou un propos qui suscite la réaction du locuteur. L'élément de surprise peut certes être un corrélat de cette réaction expressive, mais ne constitue pas un élément définitoire du sens ou de l'emploi de ce PhPex.

Enfin, nous pouvons observer que les résultats obtenus sur le distracteur sont tout à fait intéressants. En effet, bien qu'il ne semble y avoir aucune ambigüité quant à la nature dénotative de cette séquence (LN ont répondu avec un score de 100% et l'intonation de l'extrait audio ne laissait aucun doute), 29% des LNN n'ont pas choisi le sens adéquat dont 26% ayant choisi le sens phraséologique observé dans l'activité 2 (expérience du massage) et 3% ayant choisi le contresens. Les choix orientés vers le sens phraséologique montrent ici que le distracteur a bien fait son office, suggérant ainsi que les LNN se sont appuyé sur le sens observé dans le premier dialogue de l'activité 2 pour sélectionner le sens dans ce dialogue. Il serait judicieux de penser ici que les LLN ayant choisi cette alternative ne disposent pas vraiment d'une connaissance approfondie de cette expression, mais semblent avoir basé leur jugement sur des paramètres externes liés au contexte et aux emplois précédemment rencontrés dans le test. Cela suppose aussi qu'il y a une stratégie tout à fait intéressante d'adaptabilité fondée sur l'inférence et la déduction de l'emploi de l'expression inconnue à partir des données contextuelles et conversationnelles disponibles, mais aussi à partir des données expérientielles acquises lors du test. Il y a donc une posture d'analyse métapragmatique qui semble émerger de ces résultats.

---

[184] http://ortolang107.inist.fr/?locale=fr

[185] Nous avons cependant remarqué que sur Twitter, l'expression semblait encore très vivace, avec notamment trois occurrences par heure en moyenne (site consulté le 21/06/2022 à 12h42). Cela dit, le format écrit peut avoir favorisé l'usage de cette tournure, les tweets étant moins spontané et plus réfléchis que la conversation orale.

5.3. Synthèse intermédiaire activité 5

Les résultats enregistrés pour l'activité 5 ont montré un certain nombre de faits intéressants. En premier lieu, nous avons pu observer que la compétence de compréhension du locuteur non natif pouvait s'appuyer à la fois sur le contenu linguistique de l'expression analysée et sur les différents éléments contextuels qui constituent la situation d'interaction. Les résultats ont aussi montré que les locuteurs avaient parfois certaines difficultés à détacher le sens de l'expression de ses effets sur la situation d'énonciation et sur les locuteurs, c'est-à-dire de différencier l'illocutoire (visée actionnelle du PhPex) du perlocutoire (influence du PhPex sur l'interlocuteur).

L'usage du distracteur nous a permis de constater que certains des LNN qui ne connaissaient pas l'expression *c'est le pied !* s'étaient basés sur les éléments d'emplois constatés dans le premier dialogue de l'activité 2 pour tenter de sélectionner la bonne proposition de sens, ce qui les a induits en erreur, mais qui montre également qu'il y a toute une réflexion métapragmatique inconsciente qui s'articule autour de la prise en compte des informations lexicopragmatiques précédemment acquises. Dans la section qui suit, nous allons synthétiser de manière globale les résultats observés pour la deuxième phase de notre travail de recherche, puis nous passerons à la troisième phase qui consistera à apporter un éclairage du point de vue des informateurs sur les réponses analysées dans ce chapitre à l'aide d'entretiens post-test.

6. Synthèse des résultats du test lexicopragmatique

Afin de donner un premier aperçu synthétique des résultats, voici un tableau de synthèse des différentes statistiques lexicales mesurées sur les données des LN et des LNN :

	Variabilité Lexicale	Variabilité registre		Indice phraséologique	Acceptabilité réponse (taux de réussite)		Taux de compréhension
LNN	FR : 46%	**Courant**	FR : 41%	FR : 56%	**Acceptable**	FR : 67%	86%
			JP : 6%			JP : 77%	
		Familier	FR : 56%		**Inacceptable**	FR : 28%	
	JP : 35%		JP : 88%	JP : 72%		JP : 17%	
		Très familier	FR : 3%		**Sans réponse**	FR : 5%	
			JP : 0%			JP : 6%	
LN	54%	**Courant**	14%	89%	**Acceptable**	95%	
		Familier	81%		**Inacceptable**	4%	
		Très familier	5%		**Sans réponse**	1%	

Tableau 20 : Synthèse générale des statistiques lexicales mesurées sur l'ensemble du test

Bardovi-Harlig (2009) a montré que le LNN peut très bien reconnaitre et comprendre des phraséologismes pragmatiques sans pour autant pouvoir répliquer leur emploi de manière effective en situation de communication. De manière similaire, nos données ont montré que les LNN observés avaient une compréhension avancée des schémas conversationnels proposés, mais ne disposaient pas de la palette lexicale ou de la connaissance métapragmatique nécessaire pour produire l'expression attendue dans certains des dialogues proposés. Ce décalage entre compréhension et production s'observe notamment en comparant le taux de compréhension global du test chez les LNN qui est haut (86%) avec le taux de réussite général au test qui est plus faible (67%). L'indice phraséologique et la variabilité lexicales nous permettent d'affiner cette observation en montrant que 1) la variabilité lexicale est plutôt élevée (46%), 2) le taux d'usage des séquences préfabriquées est en revanche moyen. D'après l'analyse qualitative, nous avons pu constater que cette variabilité lexicale élevée était parfois liée à un faible consensus dans le lexique choisi qui lui-même traduisait des difficultés à proposer une séquence lexicopragmatique adéquate. Par ailleurs, ce faible consensus était en général corrélé à un indice phraséologique bas, ce qui accrédite d'autant plus l'hypothèse d'une difficulté lexicopragmatique au niveau de la production et non pas au niveau de la compréhension. Cette hypothèse est d'ailleurs confirmée par les résultats en japonais obtenus dans l'activité 3 (équivalence) qui montre une forte similarité des statistiques lexicales (notamment au niveau du registre, de l'indice phraséologique et du taux de réussite) et des stratégies pragmatiques employées par les LNN avec celles des LN. Les difficultés principales se trouveraient donc en production, au niveau du choix du lexique, de sa combinatoire et de la compréhension de l'illocutoire que peut actualiser ce lexique en contexte. Nous allons affiner cette vue globale des résultats dans la section suivante en observant une synthèse des caractéristiques et des difficultés associées à la variété non native observée dans le cadre de ce test lexicopragmatique.

6.1. Typologie des difficultés des LNN observées dans le test lexicopragmatique

L'analyse des résultats obtenus à l'issue de la passation du test lexicopragmatique a permis d'observer des traces linguistiques concrètes des difficultés rencontrées en compréhension et en production par les LNN dans l'usage des PhPex. En nous fondant sur ces traces linguistiques des difficultés observées, nous avons établi plusieurs catégories que nous avons réparties en fonction du fait qu'elles relèvent de la production ou de la compréhension. Nous avons également ajouté dans le tableau les différents niveaux de la compétence

lexicopragmatique (cf. Partie Cadre Théorique, Chap. 6, Section 3) associés à chaque catégorie de difficulté. Nous sommes ainsi parvenus au tableau suivant :

	Type de difficulté	Niveau de la compétence lexicopragmatique associé à la difficulté		
		Compétence pragmalinguistique	**Compétence sociopragmatique**	**Compétence métapragmatique**
Production	Difficulté ajustement du registre en fonction du contexte et des participants		X	
	Difficulté ajustement de l'intensité expressive / de la modalité (aléthique, épistémique)	X	X	X
	Difficulté combinatoire / choix lexicale et actualisation du sens	X		
	Difficultés sur le plan formel et contresens	X		
	Lacune lexicale qui engendre la production d'énoncés déclaratifs ou paraphrase	X		
Production et compréhension	Inadéquation de la visée pragmatique ou de la stratégie communicative	X	X	X
Compréhension	Difficultés à discerner le sens littéral du sens phraséologique	X		X
	Difficulté à discerner la visée pragmatique de l'expression de son effet sur le contexte et les participants		X	X
	Difficulté à différencier des PhPex qui semblent proches sur l'emploi ou différents emplois phraséologiques d'un même PhPex	X		X

Tableau 21 : Synthèse des catégories de difficulté associées aux trois niveaux de la compétence lexicopragmatique

Nous allons dans les sections qui suivent tenter de caractériser chacune des difficultés présentées dans le tableau en précisant sur quel type de traces linguistiques le diagnostic de ces difficultés se fonde.

6.1.1. En production

Première difficulté : ajustement du registre

La première difficulté récurrente constatée est celle de l'ajustement registre de langue employée. Cette difficulté s'est traduite par des emplois corrects et intelligibles sur la forme, mais qui paraissaient peu naturels au vu du registre observé dans le dialogue (cf. Activité 3 (équivalence), *c'est insensé !* employé pour *c'est pas possible !*. Nous avons aussi pu

remarquer au niveau morphosyntaxique que les productions des LNN ne réalisaient pas les phénomènes fréquemment observés dans les emplois au registre familier des LN, c'est-à-dire l'omission de la double négation, la contraction morphologique de plusieurs unités (*il y a -> y'a*) ou la compression des expressions de base en expressions minimales (*c'est chaud ! -> chaud !*). Sur ce point, les résultats de ce test tendent à rejoindre ce que nous avions préalablement observé dans l'étude diagnostique du chapitre précédent. Cependant, l'absence de ces spécificités morphologiques ne constitue pas en soi la trace d'une compétence insuffisante, mais est en revanche une marque typique de l'oral en français et fréquemment retrouvée dans le registre familier. Comme évoqué dans l'analyse, certains schémas conversationnels nécessitent l'abaissement de la distance relationnelle qui s'effectue au niveau linguistique par le passage au registre familier. L'absence de cet abaissement peut donner une impression de non-naturel, voire provoquer un échec pragmatique. Pour faire écho à ce que nous avions évoqué plus tôt, le type d'exposition, notamment à une pratique ou un apprentissage du français en milieu académique, peut avoir joué un rôle dans cette tendance à employer un registre courant, voire formel. En comparant les résultats de l'accord interjuge employé pour la variabilité du registre et le profil des locuteurs lié aux réponses classées en tant que registre courant, nous nous sommes aperçu qu'un nombre important de ces réponses était proposé par des doctorants (67% des réponses relevant du registre courant). Néanmoins, les autres locuteurs ayant utilisé des réponses du registre courant (33% des réponses relevant du registre courant) pratiquent le français dans le milieu professionnel ou familial. De même, la mesure statistique concernant le milieu d'exposition n'avait pas montré de significativité de cette variable sur les résultats au test. Cette piste du lien entre contexte de pratique du français et difficulté à abaisser le registre semble donc intéressante, mais doit donc être prise en compte avec précautions.

Deuxième difficulté : ajustement de l'intensité expressive et de la modalité

Le deuxième type de difficulté rencontré dans notre analyse concerne l'ajustement de l'intensité expressive et de la modalité. En effet, l'analyse qualitative a montré que certains LNN avaient tendance à tantôt manquer d'intensité dans leur réponse, tantôt à être trop expressifs par rapport aux propositions observées chez les LN. De même, les dispositifs de modulation de la modalité aléthique, épistémique ou du marquage de la subjectivité du locuteur étaient beaucoup moins présents dans les productions des LNN que celles des LN. L'un des exemples les plus intéressants de cette difficulté est le cas de l'expression *c'est pas*

gagné ! (Activité 3, dialogue 3, item 1). Cette expression suppose une faible chance de concrétisation d'un évènement, sans pour autant l'affirmer catégoriquement. Or, nous avons pu observer un certain nombre de réponses où la modalité épistémique était absente (ex. : *c'est dommage !*) ou des réponses expriment un degré de certitude trop élevé (*c'est impossible !, il viendra pas !*). En revanche, les réponses proposées en japonais dans l'activité 3 (équivalence) ont montré de grandes similarités avec les réponses en français des LN. Cette similarité des stratégies pragmatiques mobilisée montre encore une fois que les LNN comprennent bien le besoin communicatif, savent l'exprimer dans leur LM, mais ne possèdent pas une palette lexicopragmatique assez développée dans la LC pour produire un PhPex adéquat.

Troisième difficulté : combinatoire et actualisation du sens et de la fonction

Le troisième type de difficulté concerne les choix combinatoires et lexicaux effectués lors de la production, et notamment leurs effets sur l'actualisation de la fonction pragmatique dans les dialogues proposés. L'analyse a montré à plusieurs reprises que la connaissance au niveau combinatoire (compétence pragmalinguistique) permettant d'appréhender les différents changements de sens et de fonction pragmatique qui peuvent intervenir lors de la suppression, de l'insertion ou de la permutation d'éléments dans la séquence phraséologique employée (compétence métapragmatique) n'était pas complètement maitrisée. Cela avait pour conséquence d'actualiser un sens inadéquat, voire un contresens. Il s'agissait par exemple du cas de *certainement pas !* qui avait été proposé en tant qu'équivalent de *c'est pas gagné !*. Le locuteur ayant proposé cet équivalent voulait probablement exprimer que l'ami dont il était question ne viendrait certainement pas. Toutefois, en employant cette séquence telle quelle et avec un point d'exclamation qui plus est, la fonction actualisée est celle de refus et non pas de modulation du degré de certitude.

Quatrième difficulté : lacunes sur le plan formel et contresens

Le quatrième type de difficulté observé concerne les erreurs sur le plan formel (grammaire, syntaxe, morphologie) et des incohérences. Nous avons regroupé ces deux catégories en une seule, car elles relèvent également de la compétence générale dans la langue cible. Ainsi, les erreurs portant sur le choix du pronom, de la préposition ou alors sur des constructions syntaxiques particulières peuvent engendrer des changements de sens ou de fonction

pragmatique, d'autant plus si le phraséologisme dont il est question est fortement figé sur le plan syntaxique. Nous avons par exemple les cas de *on ne peut pas assurer qu'il vient* et *ça passera pas comme on a prévu* proposé dans l'activité 3 (équivalence) comme équivalent de *c'est pas gagné !.* L'absence de la pronominalisation du verbe ici actualise un sens qui n'est pas compatible avec l'expression de base. Toujours dans le même exemple, certains LNN ont proposé les séquences *c'est pas réussi* et *ça marche pas* qui semblent incohérentes avec la fonction de l'expression *c'est pas gagné !.* Dans ces deux derniers cas, il nous a semblé que ces emplois n'étaient pas fortuits et résultaient selon nous soit d'une analogie malheureuse avec la LM, soit d'une interprétation compositionnelle du PhPex proposé. En effet, le verbe *gagner* semble relativement proche sur le plan sémantique avec *réussir* et *marcher* (dans le sens de *fonctionner, aller comme sur des roulettes*). En outre, dans un certain nombre de réponses analysées, nous nous sommes rendu compte que les LNN qui ne disposaient pas du phraséologisme attendu ou d'une expression proche sur le plan fonctionnel avaient tendance à produire des énoncés déclaratifs ou des périphrases. C'est notamment, la forte récurrence de ce type d'énoncé combiné à un faible taux de réussite dans les résultats en japonais et un faible indice de phraséologisation qui nous avait donné les indices concernant les difficultés liées à l'interprétation de *c'est chaud !* ou de *c'est pas gagné !* dans l'activité 3 (équivalence).

Cinquième difficulté : adéquation de la visée pragmatique ou de la stratégie communicative

La difficulté suivante touche à la fois à la compétence de compréhension et de production. Elle concerne également les trois niveaux de la compétence lexicopragmatique. Il s'agit de l'inadéquation de la visée pragmatique ou de la stratégie communicative employée par les LNN avec l'illocutoire attendu dans le dialogue proposé. Ce type de difficulté se traduit notamment par la production de séquences phraséologiques dont l'illocutoire, la stratégie discursive ou la posture communicative ne semblaient pas naturels dans le dialogue proposé, voire inadaptés. Cette difficulté portait le plus souvent sur des PhPex polyfonctionnels (ex. : *c'est chaud !*) ou fortement idiomatiques (ex. : *c'est pas gagné !*) pour lesquels on remarque qu'il n'existe bien souvent pas d'équivalent direct dans la LM. Par exemple, dans l'item 2 du dialogue 3 de l'activité 2 (reformulation), nous avons pu observer l'emploi de *j'ai le droit d'être là* dans une situation où le locuteur qui vient de faire tomber le portable de son frère est censé se dédouaner en utilisant l'expression *j'y suis pour rien.* Or, l'emploi proposé par

le LNN n'a pas été validé par les juges, car l'illocutoire véhiculé est pragmatiquement inadaptée à la situation et va au contraire dans la provocation et la revendication plutôt que dans le dédouanement. De même, nous avons pu observer certains énoncés qui suggéraient une stratégie communicative en décalage avec ce qui était observé habituellement avec les LN. Nous avions par exemple observé que les LNN avaient plus tendance à adopter une posture empathique dans leur production qu'à s'indigner.

6.1.2. En compréhension

Première difficulté : discerner le sens phraséologique du sens littéral

La première difficulté en compréhension que nous avons pu observer concerne la faculté à discerner le sens littéral du sens phraséologique de l'expression proposée dans le dialogue. En effet, l'utilisation des distracteurs a permis d'observer que certains des LNN n'avaient pas conscience du sens phraséologique de l'expression visée et *vice versa*. Nous avons par exemple observé cela sur l'expression *c'est le pied ?* pour laquelle un nombre significatif des LNN avait projeté la fonction de réaction émotionnelle positive observée dans le dialogue 1 de l'activité 2 (expérience du massage) sur le sens compositionnel observé dans le dialogue deux de l'activité 5 (le locuteur désigne son pied). Un autre exemple intéressant est observable dans l'activité 3 (équivalence, dialogue à propos de l'ami accidenté). L'un des LNN a proposé l'expression *c'est magnifique !* comme équivalent à l'expression *c'est chaud !* qui était ici employé comme une réaction évaluative négative comportant de l'empathie suite à la mauvaise nouvelle de l'ami accidenté. En regardant l'équivalent proposé en japonais *geki atsu* (*c'est énorme ! c'est super !*), nous en avons déduit que le LNN avait d'une part retenu le sens littéral de *c'est chaud !* (température élevée), d'autre part avait commis un contresens en interprétant de manière compositionnelle l'expression. En effet, *geki atsu* signifie littéralement 'extrêmement chaud', et on peut supposer que le choix lexical s'est effectué par analogie avec le sens de chaleur élevée en français. Il s'agit manifestement ici d'une difficulté liée à une confusion entre le sens littéral et le sens phraséologique.

Deuxième difficulté : discerner la visée pragmatique et ses effets sur le contexte

La seconde difficulté observée au niveau de la compréhension concerne l'amalgame fréquemment observé dans les réponses à l'activité 5 entre la visée pragmatique de l'expression (illocutoire) et son effet sur le contexte et les participants (perlocutoire). Par exemple, dans l'activité 5, dialogue 1, item 3 (*c'est pas possible*), un certain nombre de locuteurs ont sélectionné le sens d'indignation ou de surprise alors qu'il s'agissait d'un distracteur dont le sens n'exprimait rien d'autre que l'incapacité de réaliser une action. Étant donné que dans l'autre dialogue où *c'est pas possible !* avait son sens phraséologique (Activité 2 (reformulation), Dialogue 4, item 1), le taux de compréhension était haut (90%), nous en avons déduit qu'il ne s'agissait pas d'un problème de compréhension entre les deux acceptions de *c'est pas possible !*. Il y a en revanche une confusion entre l'effet produit par l'expression au moment de son énonciation (marque l'impossibilité) et l'effet sur le locuteur (indignation et surprise). Ainsi, le locuteur qui exprime l'impossibilité de réaliser la tâche qui lui est demandée ressent probablement de l'indignation et de l'étonnement, mais la séquence employée en elle-même n'exprime pas l'indignation.

Troisième difficulté : différencier la proximité fonctionnelle ou la polyfonctionnalité d'un PhPex

La dernière difficulté concerne la capacité à différencier des PhPex qui semblent proches sur l'emploi ou alors les différents emplois phraséologiques d'un même PhPex. Dans le premier cas, nous avons pu par exemple remarquer certaines confusions entre les PhPex *c'est le bouquet !* et *c'est une horreur !* qui bien que proche sur l'illocutoire qu'ils véhiculent se différencient sur le processus qui conduit à la réaction émotionnelle négative. En effet, l'usage de *c'est le bouquet !* est contraint à des situations d'accumulation d'évènement négatif alors que *c'est une horreur !* est plus générale dans son emploi. Dans le deuxième cas, nous avons observé certaines difficultés des LNN à différencier les emplois qu'un même PhPex peut actualiser en fonction du contexte. Nous pensons notamment à l'expression *c'est pas vrai !* qui peut tantôt marquer l'exaspération tantôt la surprise négative provoquant une certaine inquiétude. Dans ce dernier cas, c'est bien souvent le trait prosodique, en plus des informations extralinguistiques, qui permettent de désambigüiser la fonction pragmatique et expressive dont il est question.

6.2. Autres difficultés

Concernant les autres difficultés constatées et qui ont eu, semble-t-il, un effet sur le résultat obtenu, nous avons pu observer deux types d'écueils. Le premier problème constaté, cependant inéluctable dans ce type de travail exploratoire concerne les différents biais méthodologiques constatés à postériori de la création du test, lors de l'analyse. En effet, nous nous sommes aperçu dans l'activité 2 par exemple que la consigne en elle-même était complexe et nécessitait d'accomplir un travail de reformulation qui s'avère exigeant et dont la réalisation implique un temps de réponse plus important et un développement de la consigne plus explicite (notamment sur la différence entre reformulation et équivalence). Le même problème s'est posé pour l'activité 4 qui impliquait un grand nombre d'étapes d'analyse pour les LNN (contexte, traduction, différenciation et discrimination) pour lesquelles il aurait fallu laisser plus de temps de réponse.

Le second type d'erreur constaté est lié au phénomène de *speeding* et à la fatigue cognitive inhérente au passage du test. En effet, tout type d'activité exigeante sur le plan cognitif implique qu'un soin particulier soit appliqué à la réflexion sur l'organisation et l'enchainement des activités, ainsi que sur la durée du test. Étant donné le caractère très exploratoire de notre travail de thèse, nous avions procédé qu'à une seule étape de validation pilote, qui avait semblé n'avoir mis en relief aucune difficulté particulière liée au format. Toutefois, avec la crise sanitaire actuelle, nous avons dû revoir de manière assez précipitée le protocole de test pour l'adapter à une modalité à distance. Cette mise en place du nouveau protocole a fait émerger de nouvelles contraintes qui ont eu vraisemblablement un impact sur la passation du test. Sur le plan externe à la passation, nous avions remarqué que certains des informateurs étant rentrés au Japon en raison de la crise sanitaire, nous avions dû passer le test très tard le soir en raison d'un décalage horaire. Dans d'autres cas, cette passation tardive était liée au fait que l'informateur travaillait et ne pouvait passer le test qu'à la fin de la journée. De plus, nous avons pu constater que les informateurs qui ont passé le test à distance ne disposaient pas pour certains d'un matériel adéquat (casque audio pour écouter, bonne connexion internet, écran convenable), dont certains qui ont réalisé le test sur leur smartphone. Nous pensons donc que ces variables temporelles et matérielles ont pu jouer sur la qualité des réponses. Sur le plan interne, nous avons dû adapter un test qui au départ a été élaboré pour une modalité en présentiel, à une modalité distance au moyen d'un formulaire en ligne. Ce transfert vers une autre modalité a eu pour effet de complexifier le déroulement

du test et d'allonger sa longueur, car de nouvelles étapes sont apparues en raison des contraintes de fonctionnalité du formulaire en ligne. De ce fait, il nous a semblé que cette complexification du format et cet allongement du temps de réponse ont pu influencer là aussi la qualité des réponses.

Nous arrivons à l'issue de cette deuxième phase de notre étude. Cette deuxième phase nous a permis d'une part d'étudier plus finement les difficultés liées à la compétence lexicopragmatique repérées dans la première phase diagnostique. D'autre part, nous avons pu observer de manière systématique les spécificités d'emploi des PhPex chez des locuteurs japonophones du français de niveau avancé et proposer certaines mesures (statistiques lexicales, taux de compréhension, taux de réussite) qui nous ont permis de caractériser la variété non native des LNN observés. Ce travail d'analyse fine nous a ainsi conduit à établir une typologie des difficultés rencontrées par les LNN lors de la production et de la compréhension des PhPex. Dans la partie qui suit, nous allons aborder la troisième phase de cette étude qui constitue un éclairage des données du point de vue des informateurs. Cette troisième phase prend la forme d'une étude orientée vers le niveau métapragmatique de la compétence lexicopragmatique et fondée sur une série d'entretiens semi-directifs menés après la passation du test pour l'ensemble des 35 LNN engagés dans notre protocole expérimental. Elle nous permettra notamment d'affiner certaines des observations faites au cours de la deuxième étape.

Troisième phase du protocole expérimental :

Entretiens semi-directifs d'élicitation métapragmatique des réponses au test lexicopragmatique de locuteurs japonophones du français de niveau avancé

Point étape 2 du protocole expérimental

L'étape diagnostique de notre protocole expérimental a permis de proposer un premier état de la compétence lexicopragmatique en production des locuteurs japonophones de niveau intermédiaire en contexte d'interaction exolingue informelle avec des LN du français. À l'aide d'un corpus d'interaction orale semi-élicitée qui a été transcrit, annoté et rigoureusement analysé, nous avons pu proposer des observations préliminaires concernant la compétence lexicopragmatique des LNN japonophone locuteurs du français. Ces observations étaient les suivantes :

- Un sous-emploi des PhPex compensé par le suremploi du rire ou de réaction vocale
- Des phénomènes d'appui palliatifs sur la LM traduisant une lacune lexicopragmatique
- Les difficultés d'ajustement de la stratégie pragmatique en situation d'interaction avec notamment un suremploi des formes phatiques
- Les difficultés liées à la maitrise de la combinatoire lexicale et de ses effets sur l'actualisation de la fonction pragmatique de l'énoncé en contexte.
- La motivation et le but communicatif jouent un rôle fondamental dans le développement la palette lexicopragmatique du LNN.

La deuxième étape s'est articulée autour d'un test permettant de mesurer la compétence lexicopragmatique à la fois en production et en compréhension. Cette deuxième étape nous a permis d'étendre ces différentes observations aux spécificités d'usage des PhPex chez les LNN de niveau avancé. Nous avons ainsi dressé une typologie des difficultés d'emploi et de compréhension rencontrées lors de la passation du test. Cette deuxième phase a également permis de développer différents types de mesure de la compétence lexicopragmatique, dont les mesures de statistiques lexicales et la mesure du taux de compréhension. Ces mesures permettent notamment d'orienter le chercheur sur des phénomènes qui seraient d'ordinaire invisibles à l'œil nu. Nous avons ainsi pu montrer qu'une approche outillée fondée sur l'analyse fine des productions des LNN est tout à fait pertinente et utile à la compréhension de l'acquisition des compétences lexicopragmatiques nécessaires à l'emploi des PhPex. L'une des autres originalités de cette seconde phase était la dimension contrastive prise en

compte dans l'interprétation des résultats et qui s'est avéré d'une grande utilité dans la compréhension des difficultés rencontrées par les LNN, notamment en compréhension.
Cette troisième étape a pour objectif d'apporter un éclairage supplémentaire sur les résultats obtenus dans la seconde phase, et de manière plus générale, sur les difficultés, les stratégies d'usage et de compréhension des PhPex. Cette étape s'appuie sur l'analyse de 35 entretiens effectués directement après la passation du test lexicopragmatique. Une des visées est d'explorer le potentiel heuristique de l'approche par entretien pour affiner les observations faites jusqu'à présent concernant la compétence lexicopragmatique des LNN japonophones. Nous ciblerons ici la compétence métapragmatique (Cadre théorique, Chapitre 7, section 2) à l'aide d'une approche semi-directive croisant discussion guidée sur les résultats du test lexicopragmatique et élicitation conjointe des stratégies de compréhension et de sélection des PhPex. Afin de mener cette élicitation conjointe, nous nous fondons sur deux activités de mise en situation orientées tantôt vers la compréhension, tantôt vers la production. Nous proposons ensuite une synthèse fondée sur une extraction thématique des difficultés discutées lors des entretiens et sur les réponses proposées lors des deux activités de mise en situation. Nous tenterons ensuite de faire le lien avec les motifs de difficultés récurrents dans les entretiens et les résultats observés dans les deux précédentes phases.

Dans un premier temps, nous évoquons le cadre de déroulement des entretiens ainsi que les spécificités méthodologiques liées au recueil des données. Dans un second temps, nous explicitons les modalités de traitement et d'analyse des données d'entretiens. Enfin, nous proposons une analyse thématique que nous contrasterons avec les résultats obtenus lors des précédentes étapes de notre protocole expérimental.

Chapitre 13 - Méthodologie de recueil et d'analyse des données des entretiens semi-directifs

1. Questionnement de départ et objectif

La dernière phase de ce travail de thèse est le fruit de nombreuses réflexions concernant les mécanismes psycholinguistiques et interprétatifs qui sous-tendent le choix et la compréhension des PhPex. La deuxième phase de cette étude a montré que les difficultés rencontrées par les LNN dans la production des PhPex résidaient à un niveau relativement complexe et polyfactoriel (combinatoire lexicale, actualisation du sens, ajustement de la force illocutoire, posture pragmatique, etc.), en partie inaccessible par la simple exploration des données obtenues à partir de la production écrite. En effet, il demeure complexe d'étudier de manière empirique ce qu'il se passe au niveau cognitif chez le locuteur ou de déterminer quelles sont les stratégies qu'il mobilise pour sélectionner une expression adéquate au besoin communicatif qu'il rencontre. Cette inaccessibilité à ce que nous pourrions appeler la « boite noire » lexicopragmatique semble notamment due à la grande difficulté que nous avons actuellement pour obtenir des observables dont la matérialité serait suffisante à la description des stratégies d'usage et de compréhension des PhPex chez le locuteur, qu'il soit natif ou non-natif. Un autre obstacle à l'analyse de ces stratégies réside dans le fait que les connaissances nécessaires à l'emploi des séquences préfabriquées de la conversation quotidienne dépassent le simple cadre du dire et touchent à la sphère du savoir partagé et conventionnellement admis dans une communauté de locuteur, savoir présupposé par tous, mais non marqué linguistiquement. C'est ce niveau de granularité qui nous intéresse ici et qui a guidé notre réflexion sur les stratégies linguistiques mobilisées par nos informateurs dans l'usage et la compréhension des PhPex. Se pose alors la question de comment explorer ces stratégies et du protocole à mettre en place pour pouvoir observer de manière empirique ces phénomènes. Plusieurs approches semblent possibles pour explorer cette « boite noire » lexicopragmatique, parmi lesquelles le recours à l'élicitation de phénomènes psycholinguistiques par le locuteur lui-même. Dans la partie de la thèse consacrée cadre théorique de cette étude, nous avions mentionné plusieurs protocoles permettant au locuteur d'éliciter sa production ou sa compréhension (Cadre théorique, Chapitre 7, section 2), parmi lesquels :

- Le protocole de verbalisation pendant l'activité (*think aloud protocole*, Taguchi et Kim 2016)
- L'entretien de suivi portant sur la pratique de l'objet linguistique cible (*follow-up interview*, Taguchi 2019),
- L'entretien subséquent à une activité (*retrospective interview*) (Taguchi 2011),
- Les questionnaires de jugement métapragmatique (*Metapragmatic assessment questionnaires*, Rose et Kg Kwai-fun, 2001 p 157-158)
- Les autobiographies d'apprenants (ou journal de suivi, Hassall 2006).

Néanmoins, la configuration du protocole à distance ne nous permet pas de mettre à profit ces différentes approches du fait des exigences de temps, de format et de la fatigue cognitive inhérente à la passation du test. Par exemple, comme évoqué dans la partie portant sur le cadre théorique, le protocole de verbalisation est très difficile à mettre en place et comporte de nombreuses contraintes cognitives pour le locuteur. Étant donné que cette phase d'élicitation par l'informateur devait nécessairement faire suite à 30 minutes de passation de test, il nous a semblé risqué de tenter cette approche. En outre, le protocole de verbalisation implique un cadrage expérimental exigeant qu'il nous était impossible d'atteindre étant donné la modalité à distance. Les entretiens de suivi ou le journal de suivi sont plus adaptés à un travail longitudinal espacé sur plusieurs mois, ce qui n'est pas le cas de notre étude. Le questionnaire de jugement peut être une option envisageable, mais il nous a semblé que cette modalité aurait été redondante avec le test lexicopragmatique. Ce faisant, il y aurait eu des risques de *speeding* ou des biais de réponse liés à la répétitivité de la tâche. La dernière option est celle de l'entretien rétrospectif. Il s'agit d'une forme d'entretien exploratoire communément utilisé en acquisition L2 pour obtenir le feedback d'un formateur après la passation d'un protocole expérimental. Cette dernière approche nous a semblé bien adaptée au besoin de la troisième phase de notre étude. Elle nécessitait cependant quelques adaptations du fait de son caractère très directif et peu interactif.

Enfin, la réflexion préalable à la mise en place de notre protocole exploratoire est grandement tributaire de l'apport des approches contrastives (comparer les deux systèmes linguistiques) et translinguistiques (étudier les deux langues et leurs influences mutuelles) développées dans la pragmatique interlangagière (Kasper et Blum-Kulka 1993) et l'analyse contrastive interlangagière (Granger 2015) que nous avons pu développer dans notre cadre théorique (Chapitre 5). Une étude de ce type ne peut être conduite sans la prise en compte

des différences, mais aussi des interactions observables entre la langue cible et la LM du LNN. Nous avons donc pris en compte ce dernier aspect dans la réflexion sur l'approche à adopter pour cette dernière phase. C'est dans ce cadre que l'*entretien d'élicitation métapragmatique semi-directif* trouve tout son potentiel heuristique. Nous allons développer plus en détail cette approche dans la section qui suit.

2. Principe et structure de l'entretien d'élicitation métapragmatique semi-directif

2.1. Spécificités de notre approche

Compte tenu des spécificités de notre objet d'étude, l'entretien d'élicitation métapragmatique semble tout à fait approprié pour appréhender la matière à la fois générale et complexe que constituent les élicitations des informateurs. Il constitue en effet un cadre rigoureux qui nous permettra de limiter les biais subjectifs tout en obtenant des données structurées et organisées en thématiques que nous pourrons ensuite soumettre à un traitement quantitatif et qualitatif.

La méthodologie d'entretien employée dans le cadre de notre étude est celle de l'entretien semi-directif, de type rétrospectif, c'est-à-dire positionné après le test en vue d'étayer les choix de réponse de manière conjointe avec l'informateur. Cette approche repose sur l'importance du cadrage de l'entretien qui est fait en amont, mais laisse tout de même une certaine liberté dans les échanges qui se produisent entre l'enquêteur et l'informateur. Cette flexibilité est propice à l'obtention d'informations de première main tout en limitant les digressions potentielles et en gardant une certaine exigence scientifique dans le recueil des données. L'entretien semi-directif constitue donc un juste milieu entre informations écologiques de bonne qualité et pertinence du relevé. L'entretien semi-directif offre également la possibilité de changer de thématique en fonction des réponses de l'informateur et de sa réceptivité aux questions tout en restant dans le cadre défini par l'enquêteur. Il nécessite cependant la prise en compte de plusieurs cas de figure afin de pouvoir rediriger plus facilement l'entretien si besoin. L'objectif n'est pas d'obtenir des réponses précises sur l'objet étudié, mais d'obtenir des indications sur les stratégies, les hypothèses ou les pratiques des LNN lors de l'usage des PhPex. La mise en confiance est décisive dans cette

approche puisqu'elle valorise l'informateur et son point de vue, ce qui permettra de dévoiler des contenus latents peu visibles pour le chercheur lors de l'analyse des données linguistiques.

Cet entretien semi-directif a été conçu dans une modalité d'élicitation, c'est-à-dire dans une modalité de verbalisation d'un processus cognitif face à un stimulus qui sera de nature linguistique dans le cadre de notre étude. La notion d'élicitation est une notion centrale en acquisition de la L2. Elle est ainsi à la base de nombreux protocoles impliquant la production du locuteur. L'élicitation qui est pratiquée lors de ces entretiens semi-directifs est de niveau métapragmatique. Elle s'intéresse donc au périmètre de la compétence métapragmatique (cf. Cadre théorique, Chap. 6, section 3) du LNN.

La structure des entretiens semi-directifs est déterminée par la/les question(s) de recherche qui permet de définir chacun des points de focalisation de la démarche. Ce faisant, le guide d'entretien qui sert de cadre à l'entretien est conçu de manière à observer 5 aspects qui nous permettront de mieux comprendre les mécanismes de compréhension ou d'usage des PhPex au niveau lexical et stratégique :

- La compétence de prototypage de l'emploi des PhPex
- Les types de difficultés rencontrés par l'informateur
- Les stratégies de contournement de la difficulté en production ou en compréhension
- Les interactions entre la LM et la LC en production ou en compréhension
- Les indices contextuels sur lesquels se fonde le LNN en production ou en compréhension

Afin de mobiliser cette posture métapragmatique lors de l'entretien, nous avons imaginé un protocole interactif basé sur le principe de *discussion métapragmatique.* Cette interactivité repose sur : a) le retour sur les réponses au test, b) des mises en situation concrètes puis élicitées et c) une discussion libre qui sera bien souvent un moment privilégié de comparaison entre le français et le japonais. La question de la prise de conscience des contraintes d'emploi supervisée de manière implicite ou non par l'enseignant semble avoir le vent en poupe depuis les années 90 sous la forme du concept de *conscience métapragmatique* (*metapragmatic awareness*) chez les chercheurs anglo-saxons (House 1996 ; Verschueren 2000 ; Kinginger et Farrell 2004 ; Culpeper, Mackey et Taguchi 2018).

La notion de *discussion pragmatique*, c'est-à-dire d'un réel échange participatif (et non pas uniquement guidé par l'enseignant) entre le locuteur expert et l'apprenant de la LE dans le but d'étayer le fonctionnement d'un phénomène pragmatique, a été très peu discutée. À notre connaissance, Harada (2007, 2020a, 2020b) demeure la seule tentative documentée de mise en place d'un protocole de ce type pour des apprenants japonophones du français. L'ajout des mises en situation dans l'entretien semi-directif a été pensé afin d'activer chez le LNN une posture d'analyse et de mise en hypothèse de ses différentes observations suite à la production et la compréhension. En mobilisant ainsi ce niveau fin d'analyse développé dans la LM du LNN et transposé à la LC, nous avons l'objectif de pouvoir explorer cette « boite noire » et de pouvoir obtenir des informations cruciales concernant l'emploi des PhPex. Dans la partie qui suit, nous allons aborder la structure de l'entretien.

2.2. Déroulement de l'entretien

Les questions posées aux informateurs ont été établies sur la base des observations formulées à l'issue de la première phase diagnostique puis étayées dans le test métapragmatique. Elles ont été élaborées de manière à explorer l'usage et la compréhension des PhPex sous la perspective de l'usager et non du point de vue externe du chercheur :

Prise de contact et début de l'entretien	
Questions proposées	Objectif(s) visé(s)
- Avant de commencer l'entretien souhaites-tu discuter à propos du test ou autre chose ?[186] - Souhaites-tu discuter de quelque chose en particulier avant que nous commencions ?	**Initier une posture autoréflexive sur les résultats de l'informateur** **Recueillir les premiers commentaires « à chaud ».**
Phase 1 : Discussion rétrospective sur les résultats au test lexicopragmatique	
Regarder les résultats de l'informateur et les questions d'autoévaluation de la compréhension posée pour repérer rapidement des expressions ayant posé des difficultés. Poser ensuite des questions sur certains énoncés. L'idée est d'essayer, malgré la distance, de discuter avec le locuteur de ses choix d'expression et des mécanismes sollicités pour répondre aux activités.	
Questions proposées	Objectif(s) visé(s)
- As-tu ressenti des difficultés pour comprendre cette expression ? - Pourquoi était-ce difficile de comprendre ? - Tu utilises souvent cette expression ? - Si non, pourquoi tu n'utilises pas cette expression ?	**Questionner les types de difficultés rencontrées (contenu et forme)** => les

[186] Nous avons privilégié le tutoiement pour mettre en confiance l'informateur et créer un climat propice à la discussion. En effet, parler de sa compétence linguistique peut placer l'informateur en situation d'insécurité linguistique. Ce faisant, la mise en confiance de l'informateur lui a permis selon nous de parler plus facilement de ses lacunes linguistiques. Le « vous » a été cependant employé dans de rares cas où les informateurs semblaient plus à l'aise avec le vouvoiement, et ce de manière à ne pas imposer la modalité de tutoiement.

- As-tu ressenti des difficultés pour comprendre certaines expressions ? (Si pas le temps de regarder les résultats du test). - Si oui, quelle était cette expression ? - Pourquoi était-ce difficile de comprendre ? - Tu utilises souvent cette expression ? - Si non, pourquoi tu n'utilises pas cette expression ? - Quelles expressions t'ont paru faciles à comprendre ? - Pourquoi penses-tu que ces expressions étaient faciles à comprendre ? - Tu utilises souvent cette expression ? - Si oui, dans quel contexte ? - Avec quel type de locuteur ?	difficultés linguistiques sont généralement une bonne entrée pour débuter une discussion métapragmatique. Le locuteur est en général plus sensible aux mécanismes d'usage des emplois qui lui posent des difficultés plutôt que ceux qui sont maitrisés et quasi automatisés.
- Quels sont les éléments qui te permettent de comprendre les expressions du test ? - Les mots qui composent l'expression sont-ils utiles pour comprendre ? - Si oui, traduis-tu souvent les expressions en japonais pour mieux les comprendre ? - Penses-tu que les expressions dont on ne peut pas comprendre le sens avec les mots qui la composent sont plus dures à comprendre ? - Quand tu rencontres une expression très compliquée et que tu ne comprends pas, que fais-tu ? - Le contexte est-il utile pour comprendre ? - Si oui, quels sont les éléments du contexte qui t'aident à comprendre ? - Le contexte est-il utile pour choisir une expression appropriée ? - Si oui, quels sont les éléments du contexte qui t'aident à choisir ? - Est-ce que le fait d'entendre la façon dont les personnes parlent, leur intonation t'a aidé à comprendre l'expression ? - Si oui, peux-tu donner un exemple ? - Est-ce plutôt l'intonation ou l'émotion présente dans la voix du locuteur qui t'ont aidé à comprendre ? - Le script suffit-il pour comprendre ? - Si non, pourquoi le script ne suffit pas ? - As-tu à certains moments compris le contexte, mais pas réussi à répondre, car tu ne connaissais pas l'expression appropriée ? - Si oui, peux-tu donner un exemple dans le test ? -Pourquoi tu as senti que c'était difficile de trouver une réponse pour cette expression ? - Penses-tu que ce serait aussi difficile de répondre en japonais ? - Si oui, pourquoi est-ce difficile même en japonais ?	**Questionner les indices contextuels sur lesquels reposent l'interprétation et la production.** **Comprendre l'importance que peuvent avoir les paramètres prosodiques et extralinguistiques sur la compréhension ou la production.** **S'interroger sur les stratégies analytiques et les différentes hypothèses que peut faire l'informateur en se fondant sur les indices de contextualisation qui sont à sa portée.** **S'interroger sur le rôle du contexte sur le processus interprétatif.**
- As-tu à certains moments compris le contexte dans le test, mais tu n'avais la réponse qu'en japonais ? - Si oui, te souviens-tu de quelle expression en français il s'agit ? - Te souviens-tu quelle était cette réponse en japonais ? - As-tu trouvé l'activité 3 visant à trouver un équivalent japonais des expressions proposées complexes ? - Si oui, pourquoi ? - De manière générale, dirais-tu que le plus facile dans l'activité 3, c'est de donner des équivalents en japonais ou en français ? - Si français, pourquoi était-ce plus facile ? - Si japonais, pourquoi était-ce plus facile ? - Quand tu parles en français, t'arrive-t-il de penser en japonais avant de parler en français ? - Si oui, dans quels cas ?	**Questionner les différentes stratégies d'appui sur la LM mobilisées dans la compréhension et dans la production.** **Questionner les différents types d'interaction entre la langue maternelle et la langue cible et les effets possibles de cette interaction sur la production ou la compréhension.** **Questionner les effets de la manière de communiquer et de l'éthos communicatif de la LM sur la production et la compréhension en langue cible.**

- Quand l'expression en français ressemble à une expression en japonais, penses-tu que cela t'aiderait pour comprendre et employer cette expression ? -As-tu l'impression que tes réponses au test sont parfois plus proches de la manière de communiquer en japonais que de celle en japonais ? - Si oui, peux-tu donner un exemple ? - Est-ce important pour toi de communiquer parfaitement comme un locuteur natif ? - Si oui, pourquoi ? - Si non, pourquoi ?	
Phase 2 : Mises en contexte puis discussion métapragmatique	
Il s'agit ici de proposer plusieurs mises en contexte à l'informateur afin de tester en temps réel sa compétence lexicopragmatique en compréhension et en production, puis initier ensuite une discussion métapragmatique. Le principe est de pouvoir questionner les stratégies d'usage et de compréhension du locuteur directement après la mise en contexte, tant que les processus cognitifs sont encore frais dans l'esprit de l'informateur. Nous discutons ensuite les conditions d'emploi ou les contraintes contextuelles liées aux réponses proposées par l'informateur.	
Questions proposées	Objectif(s) visé(s)
En production[187]	
- Mise en situation 1 : Tu vois un homme frapper son chien dans la rue, que dis-tu ? (Indignation, réprobation) **- Mise en situation 2 :** Ton ami(e) vient de renverser son soda sur ton t-shirt tout neuf -qui t'a couté très cher. Que réponds-tu ? (Reproche) **- Mise en situation 3 :** Tu sors dehors comme tous les matins pour monter dans ta voiture et aller au travail. Une fois arrivé sur le parking et tu t'aperçois que la vitre de la voiture a été cassée pour te voler ton autoradio et des affaires, que dis-tu ? (surprise négative, indignation) **- Mise en situation 4 :** Tu regardes le journal télévisé et tu apprends que ton groupe de musique préféré s'est séparé que réponds-tu ? (tristesse, déception) **- Mise en situation 5 :** Tu marches dans la rue et quelqu'un te bouscule violemment, que dis-tu ? (Réprobation, indignation)	**L'objectif ici est d'observer la production d'une réaction expressive chez l'informateur.** Dans un second temps, il s'agit de questionner les indices contextuels qui ont permis au locuteur de sélectionner la forme proposée. En cas de difficulté, questionner les raisons de cette difficulté.
En compréhension	
Mise en situation 1 : C'est un navet ! (évaluation négative, déception) • A : Je viens d'aller voir le nouveau film d'action à la mode, *c'est un navet !* **Mise en situation 2 : Mon œil ! (incrédulité, reproche)** • A : C'est moi qui ai gagné la partie, tu es trop nul ! • B : *Mon œil !* je suis sûr que tu as triché ! **Mise en situation 3 : C'est énorme ! (évaluation positive, joie)** • A : Tu savais que Jean a gagné 1 million d'euros au loto ? • B : *C'est énorme !* il en a de la chance ! **Mise en situation 4 : Ça se fait pas ! (évaluation négative, réprobation)** • A : Tu as vu que le patron a décidé de licencier des employés au dernier moment sans les prévenir ? • B : Franchement, *ça se fait pas !* Question d'amorce pour l'élicitation : Qu'est-ce que [PhPex] signifie dans ce dialogue selon toi ?	**L'objectif ici est d'observer la compréhension de phraséologismes expressifs et évaluatifs, fortement opaques et peu prédictibles**. Dans un second temps, il s'agit de questionner les indices contextuels qui ont permis au locuteur d'interpréter la forme proposée. En cas de difficulté, questionner les raisons de cette difficulté.
Phase 3 : Discussion libre	

[187] Afin de ne pas influencer les réponses du locuteur, nous lui avons laissé la possibilité de choisir de ne pas répondre. En effet, nous savions à ce stade de l'étude que les locuteurs japonais pouvaient choisir de ne pas réagir de manière explicite du fait de leur éthos communicatif en langue maternelle. Nous le verrons dans la synthèse, cela a effectivement permis d'envisager le rôle de l'éthos communicatif sur la production en langue cible.

Le locuteur est libre d'ajouter ce que bon lui semble ou de discuter du test librement.	
Questions proposées	Objectif(s) visé(s)
- As-tu d'autres choses à ajouter ? - Souhaites-tu poser d'autres questions ? - D'autres commentaires concernant le test ? - D'autres difficultés que tu souhaites discuter ?	Obtenir des informations complémentaires utiles. Laisser la possibilité au locuteur de proposer des remarques suscitées par la discussion métapragmatique.
Fin de l'entretien	

Tableau 22 : Guide d'entretien employé dans notre étude[188]

La discussion métapragmatique se déroule en trois phases principales et deux pré-phases d'ouverture et de clôture. L'ouverture permet de poser le cadre et de récolter d'éventuelles remarques « à chaud » de l'informateur suite à la passation du test. La première phase consiste en une série de questions visant à identifier les difficultés rencontrées par le LNN lors de la réalisation des activités. Ces questions se focalisaient à la fois sur le contenu (expressions ayant particulièrement posé problème, contenu du dialogue pas compris, etc.) et sur la forme (certaines consignes pas assez claires, activités trop complexes, etc.). D'autres questions plus précises sont ensuite proposées et visent à opérer un retour rétrospectif sur les réponses du LNN au test, et ce de manière à engager la discussion sur les stratégies de compréhension employées pour la contourner. La troisième phase consistait à mettre l'informateur en situation de production ou de compréhension. Les expressions étaient volontairement opaques et les situations de communication peu commune pour les LNN. Ce choix a été fait de afin de déclencher une posture analytique qui déboucherait sur une discussion métapragmatique. Cette mise en situation se divisait elle-même en deux temps, une activité de production où il était demandé aux informateurs de réagir avec un PhPex à des situations de communication de la vie quotidienne, et une activité de compréhension où les informateurs étaient exposés à un micro-dialogue contenant un PhPex qu'ils devaient définir ensuite.

Ce guide d'entretien propose un cadre de discussion que nous avons voulu un peu plus guidé que d'usage de manière à éviter de nous perdre en discussions certes intéressantes, mais peu pertinentes pour le traitement de notre question de recherche. Ce choix a notamment été motivé par le type d'analyse que nous voulions mettre en place par la suite. Les questions

[188] La mise en forme du guide d'entretien sous forme de tableau s'inspire de celle proposée par Pallanti (2021 : 147). Cette approche de modélisation du guide a le mérite selon nous de rendre compte de l'enchainement des questions et des relances, ainsi que de l'articulation des différentes étapes et thématiques abordées.

mentionnées n'ont bien évidemment pas toutes été posées, nous nous sommes adapté en fonction des impératifs de temps et du déroulé de l'échange. Chaque série de questions débute en général par une question ouverte que nous affinons avec des relances de plus en plus précises en fonction de la progression de l'échange. Enfin, ce protocole est aussi appuyé par une enquête (Annexe 11[189]) distribuée quelques semaines après le passage du test et rédigée en japonais pour faciliter une plus grande finesse dans les réponses. Cette enquête a été réalisée sur un formulaire en ligne et était facultative.

2.3. Les données recueillies

Les données ont été recueillies en français directement après la passation du test lexicopragmatique (pas de pause). Ces entretiens ont lieu par visioconférence et ont été enregistrés à l'aide de l'outil *Camstasia* (cf. Chapitre 10, section 3). Nous avons ici choisi le format vidéo, car les réactions physiques et mimo-faciales sont intéressantes et peuvent montrer des signes d'incompréhension ou d'embarras chez le LNN lors de la discussion métapragmatique ou des activités proposées. Nous avons pu remarquer dans l'analyse que les locuteurs japonophones adoptent fréquemment une attitude assez caractéristique de l'incompréhension liée à une difficulté lexicopragmatique. Cette attitude caractéristique mêlant réaction vocale, rire, gestuelle et marques faciales nous a été très utile pour repérer les moments de réelle incompréhension ou d'analyse métapragmatique[190]. À l'issue de ce protocole nous obtenus un corpus de 35 entretiens de longueurs variables pour un total de 16 heures et 22 minutes :

ID	Durée de l'entretien	Spécialité	Niveau CECRL
AM12FJP052020	27min51	SHS[191]	C1
AS13HJP042020	24min15	Études françaises / littéraires	C2
AY1MJP042020	22min49	Études françaises / littéraires	C1
ANH14FJP052020	19min34	Études françaises / littéraires	C1
DR15FJP042020	36min10	Études françaises / littéraires	C2
HY16FCH112019	16min32	ScL[192]	B2
HK17HJP062020	47min34	ScL	C1

[189] Une traduction de ce questionnaire est disponible à l'adresse : https://bit.ly/3yG7KOe
[190] Ce type de réaction n'est pas sans nous rappeler l'usage du rire qui a été observé dans la phase diagnostique (Chapitre 9, section 2.1)
[191] SHS = Sciences Humaines et Sociales
[192] ScL = Sciences du Langage

IM7FJP042020	23min52	Études françaises / littéraires	B2
IM18FJP052020	32min35	ScL	C2
KM19FJP052020	36min54	Études françaises / littéraires	B2
KY20FJP052020	32min56	Études françaises / littéraires	C1
KY21FJP042020	12min03	Études françaises / littéraires	B2
KT8FJP042020	16min08	SHS	C2
KM22FJP052020	24min46	Études françaises / littéraires	C1
KC23FJP052020	36min28	Études françaises / littéraires	C2
KM2FJP052020	50min07	ScL	C2
KK33FJP062020	41min52	ScL	C2
KH9FJP042020	18min04	ScL	C1
MR24FJP062020	48min26	ScL	B2
MAM25HJPFR042020	15min46	Études françaises / littéraires	C2
MA26FJP052020	21min44	SHS	C1
MN33FJP062020	31min04	ScL	C2
MN4FJP042020	22min54	Études françaises / littéraires	C2
NM3FJP042020	23min55	ScL	C2
NM27FJP052020	24min57	SHS	C1
NK34FJP072020	51min05	ScL	B2
OT28FJP042020	23min52	Études françaises / littéraires	C2
OK29FJP052020	38min27	ScL	C1
SN10FJP052020	27min48	SHS	C1
SY11FJP042020	20min50	Études françaises / littéraires	C2
SA30HJP052020	44min56	ScL	B2
TS6FJP052020	26min41	ScL	C1
TM31FJP052020	39min29	ScL	C1
UR5FJP042020	25min53	ScL	C2
YN32FJP042020	11min02	SHS	C2
Temps moyen	**30min65**		
Temps Total	**16h22min**		

Tableau 23 : Synthèse des données recueillies

L'enquête libre distribuée quelques semaines après la passation constitue un questionnaire en ligne constitué de 51 questions réparties en 6 catégories thématiques. Sur les 35 informateurs, 27 ont répondu à ce questionnaire, nous permettant d'avoir de nombreuses informations utiles concernant la situation sociobiographique des répondants, ainsi que sur les difficultés fréquemment rencontrées dans leur expérience de locuteur du français. Nous avons aussi proposé des questions invitant les informateurs à évoquer les différences linguistiques entre le français et le japonais, et ce de manière à obtenir des informations intéressantes sur les différences d'éthos communicatif.

3. Méthodologie d'analyse

3.1. Dépouillement et formalisation des données

Notre objectif n'étant pas l'analyse linguistique de la matérialité orale de l'échange ou une étude microsociologique de la pratique linguistique des informateurs, nous n'avons pas jugé pertinent de procéder à une transcription sur Elan comme celle que nous avons menée dans la phase diagnostique. Nous avons à la place procédé à une analyse qualitative sous la forme d'un relevé. Nous avons dans un premier temps écouté chaque entretien une première fois pour créer un premier aperçu des données à relever. Nous avons ensuite procédé à deux écoutes supplémentaires de l'intégralité des entretiens. Au cours de ces deux écoutes, nous avons relevé les éléments saillants liés au travail d'analyse métapragmatique et fait l'inventaire des réponses données aux questions posées lors des mises en situation de l'entretien semi-directif. Nous avons ensuite élaboré un tableau Excel (Annexe 12) afin de formaliser les données relevées lors des 3 visionnages et permettre un traitement plus global et organisé de ces données. Ce tableau se divise en 3 parties. La première partie est dédiée au relevé des réponses aux questions rétrospectives du guide d'entretien (phase 1 de l'entretien). La seconde partie comporte les réponses aux mises en situation, ainsi que les éléments de discussion métapragmatique qui s'ensuivirent (phase 2 de l'entretien). La troisième partie contient les éléments de discussion additionnelle qui ont été ajoutés lors de la discussion libre. Enfin des informations sociobiographiques concernant chacun des LNN (niveau CECRL, lieu de résidence, activité professionnelle, lieu d'obtention du diplôme, etc.) ont été ajoutées pour nous permettre d'affiner l'analyse. Nous avons également ajouté à cette grille les taux de réussite et de compréhension obtenus au test lexicopragmatique par chaque informateur. Pour certains entretiens qui présentaient un intérêt particulier pour notre analyse, nous avons effectué un retour sur les réponses au test afin d'affiner notre analyse et de mieux comprendre comment se recoupent les éléments évoqués lors de la discussion métapragmatique avec les réponses proposées dans le test. Nous avons ensuite dépouillé les questionnaires en langue japonaise qui ont été proposés quelques semaines après la passation du test au LNN afin d'obtenir quelques éléments d'analyse supplémentaires.

3.2. Codage du relevé effectué sur les entretiens

La formulation d'un guide d'entretien structuré a permis de donner une ligne directrice et un cadre aux échanges effectués avec les informateurs. Cette structuration des données menée en amont nous a permis d'établir une hiérarchisation des phénomènes discutés lors des entretiens et relevés lors de l'écoute. À partir de ce relevé pré-structuré par le guide, nous avons pu voir se dessiner des thématiques réunissant un certain nombre de phénomènes intéressants qui se recoupent avec les observations effectuées dans les deux phases précédentes. Ces différentes thématiques nous servent de base à la grille de codage qui est appliqué aux données. La grille de codage a été améliorée à partir de la typologie des difficultés modélisée dans la phase 2 de notre protocole expérimental. Le codage est un outil essentiel au traitement méthodique de la matière très variable que constituent les données d'entretien (Bardin 2018). Il suppose plusieurs allers-retours entre les données brutes, les relevés effectués et les hypothèses préalables à l'analyse. À partir de ce codage, nous avons établi une série d'étiquettes qui ont ensuite été appliquées au tableau des données relevées lors du visionnage des entretiens :

Catégories	Sous-types	Définition	Étiquette
Codage des difficultés			
Difficulté en production	Ajustement intensité expressive et modale	L'informateur a des difficultés à ajuster l'intensité expressive ou modale de son propos en fonction de la situation de communication qu'il rencontre.	AJUST-INTENS
	Combinatoire et actualisation du sens	L'informateur produit des énoncés qui actualisent un sens qui n'est pas adéquat dans la situation de communication qu'il rencontre.	COMBIN
	Difficulté formelle	L'informateur produit des énoncés dont la forme ou la structure syntaxique n'est pas adéquate dans la situation de communication qu'il rencontre.	FORM
	Lacune lexicale / transfert	L'informateur ne parvient pas à produire un PhPex ou a recours à un terme dans sa LM ou calqué sur sa LM pour pallier cette lacune.	LEX-TRANS
	Conflit éthos communicatif	L'informateur connait le PhPex idoine à la situation de communication qu'il rencontre, mais ne le produit pas, car celui-ci entre en conflit avec sa manière de communiquer.	CONF-COMM
Difficulté en compréhension	Discernement sens phraséologique / sens littéral	L'informateur ne parvient pas à discerner le sens littéral du sens phraséologique des PhPex qui lui sont proposés.	PHRASEO-LITT
	Compréhension fonction pragmatique	L'informateur ne parvient pas à comprendre l'illocutoire des PhPex qui lui sont proposés. L'informateur ne comprend pas les spécificités pragmatiques du contexte auquel il est confronté.	COMP-PRAG
	Interprétation compositionnelle	L'informateur fait une interprétation mot-à-mot du PhPex qui lui est proposé.	INTER-COMPO
	Analogie erronée avec la LM	L'informateur se fonde sur sa compétence pour interpréter une expression proche sur le plan linguistique avec sa LM (faux ami), ce qui l'amène à commettre une erreur de compréhension.	ERROR-LM
	Ajustement registre	L'informateur ne parvient pas à adapter son registre de manière optimale à la situation de communication qu'il rencontre ou au type d'interlocuteur avec lequel il interagit.	AJUST-REG

Difficultés stratégiques	Conflit culturel	L'informateur ne modifie pas sa stratégie de communication, car celle-ci entre en conflit avec sa manière de communiquer héritée de sa langue maternelle.	CONF-CULTU
	Inadéquation illocutoire	L'informateur adopte une stratégie pragmatique qui est inadéquate ou non-naturelle dans la situation de communication qu'il rencontre.	ILLOC
	Inadéquation stratégie communicative	L'informateur adopte une posture communicative ou un comportement au sein de l'interaction qui est inadéquat dans la situation de communication qu'il rencontre.	STRAT-COMM
Codage des stratégies			
Stratégies en production	Emprunt à la LM	L'informateur produit un énoncé emprunté à sa LM pour pallier une lacune lexicopragmatique.	EMP-LM
	Analogie à la LM	L'informateur produit un énoncé proche de sa LM pour pallier une lacune lexicopragmatique.	ANALOG-LM
	Adaptation à l'interlocuteur	L'informateur produit un énoncé en fonction du statut de son interlocuteur.	ADAPT-INTER
	Adaptation à l'illocutoire attendu	L'informateur produit un énoncé cohérent avec la fonction pragmatique attendue.	ADAPT-PRAG
	Adaptation à la situation	L'informateur produit un énoncé qui prend en compte les spécificités du contexte d'interaction.	ADAPT-SITU
Stratégies en compréhension	Analogie avec le lexique de la LM	L'informateur se base sur les connaissances lexicales de sa LM pour interpréter le sens du PhPex.	ANALOG-LEX
	Analogie pragmatique avec la LM	L'informateur se base sur les connaissances pragmatiques de sa LM pour interpréter le sens du PhPex.	ANALOG-PRAG
	Traduction vers la LM	L'informateur traduit le contenu lexical du PhPex vers sa LM pour interpréter le sens.	TRAD-LM
	Indice contextuel	L'informateur repère des indices contextuels pour interpréter le sens du PhPex.	INDIC-CONTEXT
	Indice prosodique	L'informateur repère des indices prosodiques pour interpréter le sens du PhPex.	INDIC-PROSO
	Indice stratégique	L'informateur repère des indices liés aux stratégies communicatives et aux comportements adoptés par les participants de l'interaction pour interpréter le sens du PhPex.	INDIC-STRAT

Tableau 24 : Codage thématique des relevés faits sur les entretiens semi-directifs

À partir de ce codage, nous avons dégagé plusieurs catégories thématiques qui se groupent en deux dimensions : les difficultés liées à l'usage des PhPex et les stratégies mobilisées dans la compréhension. Nous avons ensuite fait une synthèse de phénomènes fréquemment évoqués dans les entretiens pour chaque dimension. Nous avons enfin fait un dernier travail de synthèse qui nous a permis d'extraire plusieurs points clés que nous allons évoquer dans la synthèse des entretiens. Ces points clés apportent des éclairages aux analyses faites dans la phase 1 et 2 de notre étude. Il sera utile de rappeler que la synthèse qui sera proposée dans les pages qui suivent relève principalement des déclarations des LNN ayant participé à notre étude et ne saurait être représentative de l'ensemble des locuteurs japonophones avancés du français. Cette synthèse peut en revanche constituer un point de départ à d'autres études de ce type.

Chapitre 14 - Synthèse des discussions métapragmatiques

Dans ce chapitre, nous allons détailler les thématiques saillantes qui sont ressorties du traitement des données d'entretien. Nous focalisons notre attention essentiellement sur le fil conducteur de notre thèse à savoir la caractérisation des difficultés d'usage des PhPex chez les LNN japonophones de niveau avancé en français. La discussion sur les stratégies en compréhension et production ne fera pas l'objet d'une section spécifique, mais sera abordée tout au long de la synthèse des difficultés. Malgré les précautions prises et étant donné le caractère relativement subjectif des données recueillies, nous avons écarté volontairement un certain nombre de contenus, certes intéressants pour la compréhension des spécificités d'usage du français et de la communication orale chez ce type de locuteurs, mais secondaire dans le cadre de notre étude. En outre, nous avons été confronté à un problème double : d'une part, il est apparu que les LNN interrogés avaient pour une bonne partie tendance à adopter une posture d'écoute et ne semblaient pas être enclins à prendre la parole spontanément. Il y a donc une certaine dissymétrie qui s'est installée concernant l'occupation de la parole. D'autre part, cette dissymétrie a été accentuée par notre tendance d'enquêteur débutant à prendre la parole de manière trop fréquente. Cela étant, ce protocole d'élicitation métapragmatique fondé sur une approche semi-directive a donné des résultats intéressants que nous allons détailler ci-dessous.

1. Les difficultés de discernement de sens phraséologique et littéral

Une des thématiques fréquemment annotées dans les entretiens est la difficulté à différencier le sens phraséologique du sens littéral (PHRASEO-LITT) difficulté évoquée dans la section 6.1.2.1 de l'analyse du test lexicopragmatique. Cela est d'autant plus vrai pour les expressions *c'est pas gagné !*(Activité 3, reformulation), *c'est le bouquet !* (Activité 4, QCM lexical) et *les bras m'en tombent !* (Activité 5, QCM sémantique) mentionné par 77% des LNN comme étant les expressions les plus problématiques. Cette difficulté avait été remarquée lors du traitement des résultats au test lexicopragmatique. Deux raisons sont fréquemment évoquées lors des entretiens : 1) les unités qui composent les expressions sont non prédictibles et ont peu d'ancrage avec le contexte, 2) les LNN utilisent une stratégie de traduction qui bien souvent débouche sur une interprétation erronée. Dans la majorité des cas, la difficulté était donc liée au fait que ces expressions sont non compositionnelles. Les

LNN s'appuient alors sur les constituants pour tenter de déterminer au vu des éléments du contexte la fonction du PhPex. Toutefois, si l'on prend l'exemple de l'expression *c'est le bouquet !*, les LNN ayant eu des difficultés sur cette expression ont tous rapporté avoir spontanément pensé au bouquet de fleurs et notamment à l'image positive que véhicule celui-ci. Cette association prématurée a vraisemblablement engendré une erreur sur l'interprétation de la polarité de l'expression. L'expression *c'est pas gagné !* a posé une difficulté supplémentaire, celle de l'intraductibilité. Nous l'avions vu dans le chapitre précédent, cette expression a suscité de grandes difficultés chez le LNN, car il n'existe à proprement parler pas d'équivalent fonctionnel direct en japonais. Les entretiens métapragmatiques nous ont permis de constater qu'en plus de ce manque d'équivalence fonctionnelle qui demeure problématique, il semblerait y avoir le fait que ce type d'emploi reste marginal en japonais. Les informateurs nous ont plusieurs fois confié qu'ils étaient rarement amenés à rencontrer ce genre de schéma conversationnel, ce qui a fait qu'ils ne savaient pas vraiment quoi répondre en japonais. Enfin, la difficulté majeure pour *les bras m'en tombent !* était son sens très opaque qui ne permettait pas au LNN de déduire le sens de surprise négative à partir de la gestuelle des bras tombants.

2. Difficulté liée à des analogies erronées avec la langue maternelle

Les entretiens menés auprès des LNN ont mis en évidence des difficultés éprouvées de manière systématique pour la compréhension de l'expression *c'est le pied !*. Cette difficulté de compréhension a été remarquée chez 80% des informateurs, dont ceux qui présentaient les résultats les plus élevés au test lexicopragmatique. Dans la plupart des cas, cette difficulté a été évoquée d'emblée et spontanément, sans intervention particulière de notre part. Ce cas semble à première vue relever d'une difficulté à différencier le sens phraséologique du sens littéral, comme dans le précédent cas de *les bras m'en tombent !*. Toutefois, si l'on observe les résultats obtenus pour ce PhPex (Activité 2, dialogue 1, item 1), on s'aperçoit que le taux de réussite est élevé (77%), ainsi que le taux de compréhension (91%), ce qui semble aller à l'encontre de cette difficulté évoquée en masse par les LNN. En regroupant les relevés effectués concernant cette difficulté, nous nous sommes aperçu que les codages ERROR-LM, ANALOG-LEX et TRAD-LM revenaient fréquemment. L'analyse qualitative des entretiens a effectivement confirmé que la difficulté se trouvait en compréhension et notamment en lien avec le recours à la LM pour interpréter *c'est le pied !*. La discussion

métapragmatique a donné lieu à une interaction nourrie au cours de laquelle nous avons remarqué que les LNN associaient de prime abord l'expression à une polarité négative. Ainsi, dans la grande majorité des cas, les LNN ont confirmé qu'ils avaient une tendance forte à interpréter de manière négative l'expression *c'est le pied !*, puis ont finalement changé leur réponse pour une réaction positive après avoir observé le contexte plus en profondeur. La discussion a ensuite montré que le motif du pied est considéré au Japon comme une zone du corps fondamentalement associé à l'impureté, car proche du sol qui lui-même est considéré comme souillé. En outre, le pied désigne métaphoriquement la position basse, rappelant une certaine position d'infériorité et de soumission, la position proche du sol étant fréquemment associée à une dépréciation de soi[193]. La difficulté semble donc double ici : elle est imputable à 1) la stratégie de traduction et 2) l'analogie sémantique avec le motif du pied en japonais, stratégies palliatives qui ont été mobilisées pour interpréter l'emploi du PhPex inconnu au moment du test[194].

Un autre exemple est un commentaire laissé par l'un des informateurs dans le questionnaire en japonais (Informateur UR5FJP042020). Dans ce commentaire[195], le locuteur évoque l'emploi du marqueur de ratification *je vois* et du marqueur discursif d'étayage *tu vois ? / vous voyez ?*. Malgré la relative facilité pour le locuteur d'utiliser le premier, le second semble poser grandement problème. En effet, bien que ce locuteur connaisse les modalités d'usage de *tu vois ? / vous voyez ?*, et bien qu'il emploie fréquemment des équivalents de ces marqueurs en japonais, il ne parvient pas à employer ce marqueur à l'oral en français. Une des raisons principales selon le LNN serait qu'à la place de la fonction d'étayage, il pense qu'il traduit malgré lui vers sa LM l'expression de manière compositionnelle et

---

[193] Cf. l'exemple du 土下座 (dogeza) qui consiste à s'excuser en s'agenouillant au sol et en posant sa tête sur le sol pour s'excuser d'une faute grave, en signe d'humilité totale.

[194] Fait intéressant, la cohorte des LNN compte un locuteur métisse franco-japonais parfaitement bilingue dans les deux langues, mais ayant principalement vécu au Japon. Il a rapporté connaitre l'expression, mais avoir cependant mis du temps à se remémorer le sens. Instinctivement, il avait lui aussi pensé en premier lieu à un sens à polarité négative, et ce en raison de l'influence culturelle japonaise qui selon lui est prédominante dans son cas.

[195]"je vois"とは言えるのに、"tu vois?/vous voyez?"は意味も使うタイミングも知っているのに、口に出てこない。日本語でも例えば「でしょ？」「ね？」「～じゃん？」など同じような働きをするものは頻繁に使うが、フランス語でこれらの表現は使えない。理由：tu vois ?/vous voyez?を「わかる？」さらには「見える？」と頭の中で直訳しているかもしれない。そもそも日本語で「わかる？」のように直接的に理解を確認するようなことはあまりなく、上に挙げた接尾辞などに含めてしまうことの方が多い気がする。

notamment le verbe *voir* qui en japonais ne permet pas d'exprimer la fonction d'étayage en tant que marqueur discursif inséré dans le discours.

3. Difficultés liées au repérage de la fonction pragmatique

Les informateurs ont fréquemment déclaré lors des entretiens la difficulté à repérer ou interpréter les fonctions pragmatiques des PhPex proposés. Cette même difficulté avait été repérée dans l'analyse du test à la section 6.1.2.3. Le retour sur les données de l'activité 4 (QCM lexical) a montré que les expressions *c'est le pied ! c'est le bouquet ! ça tombe bien ! tu crains !* et *tu parles !* ont été perçues comme problématiques, car peu prédictibles au niveau de leur fonction pragmatique. La traduction compositionnelle ou l'analogie avec la langue maternelle n'a pas permis aux informateurs de prédire le sens et la fonction de ces items. D'autres informateurs avaient déclaré s'être posé la question de savoir si les emplois de l'activité 4 n'impliquaient pas de dimension ironique. Ce motif de l'ironie et du sarcasme a souvent été abordé par les informateurs comme l'une des fonctions pragmatiques représentatives du français. Cette conception du français a donné lieu à des phénomènes de sur-interprétation dans l'activité 4 qui ont eu une influence dans le repérage de la fonction pragmatique adaptée aux dialogues proposés. Un autre exemple est la mise en situation de l'expression *c'est énorme !* (cf. Guide d'entretien) dans son sens de réaction évaluative positive. Nous avions lors de l'entretien lu nous-même le dialogue tout en prenant soin d'adopter une intonation expressive la plus naturelle possible. Malgré cette précaution, et le contexte relativement simple à interpréter, une bonne partie des informateurs ont opté pour le sens quantitatif de *c'est énorme !*. Après discussion avec eux sur les stratégies qui ont motivé ce choix, ceux-ci ont évoqué le fait qu'ils n'arrivaient pas à associer le terme *énorme* à la réaction expressive de joie et à une évaluation positive. Par ailleurs, nous nous sommes aperçu que ceux-ci n'avaient pas identifié le marqueur intonatif et la spécificité du contexte où le locuteur exprime sa joie pour son ami. Ils se sont essentiellement focalisés sur le lexique employé.

4. Difficulté d'accommodation du registre au contexte de communication

Une des constatations que nous avions faite dans l'analyse du test lexicopragmatique concernait le fait que les répondants avaient tendance à employer préférentiellement un

lexique représentatif du type d'input auquel ils étaient le plus souvent exposés lorsqu'ils emploient le français. Cet usage préférentiel se traduisait par exemple chez les LNN employant principalement le français en milieu académique par une propension plus forte à employer des tournures standards là où étaient en principe attendues des tournures familières et plus spécifiques de l'oral. Cette difficulté avait été remarquée dans l'analyse du test lexicopragmatique (Chap. 12, section 6.1.1), et les mesures de variabilité du registre avaient également montré chez les LNN des taux d'emploi du registre courant en augmentation pour le français par rapport au LN. Ce lien a été de nombreuses fois évoqué dans les entretiens, notamment dans la première phase de l'entretien, lors du retour sur les réponses au test. Dans la majorité des cas, il en est ressorti que les LNN attribuaient certaines de leurs difficultés à produire un énoncé familier d'une part à une exposition insuffisante à la conversation quotidienne, d'autre part, à une sous-utilisation des expressions proposées dans les dialogues du test. La discussion sur ces différents points a notamment montré que les LNN étaient conscients du fait que leurs habitudes de communication développées dans leur contexte de socialisation linguistique principal influençaient grandement le lexique employé lors de la communication orale quotidienne. Cette influence du contexte d'exposition s'est ainsi traduite par les réponses suivantes à l'activité de mise en situation en production (cf. guide d'entretien) :

- *c'est terrible !* (Mise en situation 3)
- *qu'est-ce que vous faites?* (Mise en situation 1)
- *ce n'est pas possible !* (Mise en situation 4)
- *oh quelle tragédie !* (Mise en situation 4)
- *mon dieu !* (Mise en situation 3)
- *pourquoi ?* Pour quelles raisons ? (Mise en situation 1)

Ces réponses ne sont pas inacceptables et pourraient très bien convenir sur le plan informationnel. Elles montrent par ailleurs que les LNN ont saisi le contexte de communication. En revanche, ces réponses semblent inadaptées à l'illocutoire et à l'expressivité attendus (une réponse adaptée serait *mais ça va pas la tête !* dans le cas de la situation 1) et peuvent sembler trop standards pour la situation. L'informateur KM2FJP052020 a même évoqué une anecdote concernant ses amis qui avaient tendance à lui dire qu'il « parle comme un manuel » tant il avait tendance à employer un lexique

standard. Dans les réponses au questionnaire, le locuteur NM3FJP042020 mentionne un fait similaire montrant la conscience du suremploi du registre standard (*teinei sugiru kotobadzukai o suru = j'emploie un lexique trop poli*). Enfin, d'autres locuteurs ont déclaré lors de l'entretien métapragmatique que leurs amis n'utilisaient pas souvent d'expressions familières lorsqu'ils étaient avec eux et que de ce fait, ils ne rencontraient pas beaucoup d'occasions de se familiariser à l'emploi des expressions de la conversation quotidienne. Ces observations faites lors de l'entretien semblent constituer des arguments supplémentaires concernant notre hypothèse formulée dans le diagnostic et validée par la variable *statut professionnel* (Chap. 11, section 2.2.1) mesurée dans le test lexicopragmatique. Nous avions en effet avancé que l'exposition à des situations de communication variées pouvait avoir un effet positif sur le développement de la compétence lexicopragmatique. La variable *statut professionnel* avait alors montré une corrélation significative entre augmentation des résultats au test et exposition au français en immersion professionnelle.

5. Difficultés liées aux objectifs de communication

Une thématique souvent discutée lors des entretiens métapragmatiques concerne les difficultés liées à l'objectif de communication et au besoin de développer sa palette lexicopragmatique. En effet, certains des informateurs ont évoqué lors des entretiens le fait que le français qu'ils emploient était principalement ciblé en fonction de leurs besoins communicatifs. Par exemple, l'informateur KM2FJP052020 (actuellement en doctorat en France) nous a expliqué qu'il n'avait jamais vraiment eu besoin d'apprendre le lexique propre à la communication quotidienne puisqu'il communique essentiellement dans le cadre académique avec des gens qu'il ne côtoie pas en dehors de son activité de recherche. L'informateur MN33FJP062020 qui est enseignant de japonais dans un collège japonais pour expatrié situé à Paris nous a également confié que bien qu'il vive en France depuis longtemps, son cercle familial et professionnel se constitue essentiellement de locuteurs natifs du japonais. Ce faisant, ce LNN n'interagit que très peu avec des LN dans la vie quotidienne, ce qui diminue fortement ses chances d'être exposé aux PhPex. Il nous a donc semblé ici que les besoins communicationnels immédiats sont indirectement source de difficulté dans le développement de la compétence lexicopragmatique.

6. Difficulté en production liée à l'asymétrie entre compétence de compréhension et diversité du répertoire lexicopragmatique

L'étayage conjoint des résultats de l'activité 3 (équivalence) du test lexicopragmatique a montré que la grande majorité des LNN disposaient d'une capacité d'analyse du contexte et du fonctionnement pragmatique des PhPex relativement avancée, mais n'étaient pas en mesure de proposer un équivalent en français du fait d'un manque de lexique. De plus, nous avons pu mener des discussions métapragmatiques fines sur des aspects contrastifs avec certains des informateurs sans pour autant qu'ils n'aient été en mesure de proposer une expression adéquate dans les mises en situation de production proposées dans la phase 2 de l'entretien. Pour appuyer cette constatation, nous pouvons mentionner par exemple le cas de l'informateur MR24FJP062020 lors de la passation de la mise en situation de production 3. Celui-ci, très hésitant dans sa réponse, va proposer *ah ! zut !* puis ajouter *je n'ai pas vocabulaire* pour mentionner le fait qu'il ne disposait pas du lexique suffisant pour proposer une expression adéquate. Une relance lui demandant de proposer une réponse en japonais va montrer qu'il avait effectivement bien saisi les spécificités du contexte et le type de PhPex attendu.

D'autres éléments de réflexion concernant les difficultés de production liées une palette lexicale limitée se trouve dans l'enquête libre distribuée après le test. Certains des informateurs ont mentionné le fait que lorsqu'ils interagissent avec un LN français et qu'ils doivent réagir spontanément, la lacune lexicale est généralement palliée par l'emploi d'une réaction vocale ou linguistique emprunté à la LM. Ce phénomène est par ailleurs plus fréquent si le LNN sait au préalable que son interlocuteur dispose de compétence en japonais. Par exemple, le locuteur OT28FJP042020 a mentionné dans l'enquête le fait que lorsqu'il était surpris de manière soudaine, il avait tendance à utiliser de l'alternance codique[196] et à répondre en japonais.

---

[196] Selon les termes de l'informateur : « コードスイッチングしてしまう » (*kôdo suicchingu shite shimau*) = j'utilise malgré moi l'alternance codique.

7. L'influence de l'éthos communicatif sur la production en LC

L'influence de la façon de communiquer (ou *éthos communicatif*) a été observée dans les réponses au test lexicopragmatique sous la forme de difficultés d'ajustement de la stratégie communicative et au niveau de la sélection du lexique adéquat. Dans la phase diagnostique, il s'agissait par exemple d'un suremploi du rire ou des marques phatiques. Cette influence, parfois non consciente, a beaucoup été discutée avec les LNN lors des entretiens métapragmatiques.

Un premier exemple de cette influence de la façon de communiquer sur le type de réponse produite en LC sont les réponses proposées par certains des informateurs à la mise en situation en production. Sur les 5 dialogues proposés (cf. guide d'entretien), deux des dialogues ont fréquemment suscité une absence de réponse chez les LNN. Il s'agit notamment du dialogue concernant le chien violenté par son maitre et celui concernant la bousculade par un vélo où il était attendu une réaction d'indignation ou de condamnation. En questionnant les informateurs sur les raisons de cette absence de réponse, nous avons pu constater que l'éthos communicatif de ces locuteurs faisait qu'ils n'étaient pas enclins à réagir face à ce genre d'événement. Par exemple l'informateur IM7FJP042020 formulera cette constatation de la manière suivante : *au Japon, on ne s'intéresse pas comme ça aux gens dans la rue*. Toutefois, il sera précisé lors de l'entretien que l'informateur ressent de l'indignation et de la colère, mais ne l'exprime pas. Le même phénomène est observable pour la scène de bousculade ou un certain nombre de LNN nous explique qu'ils ne diraient instinctivement rien, ni en français ni en japonais. En poussant la discussion un peu plus loin, l'informateur ANH14FJP052020 évoquera le fait que spontanément, dans ce type de situation, un locuteur japonophone aura plus tendance[197] à être coopératif et non pas indigné. Cette déclaration sera confirmée par un l'informateur HK17HJP062020 qui suite à une relance proposera en japonais la réponse suivante : *a ! sumimasen* (*excusez-moi !*). A cette réponse, le locuteur ajoutera qu'il opterait plutôt pour une posture coopérative que conflictuelle, et ce même s'il n'est pas en tort, d'où l'emploi de l'excuse.

---

[197] Nous rappelons ici que les constatations où les hypothèses mentionnées ici ne sont en aucun généralisables et se fonde de manière empirique sur la synthèse des éléments évoqués par les informateurs lors des entretiens, et non pas sur notre intuition.

Un autre exemple intéressant est observable dans les réponses proposées pour la mise en situation de production n°2 (soda renversé sur le T-shirt). Nous avions volontairement précisé la proximité de la relation ici pour observer si cet élément serait spontanément pris en compte lors de la réponse. La réaction attendue ici était soit une surprise négative, soit du reproche. Dans la plupart des cas, les informateurs ont répondu par *ce n'est pas grave* ou *t'inquiète* en français. Très peu de réponses ont montré l'usage d'un reproche ou d'une réaction exclamative comme attendu pour ce dialogue On remarque donc chez les LNN une tendance à viser la coopération et l'atténuation plutôt que le reproche ou la déception comme ce serait plus souvent le cas, nous semble-t-il, en français[198]. Cette constatation sera confirmée dans un premier temps par les relances aux réponses en français au cours desquelles nous avons demandé aux informateurs de proposer une expression en japonais. Les expressions proposées étaient principalement *ki ni shinaide ne* (*ne fais pas attention*) ou *daijôbu* (*pas de problème*).Toutefois, même si la stratégie communicative préférentiellement adoptée par les informateurs tend à la convergence et la minimisation, il n'en demeure pas moins que certains locuteurs vont tout de même ressentir une certaine déconvenue suite à l'accident, sans pour autant l'exprimer, comme nous le montre l'informateur IM18FJP052020 dans la citation suivante : *honshin wa chigakutemo toriaezu 「daijôbu da yo」 to iu* »[199] (*même si j'en pense autrement, je dirais au préalable 'pas de problème'*). Cette prudence à l'expression de ses sentiments lorsque la clarification de ceux-ci peut avoir un impact négatif sur la situation de communication semble être un trait commun des locuteurs japonais plusieurs fois soulevé lors des entretiens et dans le questionnaire[200]. À plusieurs reprises, certains informateurs ont en effet mentionné lors de l'entretien et dans le questionnaire[201] le fait que selon eux, *les Japonais n'expriment pas spontanément leurs sentiments*. Ces différentes caractéristiques propres à la façon de communiquer de la LM semblent être à l'origine de certaines difficultés à employer les réactions expressives attendues dans les mises en situation proposée lors de l'entretien. Un autre commentaire intéressant est celui de l'informateur NM3FJP042020 présent dans l'enquête. Il évoque le fait que bien qu'il maitrise l'emploi interjectif de *putain*, celui-ci éprouve tout de même

---

[198] Cette hypothèse mériterait toutefois d'être confirmée auprès d'une cohorte de locuteurs natifs du français.
[199] 本心は違くても取り合えず「大丈夫だよ」と言う
[200] « 自分の感情を抑えて周りに合わせたり意見するのを躊躇することがある[…] »
[201] « 日本語では言葉に出さなくても察してもらおうとする文化があるが、フランスでは嫌なものは嫌と言うし、感情表現もストレートだと思う。 »

beaucoup d'hésitation à utiliser ce type de terme[202], et ce en raison du conflit que l'usage provoque avec sa façon de communiquer. D'autres locuteurs mentionnent dans le questionnaire l'usage trop fréquent des marques de politesse[203] en français telles que l'excuse (*jibun ni hi ga naï noni pardon, excusez-moi o renpatsu suru*) par rapport à ce qui est habituellement employé dans la conversation quotidienne en français. Cette observation confirme la tendance de suremploi des stratégies communicatives de la LM observé dans le diagnostic concernant l'emploi des marques phatiques.

Cette section conclut donc cette analyse des données obtenues à l'issue de notre protocole exploratoire de la compétence lexicopragmatique en trois phases. Elle nous a permis d'approfondir certaines des observations formulées dans la deuxième étape de notre protocole expérimental notamment concernant les difficultés principalement rencontrées lors de la production des PhPex. La méthode de traitement et d'analyse des données d'entretien présentée dans cette dernière phase est encore perfectible. En effet, il est toujours un peu difficile de concilier l'analyse linguistique rigoureuse avec la matière complexe et subjective du relevé d'entretien. Par ailleurs, ce type d'approche nécessite une certaine expérience de la préparation et de la gestion d'un entretien que nous n'avions pas au moment du recueil. Toutefois, nous avons tenté ici de proposer une méthodologie cadrée, offrant la rigueur et la systématicité nécessaires à nos objectifs de recherche. L'originalité de cette approche est qu'elle vise avant tout à réhabiliter le point de vue de l'informateur dans le traitement des données, point de vue souvent négligé dans les études sur corpus de locuteurs non-natifs. Dans les pages qui vont suivre, nous allons retracer de manière synthétique tout le processus expérimental et les avancées principales obtenues concernant la compétence lexicopragmatique et l'emploi des PhPex, puis nous conclurons cette thèse sur les continuités possibles de ce travail de recherche.

---

[202] « 最近やっと慣れてきたが、Putain という語を発する時に少しためらってしまっていた。 »
[203] « フランス語の désolée を日本語の「ごめんごめん」の感覚で使っていたら、どうしてそんなに謝るの？と言われた。フランス語の désolée は日本語より意味が強く、フランスではちょっとしたことでは謝らないと言われた »。

Conclusion

Ce travail de recherche a tenté d'apporter des éléments d'analyse concrets concernant les spécificités d'usage des PhPex chez des LNN japonophones locuteurs du français de niveau avancé. Ces éléments d'analyse se fondent d'une part sur une réflexion méthodologique et épistémologique concernant le traitement de la compétence lexicopragmatique mobilisée dans l'usage des PhPex, d'autre part sur une méthodologie rigoureuse et basée sur les données. Nous proposerons ici une synthèse générale des résultats obtenus à l'issue d'un protocole expérimental en trois phases. Nous évoquerons ensuite certaines des limites rencontrées. Nous conclurons enfin sur les enjeux de ce domaine de recherche et les pistes qui seront explorées à la suite de ce travail de recherche.

1. Synthèse des résultats

La première phase diagnostique du protocole a permis de proposer un premier état de la compétence lexicopragmatique des locuteurs japonophones de niveau intermédiaire du japonais. Le but de cette première phase était de formuler des hypothèses exploratoires fondées sur l'observation de données authentiques prélevées en contexte d'interaction exolingue. Ces hypothèses concernant les difficultés rencontrées et les stratégies employées par les LNN en production nous ont servi de point de départ à la deuxième phase de notre protocole. Au cours de cette première phase, nous avons pu constater plusieurs résultats intéressants :

- Les LNN de notre étude emploient moins d'unités à fonctions pragmatiques que les LN du corpus de contrôle dans des conditions similaires d'interaction libre, notamment les marques complexes servant à exprimer l'intersubjectivité, à connecter les différents éléments (connecteur) de l'interaction entre eux, à moduler son assertion (marque épistémique) ou les PhPex. On a également retrouvé peu d'emploi des routines interactionnelles (*tu vois, tu sais, du coup*) ou de reformulations (*enfin, je dirais, je sais pas moi*) fréquemment employées chez les LN du corpus endolingue.

- Le rire et le recours à la langue maternelle (emprunts et transferts) étaient fréquemment employés pour pallier un manque lexicopragmatique chez le LNN. Un

exemple révélateur est le suremploi de marques phatiques largement supérieur à ce qui est observé dans le corpus endolingue.

- Les LNN éprouvaient certaines difficultés à adapter l'illocutoire visé à la situation de communication, engendrant parfois un décalage de la posture communicative avec celle qui est attendue. En outre, on observe l'influence de l'éthos communicatif sur la manière de communiquer en langue cible (ex. : faible négociation de la parole et usage de rires sociaux caractéristiques du japonais).

- L'objectif d'apprentissage et la motivation à s'engager dans la communication sont apparus comme des facteurs de développement de la compétence lexicopragmatique. On pensera notamment au cas du LNN de Tsukuba qui en dépit de son niveau certifié A2 a une certaine aisance à l'oral, car très investi dans son apprentissage et dans la communication.

La seconde phase de notre protocole est le test lexicopragmatique. Nous avons pris comme point de départ les observations faites dans la phase diagnostique, puis nous avons tenté à l'aide de ce test d'étendre notre analyse de la compétence lexicopragmatique mobilisée dans la production et la compréhension des PhPex à des locuteurs de niveau avancé. Ce test a été conçu de manière à pouvoir évaluer la production et la compréhension. En outre, nous avons intégré une dimension contrastive à cette étape (cf. Activité 3, équivalence fonctionnelle). Elle nous a permis de repérer les différences entre les deux systèmes, les stratégies d'interprétation fondées sur la LM, mais aussi de comprendre les difficultés d'emploi liées à la distance typologique entre les deux langues. Ce test combine plusieurs types de protocoles d'élicitation de la production et de la compréhension, et ce afin d'explorer trois niveaux : la compétence pragmalinguistique, la compétence sociopragmatique et la compétence métapragmatique. À l'issue de l'analyse des résultats obtenus dans cette deuxième phase du protocole, nous avons pu dresser une typologie des difficultés :

⇒ En production :

- Difficultés d'ajustement du registre en fonction des spécificités de la situation de communication.
- Difficulté d'ajustement de l'intensité expressive et de la modalité.

- Difficulté concernant les choix combinatoires qui parfois actualisent un illocutoire en décalage avec le dialogue proposé.
- Des erreurs sur le plan formel qui peuvent influencer la compréhension de la fonction communicative.
- Inadéquation de la visée pragmatique ou de la stratégie communicative employée par les LNN avec l'illocutoire attendu dans le dialogue proposé.

⇒ En compréhension :

- Difficulté à discerner le sens phraséologique du sens littéral.
- Amalgame entre fonction illocutoire et effet perlocutoire.
- Difficulté à différencier deux PhPex proches au niveau de leur fonction.
- Difficulté de compréhension concernant les PhPex polyfonctionnels.

La troisième étape de discussion métapragmatique a permis d'affiner les analyses menées sur les résultats du test. Nous avons procédé lors de ces entretiens à des retours sur les réponses proposées dans le test par le LNN puis nous avons discuté des stratégies de sélection et d'interprétation des PhPex proposés. Cette ultime étape de notre protocole nous a permis d'approfondir du point de vue du LNN le fonctionnement de la compétence lexicopragmatique et la manière dont interagissent les trois sous-compétences qui lui sont liées. En outre, nous avons proposé lors de cet entretien quelques activités d'élicitation dans lesquelles nous mettions en situation les LNN puis proposions ensuite une discussion de niveau métapragmatique sur les choix opérés en production et en compréhension. Ces différents points de vue du locuteur sur sa production au test et sur ses stratégies mobilisées lors des mises en situation ont permis de confirmer certaines des observations faites dans la phase 2 parmi lesquelles :

- La difficulté à discerner le sens phraséologique du sens littéral.
- La difficulté de compréhension et de production liée à des analogies erronées avec la langue maternelle.
- La difficulté de repérage et d'ajustement de la fonction pragmatique.
- La difficulté d'accommodation du registre.
- Le lien entre développement de la compétence lexicopragmatique et l'objectif de communication.

- La difficulté de production liée à l'asymétrie entre compréhension et ressources lexicales nécessaires pour exprimer une fonction communicative.
- L'influence de l'éthos communicatif du LNN sur sa production et le développement de sa compétence lexicopragmatique en LC.

Ce protocole expérimental nous a permis d'observer un niveau de granularité fine de la compétence lexicopragmatique en croisant d'autres protocoles déjà existants sous trois modalités d'élicitation : semi- élicitée (Phase 1), fortement élicitée (Phase 2) et élicitée par le locuteur (Phase 3). En outre, les démarches outillées et les méthodologies de formalisation employées dans les trois phases de ce test nous ont permis d'établir des mesures statistiques et quantitatives intéressantes pour caractériser l'emploi des PhPex chez les LNN observés. Nous avons ainsi pu observer qu'une baisse de l'indice phraséologique associé à une forte diversité lexicale, un taux de compréhension élevé et un score de réussite faible pouvait caractériser une difficulté en compréhension ou en production liée à l'inexistence d'un équivalent fonctionnel dans la LM du LNN (cf. Activité 3, équivalence fonctionnelle, PhPex *c'est pas gagné !*). Nous avons également pu observer une corrélation statistique significative entre les résultats observés dans le test et les variables liées au milieu d'usage du français (*professionnel* ou *académique*), la durée de séjour en France et la durée d'apprentissage. L'approche par entretien semi-dirigé constitue une perspective qui nous a semblé pertinente pour confirmer à l'aide du point de vue du LNN les hypothèses que nous avons formulées au préalable sur les processus de sélection et d'interprétation mobilisés lors de l'usage des PhPex. Les entretiens constituent selon nous une démarche utile pour limiter la subjectivité de l'analyse des résultats. Un dernier aspect intéressant de notre protocole expérimental sont les différents dispositifs de vérification des résultats mobilisés. En effet, notre protocole a croisé à la fois deux procédures. La première procédure constitue un accord interannotateur mis en œuvre pour valider la typologie d'annotation employée dans la phase 1 de notre protocole. La deuxième procédure constitue un accord interjuge pour valider les jugements d'acceptabilité portés sur les résultats des LNN au test et les catégorisations effectuées dans le cadre de la mesure des statistiques lexicales de la phase 2.

2. Limites et améliorations possibles de notre approche

Toute innovation passe inévitablement par la confrontation à un certain nombre de limites et d'améliorations nécessaires. Notre protocole ne déroge pas à cette constante et a donné lieu à un certain nombre de limites qui ont été mentionnées tout au long de l'analyse. Parmi les plus représentatives, on observe celles liées aux aspects techniques de la mise en œuvre du protocole. En effet, il n'existe à notre connaissance aucune approche de ce type qui ait été menée sur la problématique de l'acquisition du lexique préfabriqué de la conversation quotidienne chez des LNN de niveau avancé n'étant pas nécessairement engagé dans un cursus d'apprentissage du français. Ce faisant, nous n'avons pas pu nous appuyer sur des études antérieures pour élaborer notre protocole. La construction de notre test en elle-même a donc un caractère expérimental.

Concernant les limites rencontrées dans le recueil des observables elles sont de plusieurs ordres. Tout d'abord, nous avons été confronté à la faible disponibilité des données orales de type conversation quotidienne, données qui sont encore moins disponibles en situation de communication exolingue non élicitée. Nous avons donc été dans l'obligation de produire ces données nous-même. En outre, quand ces données sont disponibles, elles sont bien souvent d'une qualité trop faible pour être utilisées telles quelles. Il y a donc tout un travail de nettoyage et d'adaptation très chronophage qui doit être fait (cf. Wang, 2021). Quand ce n'est pas les données orales qui se font rares, ce sont les locuteurs testés qui sont difficiles à recruter. Dans notre cas, réunir suffisamment de locuteurs japonophones ayant un niveau avancé en français et répondant aux critères de sélection a été un processus chronophage. Enfin, dans le cadre de la phase 1 de notre protocole et notamment de l'enregistrement de corpus de conversation exolingue semi-élicité, nous nous sommes aperçu qu'il était très complexe de recréer les conditions naturelles d'une interaction lorsque le chercheur est présent (cf. *Paradoxe de l'observateur,* Labov 1973).

Concernant le traitement des données, les nombreuses entrées d'analyse que nous avons mobilisées ont nécessairement engendré un travail d'annotation et de formalisation chronophage. D'autres types de formalisations telles que l'anonymisation et la mise en place des accords interannotateur et interjuge ont demandé un temps de conception et de préparation relativement long. Dans ce dernier cas, le couple de langue français-japonais étant relativement rare, il n'a pas été simple de trouver des locuteurs experts dans les deux

langues et ayant une expertise en sciences du langage suffisante pour mener la tâche de jugement d'acceptabilité.

Enfin, concernant la phase 2 de test lexicopragmatique, nous nous sommes aperçu a posteriori de plusieurs limites lors de l'analyse des données. La première limite concerne le format de la passation. Nous avons en effet constaté que le temps de réponse ou les conditions de passation auraient nécessité quelques ajustements supplémentaires pour permettre une passation plus confortable. À titre d'exemple, la dernière activité (Activité 5, QCM sémantique) qui est assez exigeante en ressources cognitives aurait nécessité un temps de réponse plus long et une deuxième écoute pour permettre de faciliter la réponse aux questions. Cependant, cette amélioration pose également le problème de la durée du test qui, lorsque celui-ci est trop long, peut engendrer des phénomènes de *speeding* et augmenter le taux de réponses aléatoires dans le test. La deuxième limite concerne le choix d'activité et son adéquation avec les conditions de passation. Plus particulièrement, nous nous sommes aperçu que le dosage de la difficulté peut être relativement complexe, quand bien même un test pilote aurait été mené. Nous avons notamment pu observer cela lors de l'analyse de l'activité 4 (QCM lexical) où le nombre de tâches cognitives à réaliser par le locuteur était trop lourd et a vraisemblablement joué dans le faible taux de réussite enregistré pour cette activité.

3. Mot de la fin et ouverture

Cette recherche présente des enjeux théoriques, épistémologiques et formatifs. Elle a permis d'apporter une réponse concrète et opératoire à un besoin méthodologique et descriptif concernant l'acquisition du lexique préfabriqué quotidien chez un public avancé non engagé dans un processus d'apprentissage guidé. Elle participe, nous semble-t-il, aux efforts déjà entrepris par certaines recherches sur l'anglais (Bardovi-Harlig notamment) en vue de : 1) développer un modèle de l'acquisition du lexique de l'interaction orale dans une perspective en lien avec la phraséologie pragmatique des interactions, 2) fondé sur la description outillée de la production des LNN. De manière indirecte, elle aborde un autre champ émergeant de linguistique descriptive en pleine effervescence, celui de la description des phraséologismes pragmatiques fréquemment employés dans la conversation orale. La classe des PhPex est une catégorie encore peu décrite et tout un travail d'inventaire et d'approfondissement de la caractérisation de ses spécificités reste à faire. Les descriptions du fonctionnement des PhPex

proposés dans cette thèse afin d'étayer notre analyse de la production des LNN observés contribuent également à l'effort de recherche mobilisé dans ce nouveau domaine.

Les applications didactiques semblent multiples. Notre travail répond à un manque de ressources utilisables pour la construction de dispositifs de formation axés sur le développement de la compétence pragmatique et plus particulièrement de la compétence en lien avec l'interaction orale, qu'elle soit quotidienne ou plus spécifique à certains types de conversations. Cette compétence conversationnelle a actuellement le vent en poupe et semble susciter de plus en plus d'initiatives, notamment du conseil de l'Europe et du CECRL ou des organismes de formation en langue française. C'est d'ailleurs partant de ce constat que nous souhaitons au cours de prochaines années proposer le projet de développement d'un dispositif pédagogique multilingue (français, japonais, anglais) d'aide à l'acquisition des expressions familières de la conversation orale quotidienne, dispositif qui sera accessible en ligne pour permettre un auto-apprentissage à domicile. Ce dispositif s'appuiera sur une première étape consistant en une description linguistique approfondie et une modélisation d'une cinquantaine d'expressions idiomatiques françaises et japonaises dont la fonction principale est d'exprimer les sentiments d'un locuteur en réaction à une situation. Dans un deuxième temps, ces descriptions seront utilisées pour créer une fiche de définition exhaustive pour chaque expression et activité d'apprentissage. L'objectif sera de développer à travers ces fiches de définition et ces activités une posture métapragmatique chez les apprenants afin de les aider à mieux comprendre l'utilisation de ce type d'expressions et de les rendre capables d'effectuer une meilleure analyse des stratégies communicatives propres à la conversation courante.

Il semble se profiler dans les années à venir tout un champ des possibles en lien avec les nouveaux enjeux de diffusion des savoirs et de contact des langues et des cultures, ouvrant ainsi de nombreuses collaborations transdisciplinaires entre la linguistique descriptive, l'acquisition L2 et la didactique. Ces interactions naissantes présagent des perspectives indéniablement stimulantes.

Bibliographie

Abou Haidar, L. « Analyse prosodique du discours pédagogique des enseignants de FLE ». Édité par F. Neveu, S. Prévost, A. Steuckardt, G. Bergounioux, et B. Hamma. *SHS Web of Conferences* 138 (2022): 06001. https://doi.org/10.1051/shsconf/202213806001.

Aijmer, K. « Swedish modal particles in a contrastive perspective ». *Language Sciences* 18, nº 1-2 (1996) : 393-427.

Aijmer, K. « Well I'm not sure I think… The use of well by non-native speakers ». *International Journal of Corpus Linguistics* 16, nº 2 (2011) : 231-54.

Alcón Soler, E., et Sánchez Hernández, A. « Learning pragmatic routines during study abroad: A focus on proficiency and type of routine ». *Atlantis*, (2017) : 191-210.

Alcón Soler, E. *Mitigating e-mail requests in teenagers' first and second language academic cyber-consultation*. De Gruyter Mouton, (2013).

Andersen, H. L. « Marqueurs discursifs propositionnels ». *Langue francaise* n° 154, nº 2 (2007) : 13-28.

André, V., et Canut, E. « Mise à disposition de corpus oraux interactifs : le projet TCOF (Traitement de Corpus Oraux en Français) ». *Pratiques*, nº 147-148 (2010) : 35-51.

André, V. « Analyse pragmalinguistique de la co-construction du discours: le cas des énonciations conjointes et des reprises ». In *Actes du colloque XII International Symposium on Social Communication*, (2011) : 17-21.

Andreadis, I., et Kartsounidou, E. « The impact of splitting a long online questionnaire on data quality ». In *Survey Research Methods*, 14, (2020) : 31-42.

Anscombre, J.-C., et Ducrot, O. « Lois logiques et lois argumentatives ». *Français (Le) Moderne Paris* 47, nº 1 (1979) : 35-52.

Anscombre, J.-C. « Figement, idiomaticité et matrices lexicales ». *Le figement linguistique : la parole entravée*, Honoré Champion, Paris. (2011) : 17-40.

Anzai, Y. « Deux fonctionnements du marqueur français "tu vois" dans les dialogues spontanés : Relation entre les faits intonatifs et la structure morphosyntaxique ». *Actes d'IDP* 9 (2009) : 63-76.

Auchlin, A. « Réflexions sur les marqueurs de structuration de la conversation ». *Études de linguistique appliquée* 44 (1981) : 88.

Austin, J. L. *Quand dire, c'est faire = How to do things with words*. Paris : Éditions du Seuil, (1970).

Bachman, L. F. *Fundamental considerations in language testing*. Oxford University Press, (1990).

Bachman, L. F., et Palmer, A. S. *Language testing in practice : Designing and developing useful language tests*. Vol. 1. Oxford University Press, (1996).

Baider, F., Cavalla, C., et Cislaru, G., éd. *Apprentissage des langues : compétence pragmatique, interculturalité*. Le langage et l'homme. Louvain-la-Neuve : EME Éditions, (2020).

Baldauf-Quilliatre, H., Colón de Carvajal, I., Etienne, C., Jouin-Chardon, E., Teston-Bonnard, S., et Traverso, V. « CLAPI, une base de données multimodale pour la parole en interaction : apports et dilemmes ». *Corpus*, nº 15 (2016).

Bally, C. *Traité de stylistique francaise*. 3. éd., nouv. Tirage. Genève : Georg [usw.], (1909).

Bange, P. « À propos de la communication et de l'apprentissage de L2 (notamment dans ses formes institutionnelles) ». *Acquisition et interaction en langue étrangère*, nº 1 (1992) : 53-85.

Bardière, Y. « De la pragmatique à la compétence pragmatique : A la recherche d'indices dans le Cecrl ». *Recherches en didactique des langues et des cultures* 13, nº 1 (2016).

Bardin, L. *L'analyse de contenu*. PUF. (2018).

Bárdosi, V. « Contribution à l'histoire de la phraséologie française des origines jusqu'à Michel Bréal ». *Acta Universitatis Szegedinensis, Acta Romanica*, t. XIV, Studia Lexicographica Neolatina II. (1990) : 61–132.

Bardovi-Harlig, K., et Hartford, B. S. « Learning the rules of academic talk : A longitudinal study of pragmatic change ». *Studies in second language acquisition* 15, nº 3 (1993) : 279-304.

Bardovi-Harlig, K., Bastos, M.-T., Burghardt, B., Chappetto, E., Nickels, E., Rose, M., Kasper, G., Nguyen, H. T., Yoshimi, D. R., et Yoshioka, J. K. « The use of conventional expressions and utterance length in L2 pragmatics ». *Pragmatics and language learning* 12 (2010) : 163-86.

Bardovi-Harlig, K., Mossman, S., et Su, Y. « The effect of corpus-based instruction on pragmatic routines ». *Language Learning & Technology* 21, nº 3 (2017) : 76-103.

Bardovi-Harlig, K. « 13 Explicit knowledge in L2 pragmatics ? » *New Directions in Second Language Pragmatics* 356 (2021) : 255.

Bardovi-Harlig, K. « Evaluating the empirical evidence : Grounds for instruction in pragmatics ». *Pragmatics in language teaching* 1332 (2001).

Bardovi-Harlig, K. « Exploring the pragmatics of interlanguage pragmatics: Definition by design ». *Pragmatics across languages and cultures* 7 (2010) : 219-59.

Bardovi-Harlig, K. « Formulaic language in second language pragmatics research ». *Understanding Formulaic Language*, (2018) : 97-114.

Bardovi-Harlig, K. « Formulaicity and context in second language pragmatics ». *Language in the context of communication : Studies in honor of Andrea Tyler*, (2018) : 193-211.

Bardovi-Harlig, K. « Formulas, routines, and conventional expressions in pragmatics research ». *Annual Review of Applied Linguistics* 32 (2012) : 206-27.

Bardovi-Harlig, K. « Pragmatics : Speaking as a Pragmalinguistic Resource ». *The Routledge Handbook of Second Language Acquisition and Speaking*, (2022) : 243-57.

Bardovi-Harlig, K. « Routines in L2 pragmatics research ». *The Routledge handbook of second language acquisition and pragmatics*, (2019) : 47-62.

Bardovi-Harlig, K., et Dörnyei, Z. « Do language learners recognize pragmatic violations? Pragmatic versus grammatical awareness in instructed L2 learning ». *Tesol Quarterly* 32, nº 2 (1998) : 233-59.

Bardovi-Harlig, K. « Conventional expressions as a pragmalinguistic resource: Recognition and production of conventional expressions in L2 pragmatics ». *Language Learning* 59, nº 4 (2009) : 755-95.

Bardovi-Harlig, K. « Developing L2 pragmatics ». *Language Learning* 63 (2013) : 68-86.

Bardovi-Harlig, K. « Exploring the interlanguage of interlanguage pragmatics: A research agenda for acquisitional pragmatics ». *Language learning* 49, nº 4 (1999) : 677-713.

Bardovi–harlig, K., et Su, Y. « The acquisition of conventional expressions as a pragmalinguistic resource in Chinese as a foreign language ». *The Modern Language Journal* 102, nº 4 (2018) : 732-57.

Barron, A. « Using corpus-linguistic methods to track longitudinal development: Routine apologies in the study abroad context ». *Journal of Pragmatics* 146 (2019) : 87-105.

Barron, A. *Acquisition in interlanguage pragmatics: Learning how to do things with words in a study abroad context*. Vol. 108. John Benjamins Publishing, (2003).

Baude, O. « Aspects juridiques et éthiques de la conservation et de la diffusion des corpus oraux ». *Revue française de linguistique appliquée* XII, nº 1 (2007) : 85-97.

Biber, D., et Finegan, E. *Sociolinguistic perspectives on register*. Oxford University Press, (1994).

Biber, D., et Conrad, S. *Register, genre, and style*. Cambridge University Press, (2019).

Biber, D. *Variation across speech and writing*. Cambridge University Press, (1991).

Bidaud, F. *Structures figées de la conversation : analyse contrastive français-italien*. Études contrastives, v. 4. Bern : Lang, (2002).

Bigi, B., Bertrand, R. et Guardiola, M. « Recherche automatique d'hétéro-répétitions dans un dialogue oral spontané ». In *XVIIIèmes Journées d'étude sur la Parole*, 1-4, (2010).

Blanche-Benveniste, C., Bilger, M., Rouget, C., Van Den Eynde, K., Mertens, P., et Willems, D. Le français parlé. Études grammaticales, Paris, Presses du CNRS, (1990).

Bley-Vroman, R. « The comparative fallacy in interlanguage studies: The case of systematicity 1 ». *Language learning* 33, nº 1 (1983) : 1-17.

Blum-Kulka, S., et Olshtain, E. « Requests and apologies : A cross-cultural study of speech act realization patterns (CCSARP) ». *Applied linguistics* 5, nº 3 (1984) : 196-213.

Blum-Kulka, S., et Olshtain, E. « Too many words: Length of utterance and pragmatic failure ». *Studies in second language acquisition* 8, nº 2 (1986) : 165-79.

Blum-Kulka, S., House, J., et Kasper, G. « Investigating cross-cultural pragmatics: An introductory overview ». *Cross-cultural pragmatics : Requests and apologies* 31 (1989) : 1-34.

Blum-Kulka, S. « Learning to Say What You Mean in a Second Language; a Study of the Speech Act Performance of Learners of Hebrew as a Second Language. », (1982).

Blumenthal, P., Mejri, S., et Gross, G., éd. *Phraséologie et profils combinatoires : lexique, syntaxe et sémantique : hommage à Peter Blumenthal*. Bibliothèque de grammaire et de linguistique 49. Paris : Honoré Champion, (2016).

Bolly, C. « Flou phraséologique, quasi-grammaticalisation et pseudo marqueurs de discours : un no man's land entre syntaxe et discours ? » *Linx. Revue des linguistes de l'université Paris X Nanterre*, nº 62-63 (2010a) : 11-38.

Bolly, C. « La notion de compétence "phraséologique" : Verbes à haute fréquence et corpus d'apprenants anglophones avancés », In : Isabel González Rey, *Les expressions figées en didactique des langues étrangères*, E.M.E. & InterCommunications : Cortil-Wodon, (2007) : 33-51.

Bolly, C. « Pragmaticalisation du marqueur discursif tu vois. De la perception à l'évidence et de l'évidence au discours ». In *2ème Congrès Mondial de Linguistique Française*, 045. EDP Sciences, (2010b).

Bréal, M. « Une science nouvelle : la sémantique ». *Revue des Deux Mondes (1829-1971)* 141, nº 4 (1897) : 807-36.

Bregeon, D., Antoine, J.-Y., Villaneau, J., et Lefeuvre-Halftermeyer, A. « Redonner du sens à l'accord interannotateurs : vers une interprétation des mesures d'accord en termes de reproductibilité de l'annotation ». *Traitement Automatique des Langues* 60, nº 2 (2019) : 23.

Brown, L. « Identity and honorifics usein Korean study abroad ». In C. Kinginger (Ed.). *Social and cultural aspects of language learningin study abroad*, Amsterdam : John Benjamins, (2013) : 269–298.

Brown, P., et Levinson, S. C. *Politeness: Some universals in language usage*. Vol. 4. Cambridge University Press, (1987).

Bunt, H. « Multifunctionality in dialogue ». *Computer Speech & Language* 25, nº 2, (2011) : 222-45.

Burger, H. *Phraseologie. Eine Einführung am Beispiel des Deutschen*. Berlin : Erich Schmidt Verlag, (1998).

Burger, H. *Phraseologie : ein internationales Handbuch zeitgenössischer Forschung*. Vol. 28. Walter de Gruyter, (2007).

Buysse, L. « The pragmatic marker you know in learner Englishes ». *Journal of Pragmatics* 121, (2017) : 40-57.

Canale, M. « From communicative competence to language pedagogy ». In J. Richards & R. Schmidt (Eds.), *Language and communication*. London : Longman, 1983 : 2-28.

Canale, M., et Swain, M. « Theoretical bases of communicative approaches to second language teaching and testing ». *Applied linguistics* 1, nº 1 (1980) : 1-47.

Candea, M. « Contribution à l'étude des pauses silencieuses et des phénomènes dits "d'hésitation" en français oral spontané : étude sur un corpus de textes en classe de français ». Thèse de doctorat, Paris 3, (2000).

Carrell, P. « Indirect speech acts in ESL : Indirect answers ». In C. Yorio, K. Perkins, & J. Schachter (Eds.), *On TESOL '79 : The learner I focus*. Washington, DC : TESOL. (1979) : 297-307.

Casarès, J. : *Introducción a la lexicografía moderna*. Madrid, (1969 [1950]).

Catena, À., et Corral, A. « La phraséologie en classe de FLE : structure interne du sens lexical et emploi des locutions ». *çédille, revistas de estudios franceses*, nº 13, (2017).

Causa, M. « L'alternance codique dans le discours de l'enseignant. Entre transmission de connaissances et interaction ». *Les Carnets du Cediscor. Publication du Centre de recherches sur la didacticité des discours ordinaires*, nº 4, (1996) : 111-29.

Cavalla, C. « La phraséologie en classe de FLE ». *Les Langues Modernes*, nº 1 (2009).

Cavalla, C. *Lexique transdisciplinaire et enseignement aux étudiants allophones*. ISTE Éditions, (2018).

Celce-Murcia, M., Dörnyei, Z., et Thurrell, S. « Communicative competence : A pedagogically motivated model with content specifications ». *Issues in Applied linguistics* 6, nº 2 (1995) : 5-35.

Celce-Murcia, M. « Rethinking the role of communicative competence in language teaching ». In *Intercultural language use and language learning*. Springer, (2008) : 41-57.

Chanet, C. « Fréquence des marqueurs discursifs en français parlé : quelques problèmes de méthodologie ». *Recherches sur le français parlé* 18, nº 83, (2004) : 106.

Clyne, M. « Some thoughts on pragmatics, sociolinguistic variation, and intercultural communication », *Intercultural Pragmatics* 3-1, (2006) : 95-105.

Conseil de l'Europe. *Cadre européen commun de référence pour les langues – Apprendre, enseigner, évaluer*. Paris : Didier, (2001).

Conseil de l'Europe. *Un cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer. Volume complémentaire avec de nouveaux descripteurs*, (2018).

Cook, H. M. *Socializing identities through speech style : Learners of Japanese as a foreign language*. Multilingual Matters, (2008).

Corblin, F. « Existe-t-il un 'ça' impersonnel en français ? » *L'information grammaticale* 62, nº 1 (1994) : 45-47.

Corder, S.P. « Que signifient les erreurs des apprenants ? » *Langages* 14, nº 57 (1980) : 9-15.

Corpas Pastor, G. *Manual de fraseología española*. Madrid, Gredos, (1996).

Coulmas, F. *Conversational routine : Explorations in standardized communication situations and prepatterned speech*. Vol. 96. Walter de Gruyter, (1981).

Cowie, A. P. « The Treatment of Collocations and Idioms in Learners Dictionaries ». *Applied Linguistics 2* (3), (1981) : 223-235.

Cowie, A. P. « Speech Formulae in English : Problems of Analysis and Dictionary Treatment ». *GAGL : Groninger Arbeiten Zur Germanistischen Linguistik*, nº 44 (2001) : 1-12.

Cowie, A. P., et Howarth, P. « Phraseological competence and written proficiency ». *British studies in applied linguistics* 11 (1996) : 80-93.

Crible, L., et Blackwell, S. E. *Introduction : Discourse-pragmatic markers in speech and sign*. Elsevier, (2020).

Crible, L. « Discourse markers and (dis) fluency in English and French : Variation and combination in the DisFrEn corpus ». *International Journal of Corpus Linguistics* 22, nº 2 (2017) : 242-69.

Crystal, D. *The Cambridge Encyclopedia of the English Language*. 2. ed., 7. print. Cambridge : Cambridge Univ. Press, (1997).

Culpeper, J., Mackey, A. et Taguchi, N. *Second language pragmatics: from theory to research*. New York, NY : Routledge, (2018).

Cutrone, P. « A case study examining backchannels in conversations between Japanese–British dyads ». *Multilingua* 24, nº 3 (2005) : 237-74.

De Cock, S. « Foreign words in EFL learner interviewee speech : Lending learners a productive fluency helping hand ? » *Fluency and disfluency across languages and language varieties*. (2019) : 283-99.

De Cock, S. « Repetitive phrasal chunkiness and advanced EFL speech and writing ». In *20th Annual Meeting of the International Computer Archive of Modern and Medieval English (ICAME)-Corpus Linguistics and Linguistic Theory*. (1999).

Degoumois, V. « L'expression d'opinions personnelles dans les interactions en classe:(dis) continuités dans la gestion des positionnements épistémiques entre l'école obligatoire et postobligatoire ». Thèse de doctorat, Université de Neuchâtel, (2017).

Delahaie, J. « Oui, voilà ou d'accord ? Enseigner les marqueurs d'accord en classe de FLE ». *Synergies Pays Scandinaves*, nº 4, (2009) : 17-34.

Dewaele, J.-M., et Wourm, N. « L'acquisition de la compétence sociopragmatique en langue étrangère ». *Revue francaise de linguistique appliquee* Vol. VII, nº 2, (2002) : 139-53.

Dewaele, J.-M., et Regan, V. « The use of colloquial words in advanced French interlanguage ». *EUROSLA yearbook* 1, nº 1, (2001) : 51-67.

Dostie, G., et Sikora, D. Les phraséologismes pragmatiques Entre langue et discours Présentation. *Les phraséologismes pragmatiques.*, Lexique 29, Presse universitaire de Lille (2021).

Dostie, G., et Tutin, A., éd. *Les phrases préfabriquées : sens, fonctions, usages*. Cahiers de lexicologie 2019, 1 = No 114. Paris : Classiques Garnier, (2019).

Dostie, G., et Pusch, C. D. « Présentation. Les marqueurs discursifs. Sens et variation ». *Langue française*, nº 2, (2007) : 3-12.

Dostie, G. « La réduplication pragmatique des marqueurs discursifs. De là à là là ». *Langue francaise* n° 154, nº 2, (2007) : 45-60.

Dostie, G. « Paramètres pour définir et classer les phrases préfabriquées : La vengeance est un plat qui se mange froid. Bon appétit ! » *Cahiers de lexicologie* 2019, nº 114, (2019) : 27-61.

Dostie, G. *Synonymie et marqueurs de haut degré. Sens conceptuel, sens associatif, polysémie*. Classiques Garnier, (2018).

Dostie, G., « Vers une modélisation lexicographique des propriétés sémantico-pragmatiques des locutions-phrases génériques et situationnelles. La nuit porte conseil et Le chat est sorti du sac », *Lexique*, nº 29, (2021) p. 15-38.

Ducrot, O. « Analyses pragmatiques ». *Communications* 32, nº 1, (1980) : 11-60.

Ducrot, O., Schaeffer, J.-M. *Nouveau dictionnaire encyclopédique des sciences du langage.* Paris : Éditions du Seuil, (1995).

Dufour, F. « Reformulation métalinguistique et re-catégorisation du référent : du progrès civilisateur au développement ». *L'acte de nommer : une dynamique entre langue et discours*, (2007).

Dumont, A. « Fluency and disfluency : A corpus study of non-native and native speaker (dis) fluency profiles ». Thèse de doctorat. UCL-Université Catholique de Louvain, (2018).

Edmonds, A. « Conventional expressions : Investigating pragmatics and processing ». *Studies in Second Language Acquisition* 36, nº 1, (2014) : 69-99.

Edmonds, A. « Une approche psycholinguistique des phénomènes phraséologiques : le cas des expressions conventionnelles ». *Langages* 189, nº 1 (2013) : 121.

Eisenstein, M, et Bodman, J. W. « 'I very appreciate': expressions of gratitude by native and non-native speakers of American English ». *Applied linguistics* 7, nº 2, (1986) : 167-85.

Guérin, E. « Une description fondée sur l'oral (?) : pense *ça* sans *cela* », dans Paul Cappeau (dir.), *Une grammaire à l'aune de l'oral ?*, Rennes : Presses Universitaires de Rennes, (2020) : 145-156.

Félix-Brasdefer, C. « Refusing in L2 Spanish : The effects of the context of learning during a short-term study abroad program ». In *Refusals in instructional contexts and beyond*. Brill, (2013) : 147-73.

Ferre, G. « Multimodal Analysis of Discourse Markers 'donc', 'alors' and 'en fait' in Conversational French ». In *ICPhS VXII*, 671-74, (2011).

Fillmore, C. J., Kay, P., et O'connor, M. C. « Regularity and idiomaticity in grammatical constructions : The case of let alone ». *Language* 64, nº 3 (1988) : 501-38.

Firth, J. R. « Descriptive linguistics and the study of English » in Palmer F.R. (éd.), (1968 [1956]) : 96-113.

Fonagy, I. *Situation et signification*. Pragmatics & beyond, III :1. Amsterdam ; Philadelphia: J. Benjamins, (1982).

Forsberg, F. « Using conventional sequences in L2 French », *IRAL* 48, Walter de Gruyter, (2010) : 25-51.

Fraser, B. « Idioms within a transformational grammar ». *Foundations of language*, (1970) : 22-42.

Fukuya, Y., Reeve, M., Gisi, J. et Christianson, M. « Does Focus on Form Work for Teaching Sociopragmatics ? », *12th Annual Meeting of the International Conference on Pragmatics and Language Learning*, University of Illinois, Urbana-Champaign, (1998).

Fung, L., et Carter, R. « Discourse markers and spoken English : Native and learner use in pedagogic settings ». *Applied linguistics* 28, nº 3, (2007) : 410-39.

Furniss, E. A. « Teaching the pragmatics of Russian conversation using a corpus-referred website », *Language Learning & Technology*, *20*(2), (2016) : 38-60.

Gadet, F. « Niveaux de langue et variation intrinsèque ». *Palimpsestes. Revue de traduction*, nº 10, (1996) : 17-40.

Gadet, F. *Le français ordinaire*. Paris : Armand Colin, (1989).

Geeraerts, D. « Cultural models of linguistic standardization ». *Cognitive models in language and thought. Ideology, metaphors and meanings* 24, (2003) : 25-68.

Gharbi, N. « Analyse sémantico-pragmatique et discursive : les formules expressives de la conversation ». Thèse de doctorat. Université Grenoble Alpes ; Université de Sfax, (2020).

Gilquin, G. « Discourse markers in L2 English ». *New Approaches to English Linguistics*, (2016) : 213.

Gilquin, G. « Hesitation markers among EFL learners: Pragmatic deficiency or difference ». *Pragmatics and corpus linguistics: A mutualistic entente*, (2008) : 119-49.

Gläser, R. « Phraseologie Der Englischen Sprache». *Tübingen: Niemeyer. DOI* 10, (1986).

Goldberg, A. E. *Constructions : A Construction Grammar Approach to Argument Structure*. Chicago : Chicago University Press, (1995).

González Rey, I. *La nouvelle phraséologie du français*. 3e éd. Revue et Augmentée. Interlangues. Toulouse : Presses universitaires du Midi, (2021).

González Rey, I. *La phraséologie du français*. Interlangues Linguistique et didactique. Toulouse : Presses Univ. Du Mirail, (2002).

González-Rey, I. « La phraséodidactique : état des lieux ». *Repères DoRiF*, phraséodidactique : de la conscience à la compétence, nº 18, (2019).

González Rey, I. *La didactique du français idiomatique*, EME Éditions, (2015).

González-Rey, I. « La phraséodidactique : état des lieux ». *Repères DoRiF*, phraséodidactique : de la conscience à la compétence, nº 18, (2019).

Granger, S. 1996. « From CA to CIA and back : An integrated contrastive approach to computerized bilingual and learner corpora ». In K. Aijmer, B. Altenberg & M. Johansson (Eds.), *Languages in Contrast. Text-based cross-linguistic studies*. Lund : Lund University Press, (1996) : 37–51.

Gréciano, G. *Signification et dénotation en allemand : la sémantique des expressions idiomatiques*. Vol. 1. Klincksieck, (1983).

Grezka, A, et Kijima, A. « L'expression de la perception visuelle : regard franco-japonais ». *Lexis. Journal in English Lexicology*, nº 13, (2019).

Gross, G. *Les expressions figées en français : noms composés et autres locutions*. Éditions Ophrys, (1996).

Gross, M. « Les phrases figées en français ». *L'Information Grammaticale* 59, nº 1, (1993) : 36-41.

Gülich, E, et Kotschi, T. « Les marqueurs de la reformulation paraphrastique ». *Connecteurs pragmatiques et structure du discours ; actes du 2ème Colloque de Pragmatique de Genève* 5, (1983).

Gülich, E, et Krafft, U. « Le rôle du « préfabriqué » dans les processus de production discursive ». In *La locution entre langue et usages*, (1997).

Gumperz, J. J. *Discourse strategies*. Cambridge University Press, (1982).

Abou Haidar, L. « Analyse Prosodique Du Discours Pédagogique Des Enseignants de FLE ». *Congrès Mondial de Linguistique Française*, SHS Web of Conferences 138, (2022) : 06001.

Harada, S. « La force intermédiale du film en classe de FLE au Japon ». *Langues modernes*, nº 1, (2019) : 6.

Harada, S. « Pour une discussion métapragmatique en classe de FLE ». *Revue japonaise de didactique du français* 2, nº 1, (2007) : 53-70.

Harada, S. « Les stratégies de refus : excessives ou insuffisantes ? » dans Baider, F., Cavalla, C. et Cislaru, G. éd. *Apprentissage des langues : compétence pragmatique, interculturalité*. Le langage et l'homme. Louvain-la-Neuve : EME Éditions, (2020).

Hashimoto, C., Sato, S., et Utsuro, T. « Detecting Japanese idioms with a linguistically rich dictionary ». *Language Resources and Evaluation* 40, nº 3, (2006) : 243-52.

Hassall, T. « Learning to take leave in social conversations : A diary study ». *Language learners in study abroad contexts* 3158, (2006).

Häusermann, J. Phraseologie. Hauptprobleme der deutschen Phraseologie auf der Basis sowjetischer Forschungsergebnisse. Tübingen, (1977).

Hausmann F. J. Le dictionnaire de collocations. In Hausmann F.J., Reichmann O., Wiegand H.E., Zgusta L. (eds), *Wörterbücher : ein internationales Handbuch zur Lexicographie. Dictionaries. Dictionnaires*. Berlin/New-York, De Gruyter, (1989) : 1010-1019.

Hayashi, R. « Simultaneous talk—from the perspective of floor management of English and Japanese speakers ». *World Englishes* 7, nº 3, (1988) : 269-88.

Heritage, J. « Epistemics in action : Action formation and territories of knowledge ». *Research on Language & Social Interaction* 45, nº 1, (2012) : 1-29.

Higashi, T. « Convergence émotionnelle dans la pratique communicative des Japonais–Analyse de la communication exolingue entre Japonais et Français ». *LIDIL-Revue de linguistique et de didactique des langues*, nº 5, (1992) : 13-30.

Hinai K. et Ladreyt A. "Typologie préliminaire des phraséologismes spécifique au parler des gamers" (ゲーマー語にみられる慣用表現の特徴に関する予備的類型論). *Bulletins de linguistique et de littérature françaises de l'Université de Tsukuba*, vol.35, Presses Universitaires de Tsukuba, (2020).
Hinaï K. et Ladreyt A. « Les phraséologismes pragmatiques de la conversation quotidienne : typologie et étude contrastive français / japonais des expressions fondées sur le motif de la mort », Colloque International CEDIL 2022, Grenoble, (2022).

Hoffmann, T., et Trousdale, G. *The Oxford handbook of construction grammar*. Oxford University Press, (2013).

House, J. « Developing pragmatic fluency in English as a foreign language : Routines and metapragmatic awareness ». *Studies in second language acquisition* 18, nº 2, (1996) : 225-52.

Howarth, P. *Phraseology in English Academic Writing*. Tübingen : Max Niemeyer, (1996).

Huang, L.-F. « A corpus-based exploration of the discourse marker well in spoken interlanguage ». *Language and speech* 62, nº 3, (2019) : 570-93.

Hudson, T., Brown, J. D., et Detmer, E. *Developing prototypic measures of cross-cultural pragmatics*. Vol. 7. Natl Foreign Lg Resource Ctr, (1995).

Hymes, D. « On communicative competence ». *sociolinguistics* 269293, (1972) : 269-93.

Hyvärinen, I. « Zur Abgrenzung und Typologie pragmatischer Phraseologismen–Forschungsüberblick und offene Fragen ». In *Beiträge zur pragmatischen Phraseologie: Finnische Beiträge zur Germanistik*. Peter Lang, (2011).

Ifantidou, E. « Genres and pragmatic competence ». *Journal of Pragmatics* 43, nº 1, (2011) : 327-46.

Ifantidou, E. *Pragmatic Competence and Relevance*. Pragmatics & Beyond, N.S., 245. Amsterdam : Benjamins, (2014).

Ike, S. « Backchannel : a feature of Japanese English ». In *JALT 2009 Conference Proceedings*, (2010) : 205-215.

Ishida, P. « On the Use of Syntactic Operations to Measure the Semantic Frozenness of Verb Phrase Idioms ». *Studies in the Japanese language* 55, nº 4 (2004) : 42-56.

Iwasaki, N. « Learning L2 Japanese" politeness" and" impoliteness" : Young American men's dilemmas during study abroad ». *Japanese language and literature*, (2011) : 67-106.

Iwatsuki, K. « Towards Extracting Formulaic Expressions from Japanese Scholarly Papers », In Hämäläinen, M., Partanen, N., Alnajjar, K. (eds.) Multilingual Facilitation, (2021) : 74-82.

Jacques, M.-P., et Tutin, A. *Lexique transversal et formules discursives des sciences humaines*. ISTE Group, (2018).

Jakobson, R. « Essais de linguistique générale ». *Les Etudes Philosophiques* 18, nº 4, (1963).

Kamio, A. *Territory of information*. J. Benjamins Pub., (1997).

Kasper, G. « Data collection in pragmatics research ». In H. Spencer-Oatey (Ed.), *Culturally speaking : Culture, communication and politeness theory* (2nd ed.). London : Continuum. (2008) : 279-303.

Kasper, G., et Roever, C. « Pragmatics in second language learning ». *Handbook of research in second language teaching and learning*, (2005) : 341-58.

Kasper, G., et Rose, K. R. *Pragmatic Development in a Second Language*. Language Learning, 52.2002, Suppl. 1. Malden, Mass : Blackwell, (2002).

Kasper, G., et Dahl, M. « Research methods in interlanguage pragmatics ». *Studies in second language acquisition* 13, nº 2, (1991) : 215-47.

Kasper, G., et Schmidt, R. « Developmental issues in interlanguage pragmatics ». *Studies in second language acquisition* 18, nº 2, (1996) : 149-69.

Kasper, G., et Blum-Kulka, S. « Interlanguage pragmatics : An introduction ». *Interlanguage pragmatics* 3, (1993) : 15.

Kasper, G., Maeshiba, N., Yoshinaga, N., et Ross, S. « Transfer and proficiency in interlanguage apologizing ». *Speech acts across cultures : Challenges to communication in a second language*, (1996) : 155-87.

Kasper, G. « Four perspectives on L2 pragmatic development ». *Applied linguistics* 22, nº 4, (2001) : 502-30.

Kauffer, M. « La pragmatique est-elle soluble dans la phraséologie ? » In *La Phraséologie Française : Débats théoriques et dimensions appliquées (didactique, traduction et traitement informatique)*, Hermann, (2017) : 15.

Kauffer, M. « Le figement des « actes de langage stéréotypés » en français et en allemand ». *Pratiques. Linguistique, littérature, didactique*, nº 159-160, (2013) : 42-54.

Kauffer, M. « Les “actes de langage stéréotypés” : essai de synthèse critique ». *Cahiers de lexicologie* 2019, nº 114, (2019) : 149-71.

Kauffer, M. « Les actes de langage stéréotypés ». Présentation lors du séminaire DéLiCorTAL du LIDILEM, Grenoble, (2016).

Kauffer, M. « Plaidoyer pour une phraséologie pragmatique sur la base des ALS français et allemands ». In *II International Congress of Phraseology and Paremiology*, 2 :192-208. Pontes Editores, (2011).

Kauffer, M. « Qu'est-ce qu'un ALS ». *Verbum XL* 1, (2018) : 35-50.

Kauffer, M. *Phraséologismes et actes de langage*. Champion, (2018).

Kauffer, M. *Réflexions sur le figement des 'phraséologismes pragmatiques' en français et en allemand*. Peter Lang, (2014).

Kecskes, I., Obdalova, O., Minakova, L., et Soboleva, A. « A study of the perception of situation-bound utterances as culture-specific pragmatic units by Russian learners of English ». *System* 76, (2018) : 219-32.

Kecskes, I. « Conceptual fluency and the use of situation-bound utterances in L2 ». *Links & Letters*, (2000) : 145-61.

Kecskes, I. *Intercultural pragmatics*. Oxford University Press, (2014).

Kerbrat-Orecchioni, C. *Les interactions verbales*. Linguistique. Paris : Armand Colin, (1990).

Kerbrat-Orecchioni, C. : *L'énonciation, De la subjectivité dans le langage*, Paris, Armand Colin, (1980).

Kerbrat-Orecchioni, C., et Mitterand, H. *Les actes de langage dans le discours : théories et fonctionnement*. Cursus. Paris : Armand Colin, (2009).

Kerbrat-Orecchioni, C. « L'analyse des interactions verbales : La notion de « négociation conversationnelle » : Défense et illustrations ». *L'analyse des interactions verbales : La notion de « négociation conversationnelle » : Défense et illustrations*, nº 20, (2000) : 63-141.

Kerbrat-Orecchioni, C. *Le discours en interaction*. Collection U Lettres - Linguistique. Paris : Armand Colin, (2005).

Kim, H. Y. « Learner investment, identity, and resistance to second language pragmatic norms ». *System* 45 (2014) : 92-102.

Kinginger, C., et Blattner, G. « Histories of engagement and sociolinguistic awareness in study abroad : Colloquial French ». In L. Ortega & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities*. London : Routledge, (2008) : 223-246.

Kinginger, C., et Farrell, F. « Assessing development of meta-pragmatic awareness in study abroad ». *Frontiers : The interdisciplinary journal of study abroad* 10, nº 1, (2004) : 19-42.

Kita, S. *Aizuchi to unazuki kara mita nihonjin no taijin communication [Japanese face to face communication : data from aizuchi and headnods]*. Nihongogaku, (1996).

Kita, S., et Ide, S. « Nodding, aizuchi, and final particles in Japanese conversation : How conversation reflects the ideology of communication and social relationships ». *Journal of Pragmatics* 39, nº 7, (2007) : 1242-54.

Klein, J.-R., et Lamiroy, B. « Routines conversationnelles et figement ». In *Le figement linguistique : la parole entravée*. Honoré Champion, (2011) : 195-214.

Kobayashi, H., et Hirose, K. « Japanese learners' repetition in conversation in relation to English proficiency level ». *JALT n°1-17*, (1995) : 53-73.

Krzyżanowska, A., Grossmann, F., et Kwapisz-Osadnik, K., éd. *Les formules expressives de la conversation : analyse contrastive : français - polonais - italien*. Wydawnictwo Episteme. Lublin, (2021).

Krzyżanowska, A. « Les formules expressives de la conversation. Analyse contrastive : français polonais ». Présentation au Séminaire DéLiCorTAL, LIDILEM, Université Grenoble Alpes, 13 novembre 2020.

Labov, W. « The linguistic consequences of being a lame ». *Language in society* 2, nº 1, (1973) : 81-115.

Ladreyt A. « À propos de l'emploi des formules de salutation en japonais —Spécificité et contraintes — (日本語の挨拶表現の使用について—特徴と使用制限—) », *Bulletins de linguistique et de littérature françaises de l'Université de Tsukuba*, vol.34, Presses Universitaires de Tsukuba, (2019).

Ladreyt, A. 日常会話の感情を表す慣用表現と語用論的機能 —驚きを表示する慣用表現のコーパス研究をめぐって— *Bulletins de linguistique et de littérature françaises de l'Université de Tsukuba*, vol.36, (2021).

Ladreyt, A. « Les Phraséologismes pragmatiques à fonction expressive de la conversation quotidienne : spécificités linguistiques et dynamiques d'usage ». *Congrès Mondial de Linguistique Française*, SHS Web of Conferences 138 (2022a) : 04009.

Ladreyt A. « Application contrastive du modèle des phraséologismes pragmatiques à fonction expressive au cas des expressions à substrat physiologique exprimant un sentiment : le cas du motif de la colère », Revue Littera, (2022b). ISSN : 2432-3160

Lamiroy, B., et Klein, J.-R. « Le problème central du figement est le semi-figement ». *Linx*, nº 53, (2005) : 135-54.

Lamiroy, B. « Le figement : à la recherche d'une définition ». *ZFSL, Zeitschrift für französische Sprache und Literatur* 36, (2008) : 85-99.

Larsen-Freeman, D. « Another step to be taken: Rethinking the endpoint of the interlanguage continuum ». *Interlanguage : Forty years later* 39, (2014) : 203-20.

Leech, G. *The Principles of Pragmatics. New York, America*. Longman Group Limited, (1983).

Lefeuvre, F. « ‘Eh bien’ comme évaluateur de discours l’oral spontané ou représenté », *Travaux de linguistique, Bien en perspective* (Moline ed.), n° 65, (2012) : 123-143.

Lefeuvre, F. « Les marqueurs discursifs averbaux résomptifs ». (2021). Halshs-03143412

Lefeuvre, F. et Legallois, D. : « Les marqueurs discursifs rédupliqués : le cas de oui oui oui oui », *Représentations du sens linguistique : modalité intra- et extra-phrastique*, Dostie et Larrivée eds, Caen : Presses universitaires de Caen, (2020) : 181-198.

Legallois, D., et Patard, A. « Les constructions comme unités de la langue : illustrations, évaluation, critique ». *Langue française* 194, nº 2, (2017) : 5-14.

Legallois, D., et Tutin, A. « Présentation : Vers une extension du domaine de la phraséologie ». *Langages* 189, nº 1, (2013) : 3.

Liu, N., et Yang, L. « The Effect Of Study Abroad And Intensity Of Interaction On The Development Of L2 Pragmatic Routines By Chinese Learners Of English », (2020).

LoCastro, V. « A sociocultural functional analysis of fragmentation in Japanese ». *Multilingua* 18, nº 4, (1999) : 369-90.

LoCastro, V. « Aizuchi : A Japanese conversational routine ». *Discourse across cultures*, (1987) : 101-13.

Lopes, A., Martins de Matos, D., Cabarrão, V., Ribeiro, R., Moniz, H., Trancoso, I., et Mata, A. I. « Towards using machine translation techniques to induce multilingual lexica of discourse markers ». *arXiv preprint arXiv:1503.09144*, (2015).

López Simó, M. *Fórmulas de la conversación. Propuesta de definición y clasificación con vistas a su traducción español-francés, francés-español*, Thèse de doctorat, Université d’Alicante, (2016).

Lorrillard, O. « Les « tics de langage » ont-ils leur place dans l’enseignement du français oral ? » 広島大学大学院文学研究科論集, 78 (2018) : 58.

Lüger, H. H. « Some aspects of ritual communication ». *Journal of Pragmatics*, *7*(6), (1983) : 695-711.

Lundell Forsberg, F., et Erman, B. « High-level requests : A study of long residency L2 users of English and French and native speakers ». *Journal of Pragmatics* 44, nº 6-7, (2012) : 756-75.

Machi, S. « Managing relationships through repetition : How repetition creates ever-shifting relationships in Japanese conversation ». *Pragmatics* 29, nº 1, (2019) : 57-82.

Maemura, Y. « Humor and laughter in Japanese groups : The kuuki of negotiations. » *Humor : International Journal of Humor Research* 27, nº 1, (2014).

Magri, V. « Marqueurs de reformulation : exploration outillée et contrastive dans deux corpus narratifs ». *Langages* N° 212, nº 4, (2018) : 35-50.

Maillard, M. « Concurrence et complémentarité de il et ça devant les prédicats impersonnels en français contemporain ou comment distinguer une phrase asubjectale d'une phrase à sujet indistinct ? » *L'information grammaticale* 62, nº 1, (1994) : 48-52.

Malinowski, B. *The problem of meaning in primitive languages, supplement to" The meaning of meaning" by CK Ogden and IA Richards*. New York, Harcourt, Brace and Company, (1923).

Marque-Pucheu C. « Les énoncés liés à une situation : mode de fonctionnement et mode d'accès en langue 2 », *Hieronymus I*, (2007) : 25-48.

Martin, C. G. « Étude de pragmatèmes : salutations, injonctions et jurons ». Thèse de Doctorat, Université Paris-Nord-Paris XIII, (2019).

Marquillo Larruy, M. *L'interprétation de l'erreur en langue étrangère*. Clé international, Nathan, Collection Didactique des Langues, (2003) : 128.

Martin, R. « Sur les facteurs du figement lexical ». *La locution entre langue et usage, Paris, ENS éditions Fontenay Saint-Cloud*, (1997) : 291-305.

Martins-Baltar, M., éd. *La locution entre langue et usages*. Langages. Fontenay/Saint-Cloud : ENS Éd, (1997).

Masseron, C., et Luste-Chaa, O. « Typologie d'erreurs lexicales : difficultés et enjeux ». In *Congrès Mondial de Linguistique Française*, 064. EDP Sciences, (2008).

Mathet, Y., Widlöcher, A., et Métivier, J.-P. « The Unified and Holistic Method Gamma (γ) for Inter-Annotator Agreement Measure and Alignment ». *Computational Linguistics* 41, nº 3, (2015) : 437-79.

Matsumoto, M., et Boyè Lafayette, D. *Japanese Nuance in Plain English*. Tokyo : Kodansha, (2000).

Matsumoto, Y. « Functions of laughter in English-as-a-lingua-franca classroom interactions : A multimodal ensemble of verbal and nonverbal interactional resources at miscommunication moments ». *Journal of English as a Lingua Franca* 7, nº 2, (2018) : 229-60.

Maynard, S. K. « Analyzing interactional management in native/non-native English conversation : A case of listener response ». *IRAL, International Review of Applied Linguistics in Language Teaching* 35, nº 1, (1997) : 37.

Maynard, S. K. « Conversation management in contrast : Listener response in Japanese and American English ». *Journal of Pragmatics* 14, nº 3, (1990) : 397-412.

Maynard, S. K. « Interactional functions of a nonverbal sign Head movement in japanese dyadic casual conversation ». *Journal of pragmatics* 11, nº 5, (1987) : 589-606.

Maynard, S. K. « On back-channel behavior in Japanese and English casual conversation ». *Linguistics* 24, nº 6, (1986) : 1079-1108.

Mazzocconi, C. « Laughter in interaction : semantics, pragmatics, and child development ». Thèse de doctorat, Université de Paris, (2019).

McIntosh, A. & Halliday, M.A.K. « Patterns of Language ». *Papers in General, Descriptive and Applied Linguistics*, London, Longmans, (1966).

Mejri, S. « L'idiomaticité, problématique théorique ». In : Salah Mejri, dir. *L'espace euro-méditerranéen : une idiomaticité partagée*. Tunis : CERES, (2003) : 231-243.

Mejri, S., et Gross, G. « Phraséologie et profils combinatoires : lexique, syntaxe et sémantique : hommage à Peter Blumenthal ». *Phraséologie et profils combinatoires*, (2016), 1-368.

Mejri, S., Meneses Lerin, L., et Buffard-Moret, B. *La phraséologie française en questions*. Vertige de la langue. Paris : Hermann, (2020).

Mejri, S. « Introduction : polysémie et polylexicalité ». *Syntaxe et sémantique* 5, nº 1, (2004) : 13-30.

Mejri, S. « Le figement lexical : descriptions linguistiques et structuration sémantique ». *L'information grammaticale* 76, nº 1, (1998) : 50-51.

Mejri, S. « Le figement lexical : nouvelles tendances ». *Cahiers de lexicologie* 80, (2002) : 213-25.

Mekki, J., Battistelli, D., Béchet, N., et Lecorvé, G. « TREMoLo: un corpus multi-étiquettes de tweets en français pour la caractérisation des registres de langue ». In *Traitement Automatique des Langues Naturelles*, 224-32. ATALA, (2021).

Mel'čuk, I. « Phrasemes in language and phraseology in linguistics ». *Idioms: Structural and psychological perspectives*, (1995) : 167-232.

Mel'čuk, I. « Tout ce que nous voulions savoir sur les phrasèmes, mais ». *Cahiers de lexicologie : Revue internationale de lexicologie et lexicographie*, nº 102, (2013) : 129-49.

Meléndez Quero, C. « Comment expliquer la signification des particules discursives d'une langue étrangère ? » *Recherche et pratiques pédagogiques en langues de spécialité. Cahiers de l'Apliut* 29, nº 1, (2010) : 137-51.

Miura, A. « Discourse markers in spoken corpora of Japanese EFL learners. » Présenté à *Learner Corpus Research*, Louvain-la-Neuve, Belgique, (2011).

Miura, I. « Comprehension of idiomatic expressions in Japanese—Auditory and visual presentations ». *Journal of Psycholinguistic Research* 25, nº 6, (1996) : 659-76.

Miyachi, K. *Laughter by Japanese Learners of English*. SLS, (2009).

Moeschler, J., et Reboul, A. *Dictionnaire encyclopédique de pragmatique*. Seuil, (1994).

Moon, R. *Fixed expressions and idioms in English : A corpus-based approach*. Oxford University Press, (1998).

Morris, C. W. « Foundations of the Theory of Signs ». In *International encyclopedia of unified science*, 1-59. Chicago University Press, (1938).

Müller, S. *Discourse markers in native and non-native English discourse*. Vol. 138. John Benjamins Publishing, (2005).

Náray-Szabó, M. « Formes du non-dit dans les énoncés liés ». *Revue d'étude française*, nº 14, (2009).

Náray-Szabó, Márton. « Phrases figées illocutoires : quelle interprétation sémantique ? » *Verbum Analecta Neolatina* 1, nº 11, (2009) : 153-65.

Nguyen, T.T.M. « Data collection methods in L2 pragmatics research : An overview ». *The Routledge handbook of second language acquisition and pragmatics*, (2019) : 195-211.

Nishizaka, A. « The interactive constitution of interculturality : How to be a Japanese with words ». *Human studies* 18, nº 2-3, (1995) : 301-26.

Noda, H. « Intersubjectivité : modulation et ajustement. Cas des marqueurs discursifs "hein", "quoi", "n'est-ce pas" en français et "darô", "yo", "ne", "yone" en japonais ». Thèse de doctorat, Université de Franche-Comté, (2011).

Novakova, I., et Siepmann, D. « Literary style, corpus stylistic, and lexico-grammatical narrative patterns : Toward the concept of literary motifs ». In *Phraseology and Style in Subgenres of the Novel*, 1-15. Springer, (2020).

Ostman, J.-O. « The Symbiotic Relationship between Pragmatic Particles and Impromptu Speech. », (1982).

Osuka, N. « Development of pragmatic routines by Japanese learners in a study abroad context ». *Current issues in intercultural pragmatics*, (2017) : 275-96.

Paillard, D., et Vu, T. N. *Inventaire raisonné des marqueurs discursifs du français. Description. Comparaison. Didactique.* Presse de l'Université Nationale de Hanoï, (2012).

Paul, H. Prinzipien der Sprachgeschichte. Halle : Max Niemeyer. (1880).

Pausé, M. S., & Tutin, A. « Some Insights on a Typology of French Interactional Prefabricated Formulas in Spoken Corpora ». In *International Conference on Computational and Corpus-Based Phraseology,* Springer, Cham. (2022) : 190-205.

Pecman, M. « L'enjeu de la classification en phraséologie ». In *Internationalen Tagung zur Phraseologie*, édité par A. Häcki Buhofer et H Burger. Basel, Switzerland, (2004) : 29-48.

Peirce, C. S. (1931). *The Collected Papers of Charles Sanders Peirce, Vol. I : The Principles of Philosophy*. CP 1. Edited by C. Hartshorne & P. Weiss. Cambridge : Harvard University Press, (1931).

Pellet, S. H. « The development of competence in French interlanguage pragmatics: The case of the discourse marker 'donc' », Thèse de doctorat, Université du Texas, (2005).

Perdue, C. « L'analyse des erreurs : un bilan pratique ». *Langages* 14, nº 57, (1980) : 87-94.

Perkins, M. *Pragmatic Impairment*. Cambridge University Press, Cambridge, (2007).

Polat, B. « Investigating acquisition of discourse markers through a developmental learner corpus ». *Journal of pragmatics* 43, nº 15, (2011) : 3745-56.

Polguère, A. « Classification sémantique des lexies fondée sur le paraphrasage ». *Cahiers de lexicologie* 2011, nº 98, (2011) : 197-211.

Porquier, R. « Deux repères de recherche sur l'interaction exolingue ». *Linx. Revue des linguistes de l'université Paris X Nanterre*, nº 49, (2003) : 51-62.

Porquier, R., Ōshima, H., Bazantay, J. éd. *Apprentissage d'une langue éloignée : analyse des erreurs d'apprenants francophones,* actes du colloque « Analyse des erreurs commises par des francophones apprenant une langue éloignée », Centre d'études japonaises, Paris, (2020).

Quillard, G. « Relations intersubjectives et marqueurs culturels dans les publicités anglaises et françaises ». *TTR : traduction, terminologie, rédaction* 11, nº 1, (2007) : 137-55.

Roever, C. « What learners get for free : Learning of routine formulae in ESL and EFL environments ». *ELT journal* 66, nº 1, (2012) : 10-21.

Roever, C. *Language testing and evaluation: Testing ESL pragmatics: Development and validation of a web-based assessment battery*. Lang-Peter Lang, (2005).

Rogers, M. F. *Sociology, ethnomethodology and experience*. CUP Archive, (1983).

Rosa, E. M., et Leow, R. P. « Awareness, different learning conditions, and second language development ». *Applied psycholinguistics* 25, nº 2, (2004) : 269-92.

Rose, K.R. and Ng Kwai-fun, C. « Inductive and deductive approaches to teaching compliments and compliment responses ». In Rose and Kasper, *Pragmatics in Language Teaching* 145, (2001) : 145–170.

Rouget, C. « Impersonnel, quasi-impersonnel et pseudo-impersonnel en français parlé ». *Revue romane*, (2003).

Ruíz Gurillo, L. « Relaciones categoriales de las locuciones adverbiales », (1997).

Sahin Kizil, A. « Discourse Markers in Learner Speech : A Corpus Based Comparative Study ». *Journal of Language Education and Research* 7, nº 1, (2021).

Sato, S. « A dependency syntax of Japanese ». *Metataxis in Practice : Dependency Syntax for Multilingual Machine Translation, D. Maxwell and K. Schubert, Eds. (Foris, Dordrecht/Providence)*, (1989), 183-206.

Saussure, F. *Cours de linguistique générale*. Grande bibliothèque Payot. Paris : Payot, (2005[1916]).

Schiffrin, D. *Discourse markers*. Cambridge University Press, (1987).

Schmale, G. « Qu'est-ce qui est préfabriqué dans la langue ? ? Réflexions au sujet d'une définition élargie de la préformation langagière ». *Langages* 189, nº 1, (2013) : 27.

Schmidt, R. « Attention ». In P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge University Press, (2001) : 3-32.

Schmitt, N., Dornyei, Z., Adolphs, S., et Durow, V. « Knowledge and acquisition of formulaic sequences ». *Formulaic sequences: Acquisition, processing and use*, (2004) : 55-86.

Schneider, F. *Comment décrire les actes de langage ? de la linguistique pragmatique à la lexicographie : « La belle affaire ! » et « Tu m'en diras tant ! »* Linguistische Arbeiten 227. Tübingen : Niemeyer, (1989).

Schourup, L. « Discourse markers ». *Lingua* 107, nº 3-4, (1999) : 227-65.

Searle, J. R. « A taxonomy of illocutionary acts », (1975).

Searle, J. R. *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press, (1985).

Searle, J. R., Vanderveken, D., et Willis, S. *Foundations of illocutionary logic*. CUP Archive, (1985).

Sechehaye, A. « Locutions et composés ». *Journal de psychologie normale et pathologique* 18, (1921) : 654-75.

Selinker, L. « Interlanguage ». *IRAL - International Review of Applied Linguistics in Language Teaching* 10, nº 1-4, (1972).

Selinker, L. « Interlanguage 40 years on Interlanguage : forty years later », (2014) : 229-63.

Shea, D. P. « Perspective and production : Structuring conversational participation across cultural borders ». *Pragmatics* 4, nº 3, (1994) : 357-89.

Shimada, K. « Contrastive interlanguage analysis of discourse markers used by nonnative and native English speakers ». *JALT journal* 36, nº 1, (2014) : 47-68.

Shimada, K. « Discourse marker use in learners and native speakers : A corpus-based analysis of spoken English ». *ARELE : Annual Review of English Language Education in Japan* 22 (2011) : 377-92.

Shimizu, T., Fukasawa, E., et Yonekura, S. « Introductions and Practices of Speech Acts in Oral Communication 1 Textbooks : From a Viewpoint of Interlanguage Pragmatics ». *department bullentin paper*, nº 55, (2007) : 143-63.

Shirato, J., et Stapleton, P. « Comparing English vocabulary in a spoken learner corpus with a native speaker corpus : Pedagogical implications arising from an empirical study in Japan ». *Language Teaching Research* 11, nº 4, (2007) : 393-412.

Sikora, D., et Tsedryk, A. « De quelques aspects de compétence phraséologique en L2 : Locutions En Test ». *Linguistica Atlantica* 37, nº 2, (2019).

Sinclair, J. *Corpus, Concordance, Collocation*. Oxford : Oxford University Press, (1991).

Smith, A. « L'expression de la fonction phatique en français et en allemand : du concept de phaticité au pilotage du coénonciateur à l'aide des expressions phatiques ». Thèse de Doctorat, Université Nancy 2, (2007).

Steuckardt, A. « Les marqueurs de reformulation formés sur dire : exploration outillée ». *Langages* N° 212, nº 4, (2018) : 17-34.

Surcouf, C. « Les enjeux de la compréhension du français oral quotidien en FLE : création d'une base de données de français parlé annoté ». *Ela. Etudes de linguistique appliquee* N° 198, nº 2, (2020) : 241-56.

Svensson, M.-H. « Critères de figement : L'identification des expressions figées en français contemporain ». Thèse de doctorat, Umeå University, Moderna språk, (2004).

Svensson, M.-H. « L'influence du contexte sur l'interprétation des expressions à sens figuré ». *La phraséologie : sémantique, syntaxe, discours*, (2017) : 217-29.

Sweet, H. *A New English Grammar : Logical and Historical*, 2 vol., Oxford, Clarendon Press. (1891-98)

Taguchi, N., éd. *Pragmatic Competence*. Mouton Series in Pragmatics 5. Berlin : Mouton de Gruyter, (2009).

Taguchi, N., et Roever, C. *Second language pragmatics*. Oxford University Press Oxford, (2017).

Taguchi, N., et Kim, Y. « Collaborative dialogue in learning pragmatics : Pragmatic-related episodes as an opportunity for learning request-making ». *Applied Linguistics* 37, nº 3, (2016) : 416-37.

Taguchi, N., Li, Q., et Tang, X. « Learning Chinese formulaic expressions in a scenario-based interactive environment ». *Foreign Language Annals* 50, nº 4, (2017) : 641-60.

Taguchi, N. « “Contextually” speaking : A survey of pragmatic learning abroad, in class, and online ». *System* 48, (2015) : 3-20.

Taguchi, N. « Description and Explanation of Pragmatic Development : Quantitative, Qualitative, and Mixed Methods Research ». *System* 75, (2018) : 23-32.

Taguchi, N. « Instructed pragmatics at a glance : Where instructional studies were, are, and should be going ». *Language Teaching* 48, (2015) : 1-50.

Taguchi, N. « Longitudinal studies in interlanguage pragmatics ». *Pragmatics across languages and cultures*, (2010) : 333.

Taguchi, N. « Rater variation in the assessment of speech acts ». *Pragmatics* 21, nº 3, (2011) : 453-71.

Taguchi, N. *The Routledge handbook of second language acquisition and pragmatics*. Routledge New York, (2019).

Tamba, I. « Formules et dire proverbial ». *Langages*, (2000) : 110-18.

Tanaka, H., et Campbell, N. « Classification of social laughter in natural conversational speech ». *Computer Speech & Language* 28, nº 1, (2014) : 314-25.

Tateyama, Y. “Explicit and implicit teaching of pragmatic routines : Japanese sumimasen” in Rose & Kasper (eds.), (2001) : 200-222.

Thomas, J. « Pragmatic failure ». *Unpublished MA dissertation. University of Lancaster*, (1981).

Thomas, J. « Cross-cultural pragmatic failure ». *Applied Linguistics*, 4(1), (1983) : 91-112.

Traugott, E. C. « (Inter) subjectivity and (inter) subjectification : A reassessment ». *Subjectification, intersubjectification and grammaticalization* 66, (2010) : 29-74.

Traverso, V. « Aspects de la négociation dans un polylogue ». In *Structures linguistiques et interactionnelles dans le français parlé*, édité par Mosegaard Hansen Maj-Britt in Berit Hansen Anita. Museum Tusculanum Press, University of Copenhagen, Denmark, (2001).

Traverso, V. « Pratiques communicatives en situation : objets et méthode de l’analyse d’interaction ». *Recherche en soins infirmiers* N° 89, nº 2, (2007) : 21-32.

Traverso, V. *Décrire le français parlé en interaction*. Éditions Ophrys, (2016).

Traverso, V. *La conversation familière : analyse pragmatique des interactions*. Presses Universitaires Lyon, (1996).

Trévisiol-Okamura, P. et Marquillo Larruy M. « Distance interlinguistique et stratégies d'apprentissage d'une langue éloigné », Colloque « Analyse des erreurs commises par des francophones apprenant une langue éloignée », Hiroko Ōshima, Jean Bazantay, Rémy

Porquier, et Centre d'études japonaises (Paris), éd. *Apprentissage d'une langue éloignée : analyse des erreurs d'apprenants francophones : actes du colloque*, (2020).

Trévisiol-Okamura, P., et Marquilló Larruy, M. « Plurilinguisme et influence translinguistique en français L3 : quelles implications pour l'enseignement et la formation ? » *Le Français dans le monde. Recherches et applications* 61, (2017).

Trévisiol, P. « Influence translinguistique et alternance codique en français L3. Rôles des L1 et L2 dans la production orale d'apprenants japonais ». *Acquisition et interaction en langue étrangère*, nº 24, (2006) : 13-43.

Trillo, J. R. « The pragmatic fossilization of discourse markers in non-native speakers of English ». *Journal of pragmatics* 34, nº 6, (2002) : 769-84.

Trosborg, A. « Apology strategies in natives/non-natives ». *Journal of pragmatics* 11, nº 2, (1987) : 147-67.

Tuchais, S. « La modulation de l'assertion par to omou et to omowareru en japonais. Approche contrastive japonais / français ». *Syntaxe et sémantique*. Caen : Presses universitaires de Caen, (2015).

Tuchais, S. « Les modulateurs d'assertion en français et en japonais du point de vue de la constructionnalisation ». *Langages* 220, nº 4, (2020) : 43-64.

Tutin, A. « Phrases préfabriquées des interactions : quelques observations sur le corpus CLAPI ». *Cahiers de lexicologie 2019 – 1, n° 114. Les phrases préfabriquées : Sens, fonctions, usages*, (2019) : 63-91.

Ure, J. « Introduction : approaches to the study of register range », *International Journal of Sociology of Language* 35, (1982) : 5-23.

Van Compernolle, R. A., et Henery, A. « Instructed concept appropriation and L2 pragmatic development in the classroom ». *Language Learning* 64, nº 3, (2014) : 549-78.

Vanderveken, D. *Les actes de discours : essai de philosophie du langage et de l'esprit sur la signification des énonciations*. Philosophie et langage. Liège : P. Mardaga, (1989).

Verschueren, J. « Notes on the role of metapragmatic awareness in language use ». *Pragmatics* 10, nº 4, (2000) : 439-56.

Vinogradov, V. V. « Basic concepts of phraseology as a linguistic discipline ». *Leningrad University*, (1946).

Wachs, S. « Passer les frontières des registres en français : un pas à l'école ». *Synergies France* 4, (2005) : 169-77.

Wang, Y. « La compétence phraséologique de la langue seconde chez les sinophones : la compréhension des expressions réactives françaises », Mémoire de Master, Université Grenoble Alpes, (2021).

Warga, M., et Schölmberger, U. « The acquisition of French apologetic behavior in a study abroad context », (2007).

Wojciechowska, J. « Analyse linguistique des phraséologismes pragmatiques partageant la structure *C'est + dét déf + N* », Mémoire de Master, (2019).

Wolk, C., Götz, S. et Jäschke, K. « Possibilities and drawbacks of using an online application for semi-automatic corpus analysis to investigate discourse markers and alternative fluency variables ». *Corpus Pragmatics* 5, nº 1, (2021) : 7-36.

Wray, A., et Perkins, M. « The functions of formulaic language : An integrated model ». *Language & Communication - LANG COMMUN* 20 (31 janvier 2000) : 1-28.

Wray, A. *Formulaic language and the lexicon.* ERIC, (2002).

Wray, A. *Formulaic language : Pushing the boundaries*. Oxford University Press, (2008).

Wu, L. « Frequencies and Semantic Category Distribution of Idioms in Japanese ». *Journal of the Graduate School of Letters* 11 (2016) : 75-82.

Yang, J. « CFL learners' recognition and production of pragmatic routine formulae ». *Chinese as a Second Language. The journal of the Chinese Language Teachers Association, USA* 51, nº 1 (2016) : 29-61.

Yann M. « The Agreement Measure γ_{cat} a Complement to γ Focused on Categorization of a Continuum ». *Computational Linguistics* 43, (2017) : 661–681.

Yokomori, D., Yasui, E., et Hajikano, A. « Registering the receipt of information with a modulated stance : A study of ne-marked other-repetitions in Japanese talk-in-interaction ». *Journal of Pragmatics* 123, (2018) : 167-91.

Yorio, C. A. « Idiomaticity as an indicator of second language proficiency ». *Bilingualism across the lifespan*, (1989) : 55-72.

Zuluaga, A. *Introducción al estudio de las expresiones fijas*. Tubinga, (1980).

奥雅博. « 日本文解析における述語相当の慣用的表現の扱い ». 情報処理学会論文誌 31, nº 12, (1990) : 1-8. [OKU Masahiro. « Nihonbun kaiseki ni okeru jutsugo sôtô no kan'yôteki hyôgen no atsukai ». jôhôshorigakkai ronbunshi 31, nº 12, (1990) : 1-8.]

奥田靖雄. « 語彙的な意味のあり方 » 松本泰丈編『日本語研究の方法』むぎ書房 (1978) : 29-44. [OKUDA Yasuo. « Goiteki na imi no arikata » MATSUMOTO Hirotake hen 『Nihongo kenkyû no hôhô』Mugi shobô (1978) : 29-44.]

岡田祥平. « 『爪痕を残す』の『新用法』について »『日本語学会 2014 年度秋季大会 予稿集』日本語学会, (2014) : 185-192. [OKADA Yôhei. « 『Tsumeato wo nokosu』no 『shin'yôhô』ni tsuite» 『Nihongogakkai 2014 nendo akitaikai yokôshu』Nihongogakkai, (2014) : 185-192.]

沖裕子. « 日本語の身体慣用句一覧 ». 信大日本語教育研究 4, (2004) : 33-85. [OKI Hiroko. « Nihongo no shintaikan'yôku ichiran ». Shindai nihongo kyôiku kenkyû 4, (2004) : 33-85.]

加藤恵梨. « 日常会話でほめことばとして用いられる 「すごい」 について ». ことば 41, (2020) : 106-21. [KATO Eri. « Nichijôkaiwa de homekotoba toshite mochiirareru 「SUGOI」 ni tsuite ». Kotoba 41, (2020) : 106-21.]

宮地裕. « 『成句』の分類 »『語文』第 32 大阪大学国文学研究室編, (1974) : 113-121. [MIYAJI Yutaka. « 『Seiku』no bunrui »『gobun』dai32 Osaka daigaku kokubungaku kenkyûshitsu hen, (1974) : 113-121.]

宮地裕. « 慣用句と連語成句(<特集>日本語の表現—慣用語句、特別な言いまわし—) »『日本語教育』33 号日本語教育学会, (1977) : 1-10. [MIYAJI Yutaka. « Kan'yôku to rengoseiku (<tokushû> nihongo no hyôgen —kan'yôgoku、tokubetsuna iimawashi —) »『Nihongokyôiku』33gô nihongokyôiku gakkai, (1977) : 1-10.]

宮地裕.『慣用句の意味と用法』明治書院, (1982). [MIYAJI Yutaka.『Kan'yôku no imi to yôhô』meiji shoin, (1982).]

橋本力., et 河原大輔. « 日本語慣用句コーパスの構築と慣用句曖昧性解消の試み ». 情報処理学会研究報告自然言語処理 (NL), nº 67, (2008) : 1-6. [HASHIMOTO Chikara., et KAWAHARA Daisuke. « Nihongo kan'yôku koopasu no kôchiku to kan'yôku aimaisei kaisho no kokoromi». Jôhôshori gakkai kenkyû hôkoku shizengengo shori (NL), nº 67, (2008) : 1-6.]

熊谷愛未. « 「やばい」 の現状とコミュニケーション: 私たちは 「やばい」 をどう使うべきなのか ». コミュニケーション文化 8, (2014) : 185-202. [KUMAGAI Aimi. « 「YABAI」 no genjô to komyunikeeshon : watashitachi ha「YABAI」wo dô tsukau beki nanoka ». komyunikeeshon bunka 8, (2014) : 185-202.]

呉琳. « BCCWJ を用いた基幹慣用句の選定 »『研究論集』第 16 号北海道大学大学院文学研究科, (2016a) : 99-113. [WU Lin. « BCCWJ wo mochiita kikan kan'yôku no sentei »『Kenkyû ronshû』dai 16gô Hokkaidô daigaku daigakuin bungaku kenkyûka, (2016a) : 99-113.]

呉琳. « BCCWJ における慣用句の使用度数及び意味分野 »『語彙研究』13 号語彙研究会, (2016b) : 43-50. [WU Lin. « BCCWJ ni okeru kan'yôku no shiyôdosû oyobi imi bunya »『Goi kenkyû』13gô goi kenkyûkai, (2016b) : 43-50]

呉琳. « 慣用句『鳥肌が立つ』の使用実態調査—近現代の用例を中心に— »『日中語 彙研究』第 5 号愛知大学中日大辞典編纂所 , (2016c) : 47-72. [WU Lin. « Kan'yôku『Torihada ga tatsu』no shiyou jittai chosa —kingendai no yôrei wo chûshin ni— »『Nicchû goi kenkyû』dai 5gô aichidaigaku chûnichi daijiten hensansho, (2016c) : 47-72.]

呉琳. «日本語の慣用句に関する研究の概観 »『日中語彙研究』第 6 号愛知大学中日大辞典編纂所(校正中) (2017). [WU Lin. «Nihongo no kan'yôku ni kansuru kenkyû no

gaikan »『Nicchûgoi kenkyû』dai 6gô aichidaigaku chûnichi daijiten hensansho (kôseichû) (2017).]

呉琳. « コーパスに基づく日本語慣用句の研究 ». 北海道大学, (2017). [WU Lin. « koopasu ni motozuku nihongo kan'yôku no kenkyû ». hokkaidô daigaku, (2017).

高田晴夫. « 凝結表現におけるステレオタイプ ». 言語の普遍性と個別性, nº 2, (2011) : 1-12. [TAKADA Haruo. « Gyôketsu hyôgen ni okeru sutereotaipu ». gengo no fuhensei to kobetsusei, nº 2, (2011) : 1-12.]

高木一彦. « 慣用句研究のために »『教育国語』38 号むぎ書房, (1974) : 2-21. [TAKAKI Kazuhiko. « Kan'yôku kenkyû no tame ni »『Kyôiku kokugo』38gô mugi shobô, (1974) : 2-21.]

国広哲弥. « 慣用句論 »『日本語学』1 月号 Vol.4 明治書院, (1985) : 4-14. [KUNIHIRO Tetsuya. « Kan'yôkuron »『Nihongogaku』1gatsugô Vol.4 meiji shoin, (1985) : 4-14]

佐々木文彦. « コーパスを利用した言葉の意味・用法の変化の研究—『敷居が高い』を例に— »『第 4 回コーパス日本語学ワークショップ予稿集』2013 年 9 月国立国語研究所, (2013) : 1-10. [SASAKI Fumihiko. « Koopasu wo riyô shita kotoba no imi・yôhô no henka no kenkyû —『SHIKII GA TAKAI』wo rei ni— »『dai 4 kai koopasu nihongogaku waakushoppu yokôshû』2013nen 9gatsu NINJAL, (2013) : 1-10.]

坂本勉. « 慣用句と比喩—慣用化の度合の観点から— »『言語学研究』1pp.1-21 [SAKAMOTO Tsutomu. « Kan'yôku to hiyu —kan'yôka no doai no kanten kara— »『gengogaku kenkyû』1pp.1-21]

三好裕子. « 連語による語彙指導の有効性の検討 ». 日本語教育, nº 134, (2007[1982]) : 80-89. [MIYOSHI Yûko. « Rengo ni yoru goi shidô no yûkôsei no kentô ». nihongo kyôiku, nº 134, (2007[1982]) : 80-89.]

山岡政紀. « 日本語における配慮表現研究の現状 ». 日本語日本文学, nº 14, (2004) : 17-39. [YAMAOKA Masaki. « Nihongo ni okeru hairyo hyôgen kenkyû no genjo ». nihongo nihonbungaku, nº 14, (2004) : 17-39.]

若松史恵. « 話題開始部の冒頭に現れる言語形式についての一考察—談話標識 「え」「ええと」「でも」「なんか」 に着目して— ». 社会言語科学 24, nº 1, (2021) : 173-88. [WAKAMATSU Fumie. « Wadai kaishibu no bôtô ni arawareru gengo keishiki ni tsuiteno ichi kôsatsu —danwa hyôshiki 「e」「eeto」「demo」「nanka」 ni chakumoku shite— ». shakaigengo kagaku 24, nº 1, (2021) : 173-88.]

首藤公昭, 高橋雅仁, et 田辺利文. « 日本語慣用句機械辞書 ». 研究報告情報基礎とアクセス技術 (IFAT), nº 3, (2012) : 1-12. [SHUDÔ Kôshô, TAKAHASHI Masahito, et TANABE Toshifumi. « Nihongo kan'yôku kikai jisho ». kenkyû hôkoku jôhôkiso to akusesu gijutsu (IFAT), nº 3, (2012) : 1-12.]

小森和子. « 日本語学習者の語彙知識の習得に及ぼす第一言語の影響-中国語を第一言語とする日本語学習者の和語習得を通して », (2014). [KOMORI Kazuko. « Nihongo gakushûsha no goi chishiki no shûtoku ni oyobosu dai ichi gengo no eikyo-chûgokugo wo dai ichi go to suru nihongo gakushûsha no wago shûtoku wo tôshite », (2014).]

小野米一・王婉瑩・松田知子・田原佳世・ジェビットスーザン・張海蓉. « 身体語彙 を含む日本語の慣用句—中国語・英語との対照を通して— »『語文と教育』13 鳴門教育大学国語教育学会, (1999) : 66-84. [ONO Yoneichi・WANG Wan Ying・ MATSUDA Tomoko・TAHARA Kayo・Jevitt Susan ジェビットスーザン・ZHANG Hai Rong. « Shintaigoi wo fukumu nihongo no kan'yôku—chûgokugo・eigo to no taishô wo tôshite— »『gobun to kyôiku』13 naruto kyôiku daigaku kokugo kyôiku gakkai, (1999) : 66-84.]

森田良行. « 動詞慣用句 »『日本語学』1 月号 Vol.4 明治書院, (1985) : 37-44. [MORITA Yoshiyuki. « Dôshi kan'yôku »『Nihongogaku』1gatsugô Vol.4 meiji shoin, (1985) : 37-44.]

西尾寅弥. « 形容詞慣用句 »『日本語学』1 月号 Vol.4 明治書院, (1985) : 45-53. [NISHIO Toraya. « Keiyôshi kan'yôku »『Nihongogaku』1gatsugô Vol.4 meiji shoin, (1985) : 45-53.]

石田プリシラ. « 慣用句の変異形について—形式的固定性をめぐって— »『筑波応用 言語学研究』5, (1998) : 43-56. [ISHIDA Priscilla. « Kan'yôku no henikei ni tsuite —keishikiteki koteisei wo megutte— »『Tsukuba ôyô gengogaku kenkyû』5, (1998) : 43-56.]

石田プリシラ. « 動詞慣用句に対する統語的操作の階層関係 »『日本語科学』7 国立国 語研究所『日本語科学』編集委員会, (2000) : 24-43. [ISHIDA Priscilla. « Dôshikan'yôku ni taisuru tôgoteki sôsa no kaisô kankei »『Nihongokagaku』7 NINJAL『Nihongokagaku』henshû iinkai, (2000) : 24-43.]

石田プリシラ. « 動詞慣用句の意味的固定性を計る方法—統語的操作を手段として— » 『国語学』55(4)日本語学会, (2004) : 42-56. [ISHIDA Priscilla. « Dôshikan'yôku no imiteki koteisei wo hakaru hôhô —tôgoteki sôsa wo shudan toshite— » 『Kokugogaku』 55(4) nihongo gakkai, (2004) : 42-56.]

石田プリシラ.『言語学から見た日本語と英語の慣用句』開拓社, (2015). [ISHIDA Priscilla.『Gengogaku kara mita nihongo to eigo no kan'yôku』kaitakusha, (2015).]

石田プリシラ. 筑波応用言語学研究. « Japanese Idiom variants in corpus data and phraseological dictionaries », (2014). [ISHIDA Priscilla. Tsukuba ôyô gengogaku kenkyû. « Japanese Idiom variants in corpus data and phraseological dictionaries », (2014).]

石田プリシラ. « The Semantic Analysis of Idioms : An Analysis of Verb Phrase Idioms [慣用句の意味分析--《驚き》を表す動詞慣用句・一般動詞を中心に] ». *Tsukuba journal of applied linguistics*, nº 10, (2003) : 1-16. [ISHIDA Priscilla. « The Semantic

Analysis of Idioms : An Analysis of Verb Phrase Idioms [Kan'yôku no imi bunseki--《ODOROKI》wo arawasu dôshikan'yôku・ippandôshi wo chûshin ni] ». *Tsukuba journal of applied linguistics*, nº 10, (2003) : 1-16.]

石田プリシラ. « 日本語慣用句の研究: 慣用句の特性と意味を中心に ». 筑波大学, (2002). [ISHIDA Priscilla. « Nihongo kan'yôku no kenkyû : kan'yôku no tokusei to imi wo chûshin ni ». tsukubai daigaku, (2002).]

石田プリシラ. « 日本語コーパスに見られる慣用句の変化可能性 », ウェブサイト記事, https://www.kenkyusha.co.jp/uploads/lingua/prt/13/IshidaPriscilla1407.html, (2019). [ISHIDA Priscilla. « Nihongo koopasu ni mirareru kan'yôku no henka kanôsei », webusaito kiji, https://www.kenkyusha.co.jp/uploads/lingua/prt/13/IshidaPriscilla1407.html, (2019).]

村田年・山崎誠. « 『手』の慣用句を指標とした文章ジャンルの判別—現代日本語書き言葉均衡コーパスを用いて— »『日本語と日本語教育』39 慶應義塾大学日本語・日本文 化教育センター紀要, (2011) : 75-88. [MURATA Minori・YAMAZAKI Makoto. « 『TE』no kanyôku wo shihyô toshita bunshô janru no hanbetsu —gendai nihongo kakikotoba kinkô koopasu wo mochiite— »『Nihongo to nihongo kyôiku』39 keiôgijukudaigaku nihongo・nihon bunka kyôiku sentaa kikô, (2011) : 75-88.]

村木新次郎. « 慣用句・機能動詞結合・自由な語結合 »『日本語学』1 月号 Vol.4 明治 書院, (1985) : 15-27. [MURAKI Shinjirô. « Kan'yôku・kinôdôshi ketsugô・jiyû na goketsugô »『Nihongogaku』1gatsugô Vol.4 meiji shoin, (1985) : 15-27.]

大坪喜子. « 名詞慣用句—特に隠喩的慣用句について— »『日本語学』1 月号 Vol.4 明 治書院, (1985) : 54-61. [OTSUBO Yoshiko. « Meishi kan'yôku —tokuni inyuteki kan'yôku ni tsuite— »『Nihongogaku』1gatsugô Vol.4 , meiji shoin, (1985) : 54-61.]

中村明, « 語の意味と固定連語の扱い(<特集>日本語の表現—慣用語句、特別な言いまわし—) »『日本語教育』33 号日本語教育学会, (1977) : 43-54. [NAKAMURA Akira, « Go no imi to koteirengo no atsukai (<tokushu> nihongo no hyôgen —kan'yôgoku、tokubetsu na iimawashi—) »『Nihongo kyôiku』33gô nihongo kyôiku gakkai, (1977) : 43-54.]

中村明, « 慣用句と比喩表現 »『日本語学』1 月号 Vol.4 明治書院, (1985) : 28-36. [NAKAMURA Akira , « Kan'yôku to hiyu hyôgen »『Nihongogaku』1gatsugô Vol.4 meijishoin, (1985) : 28-36.]

田中弥生, 柏野和佳子, 角田ゆかり, 伝康晴, 小磯花絵, TANAKA Yayoi, KASHINO Wakako, SUMIDA Yukari, DEN Yasuharu, et KOISO Hanae. « Construction of the Corpus of Everyday Japanese Conversation : On the Methodology of Recording Everyday Conversation ». NINJAL, nº 14, (2018) : 275-92.

土屋智行. « 言語の創造性の基盤としての定型表現—慣用句およびことわざの拡張用法の調査— ». 認知科学 18, nº 2, (2011) : 370-74. [TSUCHIYA Tomoyuki. « Gengo no

sôzôsei no kiban toshite no teikeihyôgen —kan’yôku oyobi kotowaza no kakuchô yôhô no chôsa— ». ninchi kagaku 18, nº 2, (2011) : 370-74.]

土屋智行. « 日本語定型表現の体系的分類に向けて: 特に辞書記述に基づく慣用表現とことわざの分析を中心に ». 言語科学論集 15, (2009) : 55-77. [TSUCHIYA Tomoyuki. « Nihongo teikeihyôgen no taikeiteki bunrui ni mukete : toku ni jishokijutsu ni motozuku kan’yôhyôgen to kotowaza no bunseki wo chûshin ni ». gengokagaku ronshû 15, (2009) : 55-77.]

白石大二. « 慣用と慣用語—国語慣用語論覚書— » 国語学振興会編『現代日本語の研 究』白水社, (1942) : 145-167. [SHIRAISHI Daiji. « Kan’yô to kan’yôgo —kokugo kan’yôgoron oboegaki— » kokugogaku shinkôkai hen『gendai nihongo no kenkyû』hakusuisha, (1942) : 145-167.]

白石大二.『日本語のイディオム』三省堂, (1950). [SHIRAISHI Daiji.『Nihongo no idiomu』Sanseidô, (1950).]

白石大二.『国語慣用句辞典』東京堂, (1969). [SHIRAISHI Daiji.『Kokugo kan’yôku jiten』Tôkyôdô, (1969).]

白石大二. « 慣用句とその種類—高田与清『松屋筆記』を手がかりとして—(<特集>日本語の表現—慣用語句、特別な言いまわし—) »『日本語教育』33 号日本語教育学 会, (1977a) : 11-22. [SHIRAISHI Daiji. « Kan’yôku to sono shurui —TAKADA Tomokiyo『MATSUNOYAHIKKI』wo tegakari to shite— (<tokushû> nihongo no hyôgen —kan’yôgoku、tokubetsuna iimawashi—) »『nihongo kyôiku』33 gô nihongo kyôiku gakkai, (1977a) : 11-22.]

白石大二.『国語慣用句大辞典』東京堂, (1977b). [SHIRAISHI Daiji.『Kokugo kan’yôku daijiten』Tôkyôdô, (1977b).]

飯尾牧子. « 短大生の話し言葉にみる談話標識 「なんか」 の一考察 ». 東洋女子短期大学紀要, nº 38, (2006) : 67-77. [IIO Makiko. « Tandai sei no hanashi kotoba ni miru danwa hyôshiki 「NANKA」no ichi kôsatsu ». tôyôjoshi tankidaigaku kiyô, nº 38, (2006) : 67-77.]

尾嶋憲治, 佐藤理史, et 宇津呂武仁. « 日本語慣用句用例データベースの構築法 ». 言語処理学会 第 12 回 年次大会発 表論文集, (2006) : 456-59. [OJIMA Kenji, SATO Masashi, et UTSURO Takeji. « Nihongo kan’yôku yôrei deetabeesu no kôchikuhô ». gengo shori gakkai dai12 kai nenji taikai happyô ronbunshû, (2006) : 456-59.]

冨樫純一. « 談話標識 「まあ」 について ». 筑波日本語研究, nº 7, (2002) : 15-31. [TOGASHI Jun’ichi. « Danwa hyôshiki 「MAA」ni tsuite ». tsukuba nihongo kenkyû, nº 7, (2002) : 15-31.]

平井一樹. « 上級クラスにおける語彙・表現の習得拡充に関する一考察: 慣用表現から難読漢字までを俯瞰・整理する ». 大阪大学日本語日本文化教育センター授業研究 14 (2016) : 77-88. [HIRAI Kazuki. « Jôkyû kurasu ni okeru goi・hyôgen no shûtoku

kakujû ni kansuru ichi kôsatsu: kan'yôhyôgen kara nandoku kanji made wo fukan・seiri suru ». Ôsaka daigakunihongo bunka kyôiku senta- jugyô kenkyû 14 (2016) : 77-88.]

方小贇. « 第二言語としての日本語の連語習得について ». 宇都宮大学国際学部附属多文化公共圏センター年報, nº 8, (2016) : 112-19. [FANG Xiao. « Dainigengo to shiteno nihongo no rengo shûtoku ni tsuite ». utsunomiya daigaku kokusai gakubu fuzoku tabunka kôkyô senta- nennpô, nº 8, (2016) : 112-19.]

方小贇. « 日本語慣用句の成り立ち: 理論的な枠組みと発生のメカニズム ». 外国文学, nº 63, (2014) : 77-85. [FANG Xiao. « Nihongo kan'yôku no naritachi : rironteki na wakugumi to hassei no mekanizumu ». gaikoku bungaku, nº 63, (2014) : 77-85.]

堀恵子. « 身体語にまつわる慣用表現から日本語と言語文化を学ぶ授業の試み ». 日本語教育方法研究会誌 17, nº 2 (2010) : 42-43. [HORI Keiko. « Shintaigo ni matsuwaru kan'yôhyôgen kara nihongo to gengo bunka wo manabu jugyô no kokoromi ». nihongo kyôiku hôhô kenkyûkaishi 17, nº 2 (2010) : 42-43.]

籾山洋介. « 慣用句の体系的分類—隠喩・換喩・提喩に基づく慣用的意味の成立を中心に— »『名古屋大学国語国文学』第 80 号, (1997) : 29-43. [MOMIYAMA Yôsuke « Kan'yôku no taikeiteki bunrui —inyu・kanyu・teiyu ni motozuku kan'yôteki imi no seiritsu wo chûshin ni— »『Nagoyadaigaku kokugo kokubungaku』dai 80 gô, (1997) : 29-43.]

有薗智美. « 行為のフレームに基づく『目』、『耳』、『鼻』の意味拡張—知覚行為から 高次認識行為へ— »『名古屋学院大学論集言語・文化篇』第 25 巻第 1 号, (2013) : 123-141. [ARIZONO Satomi. « Kôi no fureemu ni motozuku『me』、『mimi』、『hana』no imi kakuchô —chikaku kôi kara kôji ninchi kôi e— »『nagoya gakuin daigaku ronshû gengo・bunka hen』dai 25kan dai 1gô, (2013) : 123-141.]

楊雯淇. « 手続き的意味による日本語談話標識 「なんか」 の分析 ». 国際広報メディア・観光学ジャーナル 24, (2017) : 57-74. [YANG Wenqi. « Tetsuzuki teki imi ni yoru nihongo danwa hyôshiki「NANKA」no bunseki ». kokusai kôhô media ・kankôgaku jaanaru 24, (2017) : 57-74.]

鈴木あすみ. « 日本語における慣用句の逸脱使用がもつ言語機能: 形容詞の反義語への置き換えを手がかりに ». In 言語資源活用ワークショップ発表論文集= *Proceedings of Language Resources Workshop*, 3:105-11. 国立国語研究所, (2018). [SUZUKI Asumi. « Nihongo ni okeru kanyôku no itsudatsu shiyô ga motsu gengo kinô : keiyôshi no hangigo e no okikae wo tegakarini ». In gengo shigen katsuyô waakushoppu happyô ronbunshû= *Proceedings of Language Resources Workshop*, 3:105-11. NINJAL, (2018).]

蔡嘉紘. « 感情の表現形式による分類—直接的表現から客観的表現まで— », (2019). [CAI Jia Hong. « Kanjô no hyôgen keishiki ni yoru bunrui —chokusetsuteki hyôgen kara kyakkanteki hyôgen made— », (2019).]

Sigles et abréviations utilisés

BCCWJ : Balanced Corpus of Contemporary Written Japanese

CCP : Cross-Cultural Pragmatics

CCSARP : Cross-Cultural Study of Speech Act Realization Patterns

CECRL : Cadre européen Commun de Référence pour les Langues

CEJC : Corpus of Everyday Japanese Conversation

CIA : Contrastive Interlanguage Analysis

DCT : Discourse Completion Task

FOU : Français sur Objectif Universitaire

ILP : Interlanguage Pragmatics

LC : Langue Cible

LE : Langue étrangère

LIDILEM : Laboratoire de LInguistique et de DIdactique des Langues Étrangères et Maternelles

LM : Langue Maternelle

LN : Locuteur Natif

LNN : Locuteur Non Natif

MDI : Marques Discursives et Interactionnelles

NINJAL : National Institute for Japanese Language and Linguistics

PhP : Phraséologisme Pragmatique

PhPex : Phraséologisme Pragmatique à fonction Expressive

SLA : Second Language Acquisition

SLP : Second Language Pragmatics

TAL : Traitement Automatique des Langues

TLFi : Trésor de la Langue Française Informatisé

UP : Unité Phraséologique

Table des illustrations

Figure 1 : Structure et articulation de la thèse 19
Figure 2 : Typologie simplifiée des phraséologismes 71
Figure 3 : Articulation des trois approches d'analyse de la production des LNN et des deux champs disciplinaires mobilisés dans notre travail de thèse 100
Figure 4 : Le modèle de la compétence communicative (Hymes 1972) 104
Figure 5 : Le modèle de la compétence communicative (Canale et Swain 1980) 104
Figure 6 : Les différents champs de la pragmatique selon Leech (1983) et Thomas (1981) 105
Figure 7 : Le modèle de la compétence linguistique communicative (Bachman et Palmer 1996). 106
Figure 8 : Le modèle de la compétence communicative (Celce -Murcia et al. 1995) 107
Figure 9 : Le modèle de la compétence de communication selon le CECRL (Conseil de l'Europe 2001) 108
Figure 10 : Révision du modèle de la compétence communicative de Celce-Murcia et al. (1995). (Celce-Murcia 2007) 110
Figure 11: Synthèse chronologique de l'évolution de la compétence communicative et pragmatique (basé sur le schéma de Celce-Murcia (2007)) 116
Figure 12 : Modélisation de la compétence lexicopragmatique et son articulation dans l'acte de communication 119
Figure 13 : Modélisation de notre protocole expérimental en trois phases 151
Figure 14 : Résultats de l'AIA mesuré avec la métrique gamma sur un échantillon d'annotation du corpus Exolingue-Tsukuba 173
Figure 15 : Schéma d'arborescence des annotations en partition sur ELAN 182
Figure 16 : Exemple d'arborescence annotation en partition ELAN 182
Figure 17 : Moyenne des fréquences relatives d'emploi des marques expressives par groupe de locuteurs 186
Figure 18 : Moyenne des fréquences relatives d'emploi des marques discursives et interactionnelles par groupe de locuteur 187
Figure 19 : Moyenne des fréquences relatives d'emploi des marques du registre oral par groupe de locuteurs 189
Figure 20 : Moyenne des fréquences relatives des anomalies interactionnelles entre les locuteurs non natifs des corpus de Grenoble et de Tsukuba 191
Figure 21 : Moyenne des fréquences relatives d'emploi par type de marques et par groupe de locuteurs 192
Figure 22 : Moyenne des fréquences brutes d'emploi des marques discursives et interactionnelles par groupe de locuteurs 209
Figure 23 : Architecture du test lexicopragmatique 229
Figure 24 : Extrait de la vidéo de guidage du test 238
Figure 25 : Exemple d'une question proposée dans le test papier proposé aux LN 238
Figure 26 : Exemple d'une question proposée dans le test en ligne proposé aux LNN avec l'intégration du guide vidéo 239
Figure 27 : Exemple d'une question d'autoévaluation 240
Figure 28 : Exemple du formulaire d'accord interjuge pour les réponses en français 243
Figure 29 : Exemple de synthèse des accords interjuge pour les réponses en français 244
Figure 30 : Consignes de l'accord interjuge pour les réponses en japonais 246
Figure 31 : Informations additionnelles proposées dans l'accord interjuge pour les réponses en japonais 247
Figure 32 : Définition du PhPex et dialogue ciblé dans l'évaluation 247
Figure 33 : Exemple du formulaire proposé dans l'accord interjuge pour les réponses en japonais 248
Figure 34 : Exemple de synthèse des accords interjuge pour les réponses en japonais 248
Figure 35 : Exemple du formulaire proposé dans le classement du registre pour les réponses en français 256
Figure 36 : Exemple du formulaire proposé dans le classement du registre pour les réponses en japonais 256

Figure 37 : Exemple des consignes et des définitions des 4 types de registres proposées aux évaluateurs français 258
Figure 38 : Exemple des consignes et des définitions des 4 types de registres proposées aux évaluateurs japonais 259
Figure 39 : Exemple des synthèses de classement dans les formulaires proposés aux évaluateurs français et japonais 260
Figure 40 : Exemple des consignes et des définitions des caractéristiques des structures préfabriquées proposées aux évaluateurs japonais 262
Figure 41 : Exemple des consignes et des définitions des caractéristiques des structures préfabriquées proposées aux évaluateurs français 263
Figure 42 : Exemple des questions d'autoévaluations proposées dans le test lexicopragmatique . 265
Figure 43 : Questionnaire sociobiographique proposé aux informateurs 269
Figure 44 : Tableau de synthèse des taux de réussite au test lexicopragmatique pour les deux cohortes 271
Figure 45 : Comparaison des taux de réussite par activité entre les LN et les LNN 271
Figure 46 : Taux de réussite et écart type par item et par cohorte pour l'activité 1 273
Figure 47 : Taux de réussite et écart type par item et par cohorte pour l'activité 2 273
Figure 48 : Taux de réussite et écart type par item et par cohorte pour l'activité 3 274
Figure 49 : Taux de réussite et écart type par item et par cohorte pour l'activité 4 275
Figure 50 : Taux de réussite et écart type par item et par cohorte pour l'activité 5 276
Figure 51 : Dispersion des résultats au test lexicopragmatique des LNN 277
Figure 52 : Distribution des résultats au test lexicopragmatique des LNN 280
Figure 53 : Tableau de synthèse des compétences liées à l'interaction orale issu du volume complémentaire du CECRL (2018) 292
Figure 54 : Répartition des LNN en fonction du score de réussite obtenu dans l'activité 3 sur les propositions d'équivalence FR / JP avec mise en parallèle du taux de compréhension 344
Figure 55 : Distribution des choix de réponse des LN et des LNN entre les 4 items proposés dans le dialogue 1 347
Figure 56 : Distribution des choix de réponse des LN et des LNN entre les 4 items proposés dans le dialogue 2 350
Figure 57 : Distribution des choix de réponse des LN et des LNN entre les 4 items proposés dans le dialogue 3 352
Figure 58 : Répartition des LNN en fonction de l'appariement de l'exactitude de la réponse et du type d'autoévaluation pour chaque dialogue avec mise en parallèle du taux de compréhension et du score général 354
Figure 59 : Score de réussite par cohorte pour chaque item dans le dialogue 1 356
Figure 60 : Répartition des choix de réponse par cohorte dans le dialogue 1 357
Figure 61 : Score de réussite par cohorte pour chaque item dans le dialogue 2 359
Figure 62 : Répartition des choix de réponse par cohorte dans le dialogue 2 360

Listes des tableaux

Tableau 1: Métadonnées des différents corpus employés dans l'étude diagnostique..... 163
Tableau 2 : Métadonnées des locuteurs ayant participé à l'étude diagnostique 163
Tableau 3 : Première version de la grille d'annotation basée sur une typologie préliminaire des erreurs observées chez les locuteurs japonais du français B1/B2 167
Tableau 4 : Deuxième version de la grille d'annotation incluant à la précédente version des marques spécifiques de la conversation quotidienne 171
Tableau 5 : Version finale de la grille d'annotation suite aux améliorations suggérées par le processus d'accord interannotateur 178
Tableau 6 : Les trois fonctions du rire chez les LNN observés 202
Tableau 7 : Spécificités du groupe test et du groupe contrôle 232
Tableau 8 : Structure détaillée et contenu du test lexicopragmatique 237
Tableau 9 : Statistiques générales descriptives pour les scores de réussite de chaque cohorte..... 277
Tableau 10 : Statistiques descriptives de la distribution pour la variable des scores de réussite chez les LNN 280
Tableau 11 : Inventaire des variables observées sur les résultats au test des LNN 283
Tableau 12 : Relevé statistique pour chaque variable étudiée sur les résultats au test des LNN ... 284
Tableau 13 : Tableau de comparaison entre les modalités propres à chaque activité, le taux de réussite et le taux de compréhension..... 293
Tableau 14 : Synthèse de la significativité des variables observées chez les LNN..... 294
Tableau 15 : Spécificités lexicales des réponses analysées dans l'activité 1..... 308
Tableau 16 : Spécificités lexicales des réponses analysées dans l'activité 2..... 316
Tableau 17 : Spécificités lexicales des réponses analysées dans l'activité 3..... 329
Tableau 18 : Choix de réponse proposés pour les deux expressions testées dans le dialogue 1 356
Tableau 19 : Choix de réponse proposés pour les expressions testées dans le dialogue 2 359
Tableau 20 : Synthèse générale des statistiques lexicales mesurées sur l'ensemble du test 362
Tableau 21 : Synthèse des catégories de difficulté associées aux trois niveaux de la compétence lexicopragmatique 364
Tableau 22 : Guide d'entretien employé dans notre étude 386
Tableau 23 : Synthèse des données recueillies 388
Tableau 24 : Codage thématique des relevés faits sur les entretiens semi-directifs 391

Annexes

Table des annexes[204]

Annexe 1 Tableau de synthèse des objets linguistiques apparentés aux PhPex 452
Annexe 2 Tableau de synthèse des méthodes d'élicitation de la compétence pragmatique..... 457
Annexe 3 et 4 Consentement éclairé (version FR et JP)..... 462
Annexe 6 Métadonnées des LNN du test lexicopragmatique..... 466
Annexe 7 Métadonnées des LN du test lexicopragmatique..... 469
Annexe 8 Test Lexicopragmatique 471
Annexe 5 Extraits du corpus diagnostic transcrits et annotés sur ELAN 480
Annexe 9 Réponses des LNN au test lexicopragmatique et équivalents fonctionnels en japonais de l'activité 3 (annexe 9bis)..... 480
Annexe 10 Réponses des LN au test lexicopragmatique..... 480
Annexe 11 Enquête en japonais distribuée après la passation du test 480
Annexe 12 Relevés effectués lors des entretiens semi-directifs 480

[204] L'ordre affiché dans la table des annexes n'est pas une erreur, mais est volontaire. Les annexes 5, 9, 10 11, 12 ont été mises à disposition sur une plateforme en ligne en raison de leur volume. Nous avons donc choisi de les réunir ensemble.

Annexe 1
Tableau de synthèse des objets linguistiques apparentés aux PhPex

Dénomination(s) 205	**Scientifique(s) associé(s)**	**Définition**	**Variation diamésique**	**Exemple et contre-exemple**[206]	**Spécificités observées**				
					Syntaxique	Sémantique	Pragmatique	Interactionnel	Type de figement
Phraséologie exclamative	Bally (1909)	Une sous-catégorie regroupant divers procédés linguistiques dont la principale fonction est d'exprimer la subjectivité et le positionnement émotionnel du locuteur. Perspective descriptive	Oral	EX : allons ! / C'est bien fait ! NEX : La séance est levée.	Polylexicalité variable Contrainte combinatoire forte Autonomie	Non compositionnel Non prédictible Sens idiomatique Polysémie Importance de la prosodie	Lien variable avec le contexte Subjectivité intégrée au discours Registre familier Dimension expressive	Fluidifie l'interaction Fonction relationnelle Réactif	Figement syntaxique Figement sémantique Routinisation
Routine conversationnelle / d'interaction	Coulmas (1979, 1981) Traverso (1996)	Expressions et énoncés formulaires utilisés de manière systématique dans les rites d'interaction de la vie tous les jours. Perspective interactionnelle et sociolinguistique	Oral	EX : Bonsoir NEX : c'est le pied !	Polylexicalité variable Contrainte combinatoire forte (mais pas absolues) Autonomie Récurrence forte	Compositionnalité variable Sens idiomatique Polyfonctionnalité Prédictibilité faible Importance de la prosodie	Contrainte forte du contexte (choix préférentiel et conventionnel) Peut être expressif Illocutoire fort Registre variable en fonction des paramètres socio-affectifs Polyfonctionnalité	Modalité dyadique Initiatif ou réactif Souvent en début ou en fin d'interaction Fonction relationnelle et sociale forte Caractère conventionnel	Figement syntaxique Figement pragmatique Routinisation
Énoncés liés	Fónagy (1982) Marque-Pucheux (2007)	Des énoncés caractérisés avant tout par une forte récurrence de l'usage et par un lien interprétatif fort, mais non exclusif avec un contexte prototypique d'emploi. Perspective descriptive	Oral	EX : c'est pas un peu fini ! NEX : peinture fraiche !	Polylexical (sauf interjections) Contrainte combinatoire forte (mais pas absolues)	Non compositionnel (sauf quelques cas rares) Non prédictible Sens idiomatique	Sens lié et activé par un contexte (mais pas de contrainte forte) Expressif, évaluatif	Dyadique Réactif Inséré dans le discours	Figement syntaxique variable Figement pragmatique

[205] Quelques ajouts non-traités dans la thèse ont été fait dans ce tableau. Il s'agit de classes présentant certaines proximités avec les PhPex, mais dont les développements théoriques qui ont été fait jusqu'à présent ne s'orientent pas vers une application à la conversation quotidienne. Toutefois, nous avons jugé utile de mentionner ces travaux ici pour d'éventuels états de l'art ultérieur sur les PhP.

[206] Exemple = EX / contre-exemple = NEX

					Autonomie	Axiologique	Illocutoire fort Registre conversation quotidienne		Figement sémantique Routinisation
Phrases usuelles / Énoncé de motif usuel	Martins Baltar (1995)	Unités lexicalisées fortement idiomatiques qui se caractérisent essentiellement par leur récurrence en situation et une activation en fonction d'un motif. Perspective lexicographique	Oral	EX : Tu te prends pour qui ! NEX : Vos papiers !	Polylexical Contrainte combinatoire forte (mais pas absolues) Autonomie Mémorisé en tant qu'unité	Compositionnalité variable Non prédictible Sens idiomatique Axiologique	Sens lié et activé par un contexte (mais pas de contrainte forte) Expressif, évaluatif Illocutoire fort	Dimension conventionnelle Réactif	Figement pragmatique Figement sémantique
Situation bound utterance (énoncé situationnel)	Kekcsés (1999)	Unités idiomatiques à fonction pragmatique fortement conventionnalisées et préfabriquées, dotées d'une visée sociale explicite, et dont l'apparition est liée à des situations de communication stéréotypées. Perspective descriptive et acquisitionnelle.	Oral Écrit	EX : How is it going ? NEX : you know (MD)	Polylexical Contrainte combinatoire forte (mais pas absolues) Autonomie Utilisé comme une unité lexicale (économie cognitive)	Compositionnalité variable Prédictibilité variable Idiomaticité variable	Sens lié et activé par un contexte (parfois contraintes forte) Expressif, évaluatif Illocutoire fort Prédictibilité fonctionnelle, car lien fort avec le contexte	Renforce l'identité culturelle Conventionnalité Réactif	Figement pragmatique Figement sémantique
Structures figées de la conversation	Bidaud (2002)	Expressions figées de registre principalement familier, bivalentes (selon le contexte peut revêtir de 2 à + contextes), peu imagées ou figuratives et constituant des réponses expressives à des stimulus prenant place dans des contextes de communication stéréotypés Perspective lexicographique, contrastive et didactique	Oral (quotidien)	EX : ça alors ! NEX : Défense de stationner	Polylexical Contrainte combinatoire forte (mais pas absolues) Notion de *figitude*	Non-compositionnelle Idiomaticité Opposition sens syntaxique / sens idiomatique Opacité sémantique	Expressif Contrainte forte du contexte pour l'interprétation Registre familier / populaire Bivalence fonctionnelle	Réactif Conventionnalité Toujours à l'intérieur d'un échange, jamais en initiation ou terminaison	Figement syntaxique Figement sémantique Routinisation
Clichés	Polguère (2008, 2016) Mel'čuk (2011, 2013, 2015)	Phrasèmes polylexicaux et syntaxiquement autonomes (clausatifs) qui présentent une forte idiomaticité, figés morphosyntaxiquement et sémantiquement compositionnels. Perspective lexicographique	Oral	EX : où va la France ! NEX : en joug !	Polylexical Autonome (clausatif) Contrainte combinatoire forte	Compositionnel Idiomaticité Référentiel Iconicité	Parfois expressif ou évaluatif Actualisation de la fonction pragmatique en contexte (pas de contrainte)	Récurrence Conventionnalité Fluidifie l'interaction Construit une image de locuteur légitime	Figement morphosyntaxique Figement conceptuel Figement sémantique
Phrases figées situationnelles / illocutoires	Náray-Szabó (2009)	Les *phrases figées illocutoires* constituent une classe d'expressions idiomatiques phrastiques, utilisée en réaction à une situation conversationnelle spécifique, qui inclut des phrases simples ou complexes, bien	Oral	EX : tu parles ! NEX : Les carottes sont cuites.	Polylexicales Autonome Contrainte combinatoire forte	Généralement non compositionnelles Idiomaticité	Idiomaticité illocutoire Non-compositionnalité pragmatique	Réactives Socialement stéréotypées Conventionnalité	Figement morphosyntaxique Figement pragmatique

		souvent elliptiques. Elles sont dotées d'une valeur illocutoire non déductible à partir de la seule interprétation du sémantisme de leurs constituants. Perspective lexicographique et contrastive			Généralement elliptiques		Activée par une situation (mais pas contrainte) Récurrence Contexte et cotexte		
Pragmatèmes	Mel'čuk (2011 2013) Fléchon, Frassi et Polguère (2012) Blanco et Mejri (2018)	Les pragmatèmes constituent une sous-classe d'unité phraséologique fixe sur le plan formel, fortement contrainte par le contexte d'emploi et dotée d'un sens pragmatique qui produit généralement un effet illocutoire, voire perlocutoire. Perspective lexicographique	Oral Écrit	EX : la séance est levée NEX : Quelle poisse !	Généralement polylexical Autonome Contrainte combinatoire et transformationnelle forte Statut d'énoncé suffisant à lui-même	Généralement compositionnelles Idiomaticité Prédictibilité variable	Fortement contraint par le contexte prototypique d'emploi Illocutoire fort Effet perlocutoire fréquent	Situations sociales protocolaires et conventionnelles Fonction sociale et interactionnelle	Figement syntaxique Figement pragmatique Ritualisation de l'usage
Actes de langage stéréotypés (ALS)	Kauffer (2011)	Une expression à fonction pragmatique autonome sur le plan syntaxique, figée morphosyntaxiquement et pragmatiquement, généralement polylexicale, non- compositionnelle, fortement idiomatique et qui peut comporter un substrat émotionnel. Perspective lexicographique et contrastive	Oral	EX : la belle affaire ! NEX : Garde à vous !	Autonome Généralement polylexical Statut d'énoncé	Non compositionnel Idiomatique Opposition sens littéral sens idiomatique Ancrage déictique	Fonction pragmatique forte Constitue un acte de langage Expressif Dimension actionnelle Interprétation dépendante du contexte et du cotexte	Réactif Agit sur l'interlocuteur et la situation	Figement syntaxique Figement pragmatique
Phrases situationnelles	Anscombre (2012) Klein et Lamiroy(2011, 2016)	Expressions polylexicales et autonomes, plus ou moins figées et dont l'emploi semble particulièrement congruent à certains schémas conversationnels stéréotypés. Perspective descriptive	Oral	EX : le torchon brûle ! NEX : j'y crois pas !	Autonome (mais intégrable) Polylexicalité Statut d'énoncé Phrase complète	Stéréotypie Non compositionnel Idiomatique Opposition sens littéral sens idiomatique Ancrage déictique	Emploi congruent à certaines situations de communication Fonctions pragmatiques diverses (expressive, évaluative, conative, avertissement, etc.) Épisodique	Liées à un schéma conversationnel Récurrence importante	Figement syntaxique variable Figement sémantique variable Routinisation
Formules de la conversation	López-Simó (2016)	Énoncés réactifs complets et autonomes, utilisés dans la conversation quotidienne, fixes sur le plan syntaxique et se caractérisant à la fois par une dimension rituelle, conventionnelle et comportementale.	Oral	EX : Tu te fiches de moi ! NEX : qui vole un œuf vole un bœuf.	Statut d'énoncé Polylexicalité variable Structure elliptique	Idiomaticité Compositionnalité variable	Rituel langagier Illocutoire fort	Acte de langage interactif Réactifs dans certains cas	Ritualisation Routinisation Figement pragmatique

		Perspective typologique, interactionnelle et contrastive			Fixité syntaxique Complétude syntaxique		Fonction pragmatique liée au contexte Lien avec le cadre spatio-temporel et avec les connaissances partagées	Fonctions interactionnelles diverses Structures conventionnelles et comportementales Rôle et placement spécifiques dans la structure interactionnelle	Figement morphosyntaxique variable
Phrases préfabriquées des interactions	Tutin (2019/2020)	Séquences préconstruites généralement non compositionnelles, fréquemment employées dans l'interaction quotidienne, essentiellement polylexicales, complètes sur le plan syntaxique et dotées d'un verbe à temps fini. Perspective typologique	Oral	EX : Tu m'en diras tant ! NEX : Peinture fraîche	Préfabrication Polylexicale Noyau verbal à temps fini Phrase complète	Non compositionnelle Interprétation en contexte / cotexte	Acte de langage Fonction expressive ou évaluative Degré de contrainte du contexte variable Fonction phatique ou conative	Usage en interaction Fonction interactionnelle Fonction interpersonnelle	Figement pragmatique Figement morphosyntaxique Figement sémantique
Phrases préfabriquées	Dostie (2019)	Une séquence autonome et polylexicale, mémorisée ou employée telle quelle, sujette à plusieurs degrés de contraintes sur l'axe paradigmatique / syntagmatique souvent associés à un affaiblissement de la compositionnalité sémantique, qui expriment une dimension actionnelle forte permettant la réalisation d'un acte illocutoire et dotée d'une dimension interactionnelle. Perspective typologique	Écrit Oral	EX : ça va barder ! NEX : chapeau !	Préfabrication Polylexicale Autonomie Statut d'énoncé	Dimension métaphorique Généralement non compositionnelles Interprétable en contexte Usage de déictiques fréquent	Dimension actionnelle forte Peut avoir une fonction évaluative ou expressive Permet d'exprimer la subjectivité Plusieurs niveaux de contrainte au contexte Produit un acte illocutoire	Fonction interactionnelle de régulation Fonction sociale et interpersonnelle Peut constituer un tour de parole Importance des dynamiques émetteur/récepteur	Figement pragmatique Figement morphosyntaxique
Formules expressives de la conversation	Gharbi (2020) Krzyżanowska, Grossmann et Kwapisz-Osadnik (2021)	Séquences autonomes plus ou moins figées, de compositionnalité variable, dont l'emploi est contraint sur le plan pragmatique et dont la principale fonction est d'exprimer l'état mental et affectif du locuteur face à son acte d'énonciation. Perspective lexicographique, traductionnelle et contrastive	Oral	EX : j'y crois pas ! NEX : si je puis dire ainsi.	Fixité variable Caractère préfabriqué Autonomie Polylexicalité Statut d'énoncé	Compositionnalité variable Interprétable en contexte Stéréotypie	Contrainte pragmatique d'emploi Fonction pragmatique forte Réaction subjective sur l'énonciation Évaluation subjective Polyfonctionnels	Réaction conventionnelle Dimension interactionnelle fondamentale Dimension sociale Dimension sociocognitive	Figement morphosyntaxique Routine avec degré de figement contextuel variable

							Acte de langage (au sens large)		

Annexe 2
Tableau de synthèse des méthodes d'élicitation de la compétence pragmatique

Nom du protocole	Scientifique associé	Modalité	Type	Définition	Avantages	Limites
Activité de complétion	Blum-Kulka (1982)	Écrit	Non-interactif	Le LNN doit produire la suite d'un extrait d'interaction oral en prenant en compte les éléments du contexte, mais aussi les éléments linguistiques contenus dans le dialogue.	- Simple à mettre en place - Demande peu de moyens - Permet de moduler les paramètres contextuels à tester de manière à observer précisément une hypothèse - Protocole est particulièrement bien adapté à la comparaison entre langues - Offre un cadre de comparaison rigoureux avec le groupe contrôle - Particulièrement adapté au recueil rapide et massif de données sur une grande population de répondant - Possibilité d'éliciter des phénomènes pragmatiques habituellement complexes à provoquer en situation naturelle de communication	- Dialogue fabriqué par le chercheur en se basant sur ses propres connaissances de locuteur - Parfois peu représentatif de la vie conversationnelle du locuteur - Manque d'authenticité - Permet d'observer la compétence à produire une forme pragmatique en contexte, mais ne sont pas révélateurs des choix linguistiques en contexte réel de communication articulée entre plusieurs locuteurs
Journal de suivi	Cohen (1997) Hassal (2006)	Écrit	Non-interactif	Le LNN doit produire un journal de suivi de ses interactions orales en LC en mentionnant des observations qui peuvent être de l'ordre de réflexions personnelles, d'étonnement interculturel, d'analyse ou d'hypothèses concernant l'usage d'une structure pragmatique ou une manière de communiquer	- Constitue l'une des manières les plus empiriques d'explorer la compréhension métapragmatique des LNN de par sa nature autoréflexive et analytique - Peut servir de complément à d'autres protocoles d'exploration - Particulièrement utile pour observer s'il y existe un décalage entre la connaissance de la fonction de l'expression et la capacité à réaliser un usage effectif de l'expression en question - Permet d'explorer les paramètres liés à l'emploi (par exemple des paramètres contextuels comme la relation, le niveau de politesse ou l'enchainement dans la séquence interactionnelle)	- Caractère autocentré du protocole qui peut rendre les données recueillies fortement influencées par la subjectivité du locuteur - Le LNN peut également choisir de mentionner ou non certains évènements, voire d'altérer certains éléments, notamment ceux qui pourraient attenter à son image de locuteur - Nécessite un traitement très long et qualitatif qui exclut d'emblée les mesures statistiques et la comparabilité intersujet - Mise en place du protocole nécessite un cadrage préalable de la démarche

					- permet l'exploration de la dimension interculturelle et translinguistique	qui peut être lourd et chronophage à constituer
Questionnaires à choix multiples	Carrel (1979)	Écrit	Non-interactif	Le LNN doit, en réponse à la lecture d'un dialogue, choisir entre diverses affirmations ou propositions de réponse celle qui coïncide le plus avec le phénomène pragmatique observé.	- Rapide à mettre en place et permet une collecte massive de données - Facilement formalisable et permet des traitements statistiques corrélatifs et inférentiels - Particulièrement adapté à des analyses comparatives entre différents locuteurs, différentes langues ou différents phénomènes - Grande paramétrabilité du protocole permettant d'observer un large spectre de phénomènes dans diverses conditions d'interaction	- Caractère fabriqué et artificiel qui ne permettrait pas de mesurer de manière précise la compréhension dans des conditions réelles de communication - Très exigeant, car il implique la réalisation simultanée d'une tâche de compréhension, d'analyse et de discrimination pour chaque énoncé proposé - Ne permet pas de mesurer les stratégies interactionnelles, les phénomènes de co-construction, la fluidité ou l'aisance d'interprétation ou de réponse du LNN - Le format du QCM comporte des risques de réponses aléatoires ou de précipitation liés à la fatigue cognitive - L'élaboration des items de question et de réponse est chronophage et nécessite des tests pilotes pour s'assurer que le contenu est clair et sans ambigüité
Questionnaires à échelle	Maeshiba, Yoshinaga, Kasper et Ross (1996)	Écrit	Non-interactif	Suite à la lecture/écoute d'un dialogue, d'une proposition de réponse ou à la réalisation d'une activité, le LNN doit évaluer une forme pragmatique ou les critères d'usage de cette forme à l'aide d'une échelle de Likert.	- Rapide à mettre en place et permet de mener des études statistiques à grande échelle - Permet un contrôle optimal des variables, limitant ainsi les biais ou le bruit dans les réponses - Forte standardisation et reproductibilité permettant de faciliter la comparaison entre les types d'informateurs, les phénomènes observés et les langues ciblées - Forte paramétrabilité des items de question permettant d'explorer à la fois la compétence pragmalinguistique, sociopragmatique et métapragmatique - S'intègre très facilement à d'autres protocoles	- L'élaboration du test est complexe - Le choix des items et des échelles est extrêmement important et doit être minutieusement préparé et testé, ce qui est chronophage - Précipitation ou réponse aléatoire lorsque le test est trop long ou trop répétitif - Manque d'authenticité - Ne permet pas de tester la compétence en situation réelle de communication

Verbalisation de la pensée à voix haute	Rosa et Leow (2004)	Oral	Non-interactif	Consiste à demander au LNN de verbaliser à voix haute pendant ou après la réalisation de l'activité les processus mentaux qu'il mobilise lors de l'interprétation des formes linguistiques qui lui sont proposées.	- Accès aux processus mentaux du LNN, ce qui ne serait pas possible dans des protocoles classiques ou en conditions réelles d'interaction - Permet d'observer en temps réel les stratégies et les indices de contextualisation / indexicaux utilisés par le LNN	- Nécessite la mise en place de conditions expérimentales très strictes et couteuses en ressources matérielles - Le locuteur n'est pas en mesure d'exprimer en temps réel tous ses processus mentaux => problème de fiabilité et de précision de la verbalisation - Sollicitation externe du chercheur lors du processus de verbalisation peut altérer le processus et biaiser les données - Plus la tâche à réaliser est longue, plus le locuteur ne sera plus en mesure de stocker dans sa mémoire toutes les informations qu'il souhaite verbaliser par la suite - Tâche est très lourde sur le plan cognitif est nécessite de l'entrainement - N'est pas fait pour tout le monde, certains locuteurs peuvent trouver la démarche ridicule - La langue de verbalisation et la difficulté de la tâche peuvent substantiellement altérer la qualité des données.
Jeux de rôle	Kasper et Dahl (1991)	Oral	Interactif	Consiste à faire participer les LNN à une simulation de situation d'interaction en attribuant à chacun un rôle social spécifique qui conditionnera la nature de leur production orale ou la fonction communicative attendue.	- Possibilité de simuler des interactions en condition réelle d'échange de tours de parole - Permet de recréer les conditions d'interaction en contexte authentique de communication - Permet d'obtenir des données orales dynamiques à la fois aux niveaux micro et macro	- Caractère artificiel et préparé des données - La situation de communication proposée peut parfois s'éloigner grandement des conditions d'interaction auxquelles le LNN est amené à être confronté - On peut se demander si l'activité reflète les choix réels du locuteur - Ne permet pas une pleine comparabilité des données - Un guidage trop présent peut nuire à la spontanéité des données
					- Permet d'observer un large spectre de constructions discursives et pragmatiques,	- Complexe à mettre en œuvre

Enregistrements non-élicité / semi-élicité	Kasper et Schmidt (1996)	Oral	Interactif	Consiste à enregistrer plusieurs locuteurs non natifs interagissant librement avec d'autres locuteurs non natifs ou natifs de la langue cible, sans intervention directe du chercheur.	ainsi que des phénomènes de co-construction de l'interaction et de la relation interpersonnelle - Spontanéité orale propre au protocole donne lieu à des comportements communicatifs non contrôlés ni anticipés utiles pour la description de la variété orale non native en langue cible - Permet d'observer la capacité d'adaptation en d'anticipation des besoins communicatifs chez le LNN - Permet d'observer la réalisation de l'acte illocutoire et ses effets perlocutoires sur l'interlocuteur - Permet d'observer les échecs de réalisation de l'acte communicatif et la stratégie mobilisée pour remédier à cet échec - L'observation porte sur des données extrêmement proches de ce que produirait le locuteur en situation habituelle de communication	- Ce type de recueil est bien souvent chronophage et couteux à mettre en place - Les données peuvent être volumineuses, mais avec peu d'occurrences du phénomène observé - Impossibilité de contrôler les variables ou de limiter les biais - Le traitement des données est long - Nécessite d'avoir recours à des procédures de validation interannotateur - Non-comparabilité des données sources avec celles d'un corpus contrôle
Entretiens métapragmatiques	Kinginger et Blattner (2008)	Oral	Interactif	Consiste en une discussion guidée entre le chercheur et l'informateur concernant le lien entre formes linguistiques actionnelles, cadre sociopragmatique et organisation de l'interaction.	- Permet au chercheur d'explorer les processus interprétatifs conscients du locuteur et de pouvoir observer les stratégies de reconstitution et/ou d'ajustement de la fonction s'appuyant sur la prise en compte du contexte d'énonciation - Permet d'observer comment l'identité du locuteur peut influencer le choix des formes pragmatiques employées - Constitue un des protocoles les plus empiriques et les plus rigoureux pour explorer chez le LNN les phénomènes de conscientisation et de niveau méta liés à l'usage de structures à fonction pragmatique	- Le protocole produit des données de faible comparabilité et ne permet pas d'analyse statistique - Difficulté de mesurer ou d'évaluer quantitativement le niveau de conscientisation ou la capacité à modéliser des phénomènes pragmatiques chez le LNN - Protocole très difficile à calibrer sans trop guider le processus de réflexion du LNN - La perception qu'a le locuteur des phénomènes est très subjective et difficilement généralisable - Protocole long à mettre en place, nécessitant beaucoup de préparation en amont pour le guide d'entretien

Annexe 3 et 4
Consentement éclairé (version FR et JP)

AUTORISATION
POUR
L'ENREGISTREMENT AUDIO/VIDÉO ET
L'EXPLOITATION DES DONNÉES ENREGISTRÉES

Présentation de l'enquête

Vous avez été contacté pour participer à une enquête comportant un recueil de données parlées qui seront d'abord enregistrées en vidéo et puis transcrites sous forme de texte.

Cette enquête est réalisée par Alexis Ladreyt, Doctorant au laboratoire LIDILEM, laboratoire placé sous la tutelle du CNRS, de l'Université Grenoble Alpes et de l'ILF (Institut de la linguistique française), sous la supervision d'Agnès Tutin (LIDILEM), de Tomoko Higashi (LIDILEM) et de Saburô Aoki (Université Tsukuba).

Dans le cadre de ses activités de recherche et de formation, Alexis Ladreyt analyse notamment les processus linguistiques d'utilisation des structures idiomatiques de l'oral. Ces analyses s'effectuent sur la base d'enregistrements audio et vidéo, qui, une fois transcrits, permettent d'étudier finement les mécanismes par lesquels se construit l'activité d'interaction orale.

Ils sont une base utile pour l'étude de la langue et la grammaire telles qu'elles sont effectivement utilisées dans des situations sociales informelles et non contraintes, et permettent des avancées inédites en sciences du langage.

Les transcriptions et les enregistrements seront par la suite intégrés dans des bases de données en ligne réservées à la recherche en linguistique, et toutes informations personnelles ou en lien avec l'identité de personnes existantes seront systématiquement anonymisées.

Ces recherches ne sont possibles que grâce au consentement des personnes qui acceptent d'être enregistrées. Nous vous demandons par conséquent votre autorisation à procéder aux enregistrements et à l'exploitation des données enregistrées.

Pour toute demande concernant cette autorisation, vous pouvez nous contacter à :
Ladreyt.alexis@gmail.com
alexis.ladreyt@univ-grenoble-alpes.fr

N° de déclaration CNIL : 2-12064

Université Grenoble Alpes, UFR LLASIC / Laboratoire LIDILEM, Bâtiment Stendhal, CS40700, 38058 Grenoble cedex 9

Autorisation

d'enregistrement et d'exploitation de données enregistrées

Je soussigné(e) _____

- autorise l'enregistrement audio et vidéo de l'interaction [« goûter entre amis »], la (date) _____

- autorise ainsi la captation, la reproduction, la diffusion et la représentation de ma voix et de mon image (ci-après désignées par « données »), sous leur forme enregistrée ainsi que sous leur forme transcrite et anonymisée (cf. infra) :

a) à des fins de recherche scientifique (mémoires de master ou thèses, articles scientifiques, exposés à des congrès, séminaires) ;
b) à des fins d'enseignement universitaire (cours et séminaires donnés à des étudiants avancés, à partir du niveau maîtrise, en sciences du langage et en sciences sociales) ;
c) à des fins de recherche scientifique et d'enseignement à partir du site internet d'une plateforme dédiée à la recherche sur la langue parlée en interaction.
d) à des fins de recherche scientifique et d'enseignement lors d'une diffusion éventuelle sur d'autres sites également dédiés à la recherche (par exemple dans le cadre de collaborations européennes) ;
e) à des fins de valorisation des travaux réalisés par le laboratoire LIDILEM, par tous moyens (notamment ouvrages, sites internet, colloques).

- prends acte que pour toutes ces utilisations scientifiques les données ainsi enregistrées seront *anonymisées* : ceci signifie :
a) que les transcriptions de ces données utiliseront des pseudonymes et remplaceront toute information pouvant porter à l'identification des participants ;
b) que les bandes audio qui seront présentées à des conférences ou des cours (généralement sous forme de très courts extraits ne dépassant pas la minute) seront bipées lors de la mention d'un nom, d'une adresse ou d'un numéro de téléphone identifiables (c'est-à-dire couverts par un bruit/son qui les rendra inaudibles) ;

- souhaite que la contrainte supplémentaire suivante soit respectée (floutage du visage, anonymisations supplémentaires,etc.) :

_____

_____

_____

_____

Fait à....., le .....en deux exemplaires, dont un original pour le Laboratoire LIDILEM.
Signature :

Veuillez noter ici votre adresse Mail SVP :

_____

UNIVERSITÉ Grenoble Alpes　筑波大学 University of Tsukuba　LIDILEM

会話の録音・録画へのご協力のお願い
収録データの利用法について

調査の概要

この度は日常会話のデータ収録にご協力いただき、ありがとうございます。

本調査は、CNRS とグルノーブル・アルプ大学、および ILF（フランス国語言語学研究所）の監督下にある LIDIEM 研究所に所属する Alexis Ladreyt（博士課程 1 年生）が実施するものです。また、本研究は Agnès Tutin（LIDILEM）と東伴子（LIDILEM）、青木三郎（筑波大学）の 3 教員の指導の下で進められています。

Alexis Ladreyt は日常会話の言語運用プロセスの分析のために本調査を行い、収集データは言語学、フランス語学、第二言語習得などの分野における研究に寄与するものであります。

録音、録画された音声とスクリプトは、文字化を行い、言語研究のために制限付きアクセスデータベースに加えられます。その際、被験者の個人情報、個人が特定できる情報はすべて匿名化されます。この研究は協力者の皆様のご協力のもと成り立ちます。

ご協力を心より感謝申し上げます。以上の内容をご確認のうえ、会話の録音、録画に関する同意がいただけましたなら、使用承諾書へのご署名をお願いいたします。

本調査に関するご質問等は、下記までご連絡ください：

Ladreyt.alexis@gmail.com
alexis.ladreyt@univ-grenoble-alpes.fr

宜しくお願い致します。

N° de déclaration CNIL : 2-12064

Université Grenoble Alpes, UFR LLASIC / Laboratoire LIDILEM, Bâtiment Stendhal, CS40700, 38058 Grenoble cedex 9

UNIVERSITÉ Grenoble Alpes　筑波大学 University of Tsukuba　LIDILEM

音声の録音、録画と記録データの使用についての同意書

私、（氏名／　　）は以下の項目について同意する。

ー（録音、録画の年月日：　　年　　月　　日）のビデオと音声の記録を許可する

ー以下の目的において、匿名化された被験者の音声および画像の記録、複製、資料としての配布

　ー研究目的（博士論文およびその他の論文、研究記事、学会での発表、セミナー）

　ー大学教育を目的としたもの（修士課程および博士課程の語学教育、社会科学のコースやセミナーにおいて）

　ー言語研究専門の制限付きアクセスウェブサイト

　ーLIDILEM 研究所における研究目的（いかなる方法（本・ウェブサイト・シンポジウムを含む））

ー研究目的で使用される全ての場合において、被験者のデータは匿名化されることを了承する。

　ースクリプトは本人の名前の代わりに仮名を使用し、被験者に関する個人情報などは情報を置き換え、被験者個人が特的できないものとする。

　ー実際に研究または教育場面で使用される録音データにおいて、個人情報（氏名・住所等の個人を特定できるもの）は電子音で音声を被せ、聞き取れないようにする。

ー個人情報の匿名化に関し、上記以外の制約（顔のぼかしなど）を加えたい場合はご記入ください：

_____

_____

_____

_____

本許可証は２枚印刷されます。

１枚目は被験者ご本人がお持ちください。２枚目は LIDIEM 研究所で保管致します。

同意地：

年月日：

署名　：

メールアドレス：

UNIVERSITÉ Grenoble Alpes　筑波大学 University of Tsukuba　LIDILEM

Annexe 6
Métadonnées des LNN du test lexicopragmatique

ID	Niveau[207]	Université	Spécialité	Niveau CECRL	Durée apprentissage	Cadre apprentissage	Statut	Résidence	Durée séjour FR	Contexte conversation	Fréquence conv. FR	Type d'interlocuteur
AM12FJP052020	Doctorat	Université française	SHS	C1	6 à 10 ans	Guidé académique	Académique	France	Entre 4 et 7 ans	Académique	Plusieurs fois par semaine	Familier / proches
AS13HJP042020	Master	Université japonaise	Études françaises / littéraires	C2	15 ans et plus	Autonome	Professionnel	Japon	8 ans et plus	Professionnel	Plusieurs fois par mois et moins	Familier / proches
AY1MJP042020	Doctorat	Université française	Études françaises / littéraires	C1	15 ans et plus	Autonome	Académique	Japon	Entre 4 et 7 ans	Académique	Plusieurs fois par semaine	Familier / proches
ANH14FJP052020	Licence	Université japonaise	Études françaises / littéraires	C1	11 à 15 ans	Autonome	Professionnel	Japon	Entre 1 et 3 ans	Professionnel	Plusieurs fois par mois et moins	Familier / proches
DR15FJP042020	Master	Université japonaise	Études françaises / littéraires	C2	15 ans et plus	Autonome	Professionnel	Japon	8 ans et plus	Professionnel	Plusieurs fois par mois et moins	Pro / académique
HY16FCH112019	Doctorat	Université japonaise	ScL[208]	B2	11 à 15 ans	Autonome	Académique	Japon	Entre 4 et 7 ans	Académique	Plusieurs fois par jour	Familier / proches
HK17HJP062020	Doctorat	Université japonaise	ScL	C1	6 à 10 ans	Autonome	Académique	Japon	Entre 1 et 3 ans	Académique	Plusieurs fois par semaine	Pro / académique
IM7FJP042020	Licence	Université japonaise	Études françaises / littéraires	B2	1 à 5 ans	Guidé académique	Académique	Japon	Entre 1 et 3 ans	Académique	Plusieurs fois par semaine	Pro / académique
IM18FJP052020	Master	Université française	ScL	C2	15 ans et plus	Autonome	Académique	France	8 ans et plus	Professionnel	Plusieurs fois par jour	Locuteur lambda
KM19FJP052020	Licence	Université japonaise	Études françaises / littéraires	B2	6 à 10 ans	Guidé académique	Professionnel	Japon	Entre 1 et 3 ans	Perso	Plusieurs fois par mois et moins	Familier / proches
KY20FJP052020	Licence	Université japonaise	Études françaises / littéraires	C1	11 à 15 ans	Guidé académique	Professionnel	France	8 ans et plus	Professionnel	Plusieurs fois par jour	Locuteur lambda
KY21FJP042020	Master	Université française	Études françaises / littéraires	B2	11 à 15 ans	Guidé académique	Professionnel	Japon	Entre 4 et 7 ans	Perso	Plusieurs fois par semaine	Pro / académique
KT8FJP042020	Master	Université française	SHS	C2	1 à 5 ans	Autonome	Professionnel	France	Entre 4 et 7 ans	Perso	Plusieurs fois par jour	Locuteur lambda

[207207] Le niveau *doctorat* réfère à des doctorants préparant le diplôme. Les niveaux *master* et *licence* réfèrent uniquement à des LNN ayant déjà acquis ce diplôme.
[208] ScL = Sciences du Langage

KM22FJP052020	Licence	Université japonaise	Études françaises / littéraires	C1	1 à 5 ans	Guidé académique	Académique	Japon	Entre 1 et 3 ans	Perso	Plusieurs fois par jour	Pro / académique
KC23FJP052020	Licence	Université japonaise	Études françaises / littéraires	C2	15 ans et plus	Guidé non académique	Professionnel	Japon	8 ans et plus	Professionnel	Plusieurs fois par jour	Pro / académique
KM2FJP052020	Doctorat	Université française	ScL	C2	6 à 10 ans	Autonome	Académique	France	Entre 1 et 3 ans	Perso	Plusieurs fois par jour	Locuteur lambda
KK33FJP062020	Master	Université japonaise	ScL	C2	15 ans et plus	Guidé non académique	Académique	Japon	8 ans et plus	Perso	Plusieurs fois par mois et moins	Pro / académique
KH9FJP042020	Master	Université française	ScL	C1	1 à 5 ans	Autonome	Académique	France	Entre 1 et 3 ans	Perso	Plusieurs fois par jour	Locuteur lambda
MR24FJP062020	Doctorat	Université japonaise	ScL	B2	6 à 10 ans	Guidé académique	Académique	Japon	Entre 1 et 3 ans	Professionnel	Plusieurs fois par semaine	Pro / académique
MAM25HJPFR042020	Licence	Université japonaise	Études françaises / littéraires	C2	15 ans et plus	Guidé non académique	Professionnel	Japon	Entre 4 et 7 ans	Perso	Plusieurs fois par semaine	Pro / académique
MA26FJP052020	Master	Université japonaise	SHS	C1	6 à 10 ans	Autonome	Professionnel	France	Entre 4 et 7 ans	Perso	Plusieurs fois par jour	Pro / académique
MN33FJP062020	Doctorat	Université française	ScL	C2	11 à 15 ans	Autonome	Académique	France	8 ans et plus	Professionnel	Plusieurs fois par semaine	Locuteur lambda
MN4FJP042020	Doctorat	Université française	Études françaises / littéraires	C2	6 à 10 ans	Guidé académique	Académique	France	Entre 1 et 3 ans	Académique	Plusieurs fois par mois et moins	Pro / académique
NM3FJP042020	Master	Université française	ScL	C2	15 ans et plus	Autonome	Académique	France	8 ans et plus	Perso	Plusieurs fois par jour	Pro / académique
NM27FJP052020	Master	Université française	SHS	C1	11 à 15 ans	Autonome	Professionnel	France	Entre 4 et 7 ans	Perso	Plusieurs fois par jour	Locuteur lambda
NK34FJP072020	Master	Université japonaise	ScL	B2	1 à 5 ans	Autonome	Académique	Japon	Entre 1 et 3 ans	Perso	Plusieurs fois par mois et moins	Familier / proches
OT28FJP042020	Licence	Université japonaise	Études françaises / littéraires	C2	6 à 10 ans	Guidé non académique	Académique	Japon	Moins d'un an	Perso	Plusieurs fois par semaine	Pro / académique
OK29FJP052020	Doctorat	Université japonaise	ScL	C1	6 à 10 ans	Autonome	Académique	Japon	Moins d'un an	Académique	Plusieurs fois par mois et moins	Locuteur lambda
SN10FJP052020	Master	Université française	SHS	C1	6 à 10 ans	Autonome	Académique	France	Entre 4 et 7 ans	Perso	Plusieurs fois par jour	Familier / proches
SY11FJP042020	Master	Université française	Études françaises / littéraires	C2	11 à 15 ans	Guidé académique	Académique	France	8 ans et plus	Perso	Plusieurs fois par jour	Pro / académique
SA30HJP052020	Doctorat	Université japonaise	ScL	B2	6 à 10 ans	Autonome	Académique	Japon	Moins d'un an	Académique	Plusieurs fois par mois et moins	Locuteur lambda
TS6FJP052020	Master	Université japonaise	ScL	C1	1 à 5 ans	Guidé académique	Académique	Japon	Moins d'un an	Perso	Plusieurs fois par mois et moins	Familier / proches
TM31FJP052020	Master	Université française	ScL	C1	6 à 10 ans	Guidé académique	Académique	France	Entre 4 et 7 ans	Perso	Plusieurs fois par jour	Pro / académique
UR5FJP042020	Doctorat	Université française	ScL	C2	15 ans et plus	Guidé académique	Académique	France	8 ans et plus	Perso	Plusieurs fois par jour	Familier / proches

YN32FJP042020	Master	Université française	SHS	C2	1 à 5 ans	Autonome	Professionnel	France	Entre 4 et 7 ans	Perso	Plusieurs fois par jour	Pro / académique

Annexe 7
Métadonnées des LN du test lexicopragmatique

ID	Université	Niveau d'étude	Spécialité d'étude	Lieu de résidence
CN01FFR032020	UGA[209]	L2	LEA EN/JP	GRENOBLE
DQ02HFR032020	UGA	L2	LEA EN/JP	GRENOBLE
AP03HFR032020	UGA	L2	LEA EN/JP	GRENOBLE
SN04FFR032020	UGA	L2	LEA EN/JP	GRENOBLE
PA05FFR032020	UGA	L2	LEA EN/JP	GRENOBLE
BN06FFR032020	UGA	L2	LEA EN/JP	GRENOBLE
SH08FFR032020	UGA	L2	Double Licence Eco-gé + LEA EN/JP	GRENOBLE
JB09FFR032020	UGA	L2	LEA EN/JP	GRENOBLE
CM10FFR032020	UGA	L2	LEA EN/JP	GRENOBLE
ML11HFR032020	UGA	L2	Double Licence Eco-gé + LEA EN/JP	GRENOBLE
AF12HFR032020	UGA	L2	LEA EN/JP	GRENOBLE
FJ15FFR032020	UGA	L2	LEA EN/JP	GRENOBLE
AA16HFR032020	UGA	L2	LEA EN/JP	GRENOBLE
GE17HFR032020	UGA	L2	LEA EN/JP	GRENOBLE
DT18HFR032020	UGA	L2	LEA EN/JP	GRENOBLE
BC19HFR032020	UGA	L2	LEA EN/JP	GRENOBLE
CK20FFR032020	UGA	L2	LEA EN/JP	GRENOBLE
FB21HFR032020	UGA	L2	LEA EN/JP	GRENOBLE
BV22HFR032020	UGA	L2	LEA EN/JP	GRENOBLE
LE25HFR0302020	UGA	L2	LEA EN/JP	GRENOBLE
PS26HFR032020	UGA	L2	LEA EN/JP	GRENOBLE
GV27HFR032020	UGA	L2	LEA EN/JP	GRENOBLE
TY29HFR032020	UGA	L2	LEA EN/JP	GRENOBLE

[209] UGA = Université Grenoble Alpes

RM31FFR032020	UGA	L2	LEA EN/JP	GRENOBLE
LL32HFR032020	UGA	L2	LEA EN/JP	GRENOBLE
MA33HFR032020	UGA	L2	LEA EN/JP	GRENOBLE
MA34HFR032020	UGA	L2	LEA EN/JP	GRENOBLE
ML36FFR032020	UGA	L2	LEA EN/JP	GRENOBLE
PL37FFR032020	UGA	L2	LEA EN/JP	GRENOBLE
RA38FFR032020	UGA	L2	LEA EN/JP	GRENOBLE
RR39HFR032020	UGA	L2	LEA EN/JP	GRENOBLE
BL40FFR012020	UGA	M1	LINGUISTIQUE	GRENOBLE
CM42FFR012020	UGA	M1	LINGUISTIQUE	GRENOBLE
PA43FFR012020	UGA	M1	LINGUISTIQUE	GRENOBLE
FM44FFR012020	UGA	M1	LINGUISTIQUE	GRENOBLE

Annexe 8
Test Lexicopragmatique

Questionnaire sur les expressions de l'oral quotidien

Version à distance : https://forms.gle/o1XTPMqLa3EXg5pg7

Nous sollicitons aujourd'hui votre participation dans le cadre de notre étude sur les expressions de l'oral quotidien. Cette étude comporte diverses tâches[210] *visant à étudier votre pratique linguistique dans la vie de tous les jours. Elle nous permettra de mener des analyses dans le cadre de notre travail de thèse. Nous vous remercions d'avance pour votre aide.*

Information sur le participant à l'étude :

Êtes-vous locuteur natif du français ? : oui | non

- **Âge :**
- **Sexe : Homme | Femme | Autre**
- **Occupation : Salarié | Étudiant | autres**
- **Si vous êtes étudiant, quels sont votre niveau d'étude et votre spécialité ? :**

- **Si vous n'êtes pas locuteur natif du français, quelle est votre langue maternelle ? :**
  - **Depuis combien de temps étudiez-vous le français :**
  - **Avez-vous déjà fait évaluer votre niveau (TCF/ DELF par exemple) : oui | non**
    - **Si votre réponse est « oui », pourriez-vous entourer votre niveau : A1 | A2 | B1 | B2 | C1 | C2**
  - **Avez-vous déjà séjourné en France ? : oui | non**
    - **Si votre réponse est oui, pourriez-vous indiquer la durée de votre séjour :**

[210] Certains des dialogues utilisés dans cette étude sont extraits du corpus oral en linge CLAPI (http://clapi.ish-lyon.cnrs.fr) et ont été remaniés pour les besoins de cette étude

- **De quelle manière avez-vous appris le français : à l'école | Tout seul | en vivant en France**
- **Discutez-vous souvent en français : oui | non**
  - **Si votre réponse est oui :**
    - **Combien de fois par semaine / jour ? :**
    - **Avec qui ? :**

Activité 1

Consigne : Complétez les dialogues ci-dessous en utilisant des expressions appropriées à la situation décrite. Vous pouvez répondre avec plusieurs phrases. Par exemple :

Ami : ***Salut ! ça va ?***
Vous : ***Ça va et toi ?***
Ami : ***Ça va ! alors quoi de nouveau ?***
Vous : ***Je reviens d'un concert de Coldplay,*** **[C'était ….** ***Le pied ! un truc de dingue ! j'étais au paradis ! La réalisation, le son, le groupe c'était de la folie !*** **]**

❖ **Dialogue 1 : <u>Exprimez votre colère en utilisant de préférence une expression toute faite</u> :**

Ami : Ah bah t'as l'air malheureux ! t'as pu régler ton problème avec Julie ?
Vous : Non toujours pas, ça m'énerve ! C'est …

……

……

❖ **Dialogue 2 : <u>Exprimer votre surprise ou votre indignation en utilisant de préférence une expression toute faite</u>** :

Amie : En rentrant chez moi hier soir, j'ai croisé deux gars louches qui commencent à me suivre. J'accélère pour les distancer et eux aussi. Heureusement, j'ai réussi à rentrer dans mon hall d'entrée et refermer la porte avant qu'ils n'entrent. J'ai eu très peur. J'en tremble encore…

Vous : Sérieux ? …..

……

……

Activité 2

Consigne : Observez les dialogues suivants et reformulez les expressions soulignées. Par exemple :

Léa, Ivan et Cédric discutent à propos de la sœur de Léa.

Léa : ***Ma sœur, pour le travail saisonnier, elle est allée chercher aux impôts***
Ivan : ***Ah ouais ?***
Léa : ***Et elle a été prise ! Elle a quand même une chance elle, un truc international !***
Cédric : ***C'est pas faux !***
Léa : ***donc elle a été prise aux impôts, donc elle était trop contente, mais c'était super dur hein, les impôts !***
Ivan : ***Ah bon ?***
Léa : ***ouais, et c'était une trop sale ambiance en plus***
Ivan : ***tu m'étonnes = [Tu as raison, c'est sûr !]***
Cédric : *ouais, y a moyen, j'imagine pas*

❖ **Dialogue 1**

Alice, Billie et Chris discutent à propos d'un salon de massage.

Alice : Tu te rappelles du chèque cadeau que mon copain m'a offert ?
Billie : Ah oui pour ton anniversaire, c'est ça ?
Alice : Et bah je l'ai utilisé hier dans le salon de massage qui est vers la mairie.
Billie : Ah super !
Chris : Alors, c'est comment ?
Alice : Ah bah **c'est le pied !** = _____ Ils te font la totale de la tête aux pieds avec les huiles essentielles !
Chris : Faudra que j'offre ça à ma copine alors !

❖ **Dialogue 2**

Louise et Jean discutent à propos d'un film au cinéma.

Louise : Le nouveau film du réalisateur des Harry Potter est génial, une vraie merveille !
Jean: Ah bon ? je me tâtais pour aller le voir.
Louise : Et bah tu peux y aller les yeux fermés !
Jean: Ah super ! ça te dit de le revoir avec moi genre demain ?
Louise : Demain ? **c'est pas possible =** _____
Jean: **Ah bon, dommage….**
Louise : Désolée, je suis de fermeture au boulot demain, je vais finir trop tard pour la séance, on se remet ça à une autre fois !

❖ **Dialogue 3**

Arthur et Brice discutent à propos d'un dossier à rendre pour un cours.

Arthur : Ah il est galère ce dossier ! je sais pas si je vais finir à temps.
Brice : Ouais pareil ! trente pages à rendre pour dans une semaine, **faut pas pousser !** =_____ On n'est pas des robots non plus !
Arthur : Je suis d'accord, **c'est abusé !** =_____Ils ne se rendent pas compte que l'on a d'autres cours à côté…

❖ **Dialogue 4**

Alain et Béatrice discutent à propos du téléphone d'Alain qui vient de tomber par terre.

Alain : Hey ! tu peux pas faire attention quand tu passes ? **C'est pas possible** !_____
Béatrice : bah **j'y suis pour rien** !_____c'est toi qui l'as mal posé sur le bord de la table !
Alain : Arrête de dire des bêtises, tu l'as touché en passant, c'est de ta faute !

Activité 3

Consigne : Observez les dialogues suivants et proposez une ou plusieurs expressions équivalentes (utilisable dans une situation de dialogue équivalente) pour chacune des expressions soulignées dans les dialogues ci-dessous. Pour les locuteurs non-natifs, en plus de l'équivalent en français, proposez une expression équivalente dans votre langue maternelle. Par exemple :

Adèle et Robin discutent de leur projet pour la fête d'Halloween.

Adèle : ***Moi j'ai pas fait Halloween, mais j'étais super contente***
Robin : ***Bah pourquoi ?***
Adèle : ***J'aime pas me déguiser, ça me gave ! = [ça me saoule / ça m'énerve] je vais avoir l'air con si j'arrive en bas de pyjama moi***
Robin : ***Mais non c'est pas grave !***

❖ **Dialogue 1**

Enzo, Clarence et Paul discutent à propos de la sœur de Clarence. Celle-ci travaille dans une entreprise à l'ambiance de travail assez difficile.

Enzo : Ta soeur ça se passe bien son nouveau travail ?

Clarence : Bof, elle a un peu du mal avec son patron qui est trop exigeant, il la lâche pas d'une semelle ! Dimanche dernier, il l'a appelé pour qu'elle passe au bureau récupérer des dossiers pour lui.

Paul : **J'y crois pas !** = _____ elle devrait lâcher ce boulot pour un autre plus relax.

Clarence : Ouais ta raison !

Proposez une expression dans votre langue maternelle qui peut être employée de la même manière que l'expression soulignée dans le dialogue ?

……

❖ **Dialogue 2**

Yves, Éric et Isabelle discutent à propos d'un ami commun.

Yves : Vous avez des nouvelles de Julien ? Ça fait un mois qu'il vient plus en cours…

Éric : C'est vrai qu'on le voit plus en ce moment…

Isabelle : Paraît qu'il a eu un accident de vélo, une voiture qui lui a coupé la route …

Éric : **C'est pas vrai !** = _____

Yves : **C'est chaud !** = _____

Isabelle : Je te jure…

Proposez une expression dans votre langue maternelle qui peut être employée de la même manière que l'expression soulignée dans le dialogue ?

1)……

2)……

❖ **Dialogue 3**

Julien et Rémi discutent à propos du déménagement de Rémi.

Julien : Si je passe demain vers 9 h ça te va ?

Remi : Nickel !
Julien : Et du coup Alex viendra ?
Remi : Ah bah là, **c'est pas gagné !** = _____ il est de sortie ce soir, ah mon avis il aura pas le courage de se lever.
Julien : **Tu plaisantes ?** = _____ on a vraiment besoin de lui pour conduire le camion !

Proposez une expression dans votre langue maternelle qui peut être employée de la même manière que l'expression soulignée dans le dialogue ?

1)……

2)……

Activité 4

Consigne : Observez les dialogues suivants et entourez une seule réponse qui vous semble appropriée. Par exemple :

Jean et Philipe discutent au bar.

Jean : *Des fois, je crois rêver quand j'entends dire que les professeurs des écoles sont trop payés.*
Philipe : *Oui, j'ai entendu ça à la radio, les politiciens critiquent beaucoup les enseignants en ce moment*
Jean : *Quand on les écoute, c'est limite si c'est les vacances tous les jours pour nous, [**c'est génial | c'est énorme ! | on nage en plein délire ! | la belle affaire !**]*
Philipe : *Oui, je t'avoue que c'est énervant à la longue ….*

❖ **Dialogue 1**

Florian et Évelyne discutent à propos d'un ami.
Évelyne : Je t'ai dit que j'étais allée à sa crémaillère et que ça s'est bien passé entre nous ?
Florian : Oui, t'en as parlé avant-hier me semble
Évelyne : Mais depuis pas de nouvelles de lui, genre un fantôme
Florian : C'est pas vrai ?
Évelyne : Du coup je lui ai envoyé une tonne de messages, mais il répond jamais ! Alors que je le vois tout le temps connecté sur Facebook,
[C'est génial ! | c'est le pied ! | c'est le bouquet ! | c'est une horreur !]

❖ **Dialogue 2**

Paul, Julie et Pierre discutent à propos des fumeurs au travail qui bénéficient de plus de temps de pause que les non-fumeurs.

Paul : Ah ben les fumeurs ils sont comme ça !
Julie : C'est quatre ou cinq pauses dans la journée hein
Paul : Surtout maintenant qu'on pointe plus nos horaires, c'est souvent ceux qui fument qui se font plaisir sur la pause
Pierre : Ouais ces fumeurs **[c'est génial ! | c'est pas gagné ! | ça tombe bien ! | c'est une horreur !]**
Julie : Oh oui ça c'est clair !

❖ **Dialogue 3**

Jean et Anna discutent à propos de l'actualité.

Anna : Je lisais mon journal tranquillement dans le train et là, je trouve un article disant que le maire promet de rénover la bibliothèque universitaire cette année !
Jean : **[Tu crains ! | Tu parles ! | T'inquiètes ! | Tu l'as dit !]**, jamais de la vie il mènera ce projet à terme. Chaque année, c'est toujours la même rengaine !
Anna : Oui, mais avoue que ce serait quand même le pied de bosser dans une bibliothèque toute neuve !
Jean : Si tu le dis…

Activité 5

Consigne : Observez les dialogues suivants et cochez la réponse qui vous semble appropriée dans les encadrés à droite. Par exemple :

Pierrick et Vivien discutent sur le lieu de leur travail

Pierrick : ***Tu as vu que le manager a encore changé nos horaires***
Vivien : ***Non mais je rêve ! Faut vraiment qu'il arrête de faire ça au dernier moment !***
Pierrick : ***Tu l'as dit ! c'est vraiment pas pratique, on peut jamais rien prévoir…***

Dans le dialogue ci-dessus, l'expression :

non mais je rêve!	Vivien est en train de rêver qu'il parle à Pierrick	
	Vivien est exaspéré par ce que Pierrick vient de dire	**X**
	Vivien est surpris par ce que Pierrick a dit	

❖ **Dialogue 1**

Pierre et Éric discutent à propos d'un dossier à rendre pour un cours.

Pierre : Tu as vu que Mr Brun nous a encore donné un dossier à rendre pour la semaine prochaine
Éric : **Tu es sérieux ?** On n'aura jamais le temps de finir !
Pierre : Je sais bien, mais on n'a pas trop le choix…
Éric : En vrai**, c'est pas possible**. Il nous faudra forcément une semaine de plus.

Dans le dialogue ci-dessus, l'expression :

Tu es sérieux ?	Éric demande à Pierre si c'est une personne sérieuse	
	Éric est exaspéré par ce que Pierre vient de dire	
	Éric demande à Pierre si ce qu'il vient de dire est vrai	

C'est pas possible	Éric pense que rendre le dossier dans ces délais est impossible	
	Éric est exaspéré par ces délais injustes	
	Éric est surpris par ces délais injustes	

❖ **Dialogue 2**

Paul est en train de discuter avec Éric à la pause repas :

Paul : Non, mais je ne te raconte pas, à cause des bouchons, hier j'ai mis deux heures à rentrer du boulot !
Éric : **J'y crois pas !** ça craint, T'es passé par la rocade ?
Paul : Oui, mais avec les travaux on croirait rêver, 10 km de bouchons tous les soirs …
Éric : **Les bras m'en tombent !** depuis le temps que les travaux durent, **c'est une honte !**
Paul : Ah au fait, la dernière fois tu m'avais parlé de rougeurs qui te démangeaient, c'était sur quelle partie du corps déjà ?
Éric : **C'est le pied**. Ça m'arrive surtout le soir après la douche.
Paul : Ok j'en parlerai à mon frangin, il pourra te recevoir en consultation rapidement je pense.
Éric : Merci tu gères mec !

Cocher le sens qui correspond à l'expression :

J'y crois pas	Éric ne croit pas ce que Paul lui raconte.	

	Éric est exaspéré et étonné par ce que Paul lui raconte.	
	Éric est d'accord avec ce que Paul lui raconte.	

Les bras m'en tombent !	Éric est motivé.	
	Éric est surpris et exprime son incompréhension.	
	Les bras d'Éric sont tombés.	

C'est une honte !	Éric est en colère et n'est pas d'accord avec la situation.	
	Éric a honte de lui-même.	
	Éric est étonné par la situation.	

C'est le pied.	Éric pense que les démangeaisons sont très agréables	
	Éric parle de ses pieds	
	Éric est à pied pour rentrer chez lui	

Autorisation du participant à l'usage des données recueillies

Nous vous remercions chaleureusement d'avoir pris le temps de répondre à cette enquête. Nous vous rappelons que les données collectées seront anonymisées et ne servirons que pour le strict usage de ce travail de thèse. Ces données ne seront pas réutilisées ni diffusées sans votre accord. Si vous souhaitez d'autres détails au sujet de cette étude, n'hésitez pas à nous contacter à l'adresse suivante : **alexis.ladreyt@univ-grenoble-alpes.fr**.
J'autorise l'utilisation et la diffusion des données de cette étude à des fins de recherche et/ou d'enseignement et sous couvert d'anonymat.

Date : **NOM + PRÉNOM + Signature :**

Annexe 5
Extraits du corpus diagnostic transcrits et annotés sur ELAN[211]

Annexe 9
Réponses des LNN au test lexicopragmatique et équivalents fonctionnels en japonais de l'activité 3 (annexe 9bis)

Annexe 10
Réponses des LN au test lexicopragmatique

Annexe 11
Enquête en japonais distribuée après la passation du test[212]

Annexe 12
Relevés effectués lors des entretiens semi-directifs

Les annexes 5, 9, 10, 11, 12 étant trop volumineuses, elles sont disponibles à l'adresse suivante : https://bit.ly/3yG7KOeez

Il suffit soit de cliquer sur lien dans la thèse en PDF, soit de recopier manuellement l'adresse ou de la copier-coller dans le navigateur. Vous pouvez également utiliser le QR code suivant :

[211] Il faut télécharger le logiciel ELAN au préalable pour pouvoir consulter les fichiers d'annotation/transcription.
[212] Questionnaire disponible en ligne : https://forms.gle/uzuTwFzwPU6reY98Aeer

Pour utiliser ce QR code :

1) Ouvrir appareil photo du téléphone et toucher le lien qui s'affiche à l'écran puis une fois sur la page, s'envoyer la page par mail ou sur via Airdrop.

OU

2) Télécharger une application lecteur QR code et refaire la procédure 1) à l'aide cette application.

Table des matières

REMERCIEMENTS 2

SOMMAIRE 8

NORME DE PRESENTATION DES CITATIONS EN JAPONAIS 12

INTRODUCTION 14

PARTIE 1 - CADRAGE THÉORIQUE 20

PREMIER VOLET DU CADRE THEORIQUE - DE LA PHRASEOLOGIE GENERALE AUX PHRASEOLOGISMES PRAGMATIQUES A FONCTION EXPRESSIVE 22

CHAPITRE 1 - ÉTAT DES LIEUX DU DOMAINE DE LA PHRASEOLOGIE 26

1. Le champ de la phraséologie *26*

1.1. Survol historique 26

1.2. Statut et objet de la phraséologie 28

1.2.1. La fixité combinatoire 28

1.2.2. Le statut d'unité 30

1.2.3. L'idiomaticité sémantique 31

2. Le domaine de la phraséologie au Japon *32*

2.1. Shiraïshi le précurseur (1942) 32

2.2. Stagnation puis émergence du domaine de la phraséologie au Japon 33

2.3. Les années 80 : Renouveau, structuration, consolidation 34

2.4. À partir des années 90 : essor du domaine, mais difficulté du rassemblement et de la coordination pour les chercheurs 35

2.5. Unification et stabilisation du domaine autour de la linguistique de corpus 37

3. Extension récente du domaine de la phraséologie *39*

CHAPITRE 2 - LA PHRASEOLOGIE PRAGMATIQUE : VERS UN NOUVEAU PARADIGME 42

1. Le champ de la pragmatique *42*

2. La phraséologie pragmatique au sens large *44*

3. La phraséologie pragmatique des interactions orales *45*

4. Revue des différents paradigmes de phraséologismes pragmatiques apparentés aux PhPex *46*

4.1. Les approches généralistes 46

4.1.1. Locutions exclamatives, Bally (1909) 47

4.1.2. Formules de la conversation, López Simó (2016) 48

4.1.3. Phrases préfabriquées des interactions, Tutin (2019) 49

4.1.4. Phrases préfabriquées, Dostie (2019) 50

4.2. Les approches lexicographiques 51

4.2.1. Structures figées de la conversation, Bidaud (2002) 52

4.2.2. Actes de langage stéréotypés, Kauffer (2011) 53

4.2.3. Formules expressives de la conversation, Gharbi (2020) et Krzyżanowska, Grossmann & Kwapisz-Osadnik (2021) 55

4.3. Les approches interactionnistes 58

4.3.1. Routines conversationnelles, Coulmas (1981) 58

4.3.2. Routine d'interaction, Traverso (1996) 60

4.4. Les approches contextuelles 62

4.4.1. Énoncés liés, Fónagy (1982) et Marque-Pucheu (2007) 63

4.4.2. Énoncés situationnels (*situation bound utterances*), Kecskés (2000) 64

4.4.3. Phrases figées illocutoires, Náray-Szabó (2009) 66

CHAPITRE 3 - LES PHRASEOLOGISMES PRAGMATIQUES A FONCTION EXPRESSIVE : FORME, SENS, USAGE 70

1. Choix de dénomination *70*

2. Définitions des spécificités des PhPex *74*

2.1. Spécificités morphosyntaxiques 74

2.2. Spécificités sémantiques 75

2.3. Spécificités pragmatiques 77

2.4. Spécificités interactionnelles 79

3. Les PhPex observés dans cette étude *82*

DEUXIEME VOLET DU CADRE THEORIQUE - DES PHRASEOLOGISMES PRAGMATIQUES A FONCTION EXPRESSIVE A L'ACQUISITION DE LA COMPETENCE LEXICOPRAGMATIQUE 85

CHAPITRE 4 - CONTEXTUALISATION DU CHAMP DE LA PRAGMATIQUE DE LA LANGUE SECONDE 87
1. Tour d'horizon historique et théorique 87
1.1. L'émergence de l'étude de la compétence pragmatique 87
1.2. Les années 80 à 90 : émergence des études comparatives et translinguistiques 88
1.3. À partir des années 90 : études longitudinales et généralisation des mesures statistiques 89
1.4. Des années 2000 à aujourd'hui : apport des théories de la SLA à la SLP 90
2. Le champ de la pragmatique de la langue seconde 92
2.1. Définition 92
2.2. Quels sont les objectifs de ce champ disciplinaire ? 93
CHAPITRE 5 - UNE APPROCHE INTERCULTURELLE, INTERLINGUISTIQUE ET CONTRASTIVE 95
1. La pragmatique transculturelle (Cross-Cultural pragmatics) 95
2. La pragmatique interlangagière (Interlanguage Pragmatics) 97
3. Analyse Contrastive de l'interlangue (Contrastive Interlanguage Analysis) 98
4. À propos des travaux sur l'acquisition des phraséologismes pragmatique de la conversation .. 100
CHAPITRE 6 - DE LA COMPETENCE COMMUNICATIVE A LA COMPETENCE LEXICOPRAGMATIQUE 103
1. Évolution de la définition de la compétence pragmatique 103
2. Vers une nouvelle approche de la compétence pragmatique ? 112
3. La compétence lexicopragmatique : définition et modélisation 117
3.1. Le niveau pragmalinguistique 120
3.2. Le niveau sociopragmatique 121
3.3. Le niveau métapragmatique 123
CHAPITRE 7 - METHODE D'ELICITATION ET D'EVALUATION DE LA PRODUCTION PRAGMATIQUE EN L2 127
1. Les protocoles non-interactifs 127
1.1. Activité de complétion du dialogue (*Discourse completion task* ou DCT) 128
1.2. Journal de suivi (ou autobiographie d'apprenant) 130
1.3. Questionnaires à choix multiples (QCM) 131
1.4. Questionnaires à échelle (*scaled-response questionnaires*) 134
1.5. Verbalisation de la pensée à voix haute (*Think Aloud Protocol*) 135
2. Les protocoles interactifs 136
2.1. Jeux de rôle 137
2.2. Enregistrements non-élicités / semi-élicités 139
2.3. Entretien métapragmatique 141
SYNTHESE DU CADRE THEORIQUE 145

PARTIE 2 - PROTOCOLE EXPÉRIMENTAL ET ANALYSE 149

AVANT-PROPOS 151

PREMIERE PHASE DU PROTOCOLE EXPERIMENTAL : 153

CHAPITRE 8 - CADRAGE EXPERIMENTAL DE LA PHASE DIAGNOSTIQUE 155
1. Questionnement de départ et description de l'étude 155
2. Le diagnostic de la compétence lexicopragmatique et la notion d'erreur. 157
3. Méthodologie de l'étude diagnostique 159
3.1. Le recueil des données 159
3.1.1. Le corps Exolingue-Grenoble 160
3.1.2. Le corpus Exolingue-Tsukuba 161
3.1.3. Le corpus Endolingue-Tsukuba 162
3.1.4. Matériel de recueil 163
3.1.5. Cadrage juridique des données et anonymisation 164
3.2. Transcription et annotation du corpus 164
3.2.1. La transcription des extraits 164
3.2.2. L'annotation des échantillons 166
3.3. L'accord interannotateur et l'amélioration de l'annotation 172
3.3.1. L'accord interannotateur 172
3.3.2. Résultats de l'AIA pratiquée sur le schéma d'annotation 172
3.3.3. Amélioration du schéma d'annotation 174
3.3.4. Implémentation du schéma d'annotation et utilisation sur ELAN 181
CHAPITRE 9 - ANALYSE DES RESULTATS DU DIAGNOSTIC 185

1. Synthèse des résultats quantitatifs par catégorie d'annotation 185
1.1. Marques expressives 186
1.2. Marques discursive 187
1.3. Marques de l'oralité 189
1.4. Marques d'anomalies lexicopragmatiques et formelles 191
1.5. Synthèse 192
2. Analyse qualitative des anomalies pragmatiques et interactionnelles 194
2.1. Un usage du rire propre à la variété non-native 194
2.1.1. La fonction de convivialité 196
2.1.2. La fonction de palliation 198
2.1.3. La fonction de temporisation 200
2.2. Un suremploi des formes phatiques 202
2.3. Un sous-emploi des marques discursives et interactionnelles 207
2.4. Particularités structurelles, lexicales et pragmatiques de la variété non-native observée 212
3. Diagnostic et conclusion 218
3.1. Synthèse diagnostique 218
3.2. Conclusion de la première phase du protocole expérimental 222

DEUXIEME PHASE DU PROTOCOLE EXPERIMENTAL : 225

POINT ETAPE 1 DU PROTOCOLE EXPERIMENTAL 227
CHAPITRE 10 – CADRAGE EXPERIMENTAL DU TEST LEXICOPRAGMATIQUE 231
1. Les informateurs 231
2. Le test lexicopragmatique 233
2.1. Étapes d'élaboration du test 233
2.2. Structure et spécificités du test 235
3. Traitement des données 241
3.1. Formalisation des données 241
3.2. Validation des résultats du test 242
3.2.1. Procédure d'accord interjuge des réponses en français 242
3.2.2. Procédure d'accord interjuge des réponses en japonais 245
3.2.3. Pourquoi évaluer les résultats des LN ? 249
3.3. Mesure des spécificités lexicales des réponses des LN et des LNN 252
3.3.1. Variabilité lexicale 252
3.3.2. Variabilité du registre 253
3.3.3. Indice Phraséologique 261
3.3.4. Fréquence des formes employées dans le test 264
3.3.5. Taux de compréhension 264
CHAPITRE 11 - ANALYSE QUANTITATIVE ET STATISTIQUE DES DONNEES 269
1. Étude quantitative des résultats obtenus au test lexicopragmatique 270
1.1. Analyse par activité 272
1.1.1. Activité 1 (complétion) 272
1.1.2. Activité 2 (reformulation) 273
1.1.3. Activité 3 (équivalence) 274
1.1.4. Activité 4 (QCM lexical) 275
1.1.5. Activité 5 (QCM sémantique) 275
2. Étude statistique des variables liées à l'apprentissage et les variables liées à la pratique du français 278
2.1. Variables liées aux conditions d'apprentissage 285
2.1.1. Variable niveau d'étude 285
2.1.2. Variable de la localisation de l'université 286
2.1.3. Variable de la spécialité d'étude 286
2.1.4. Variable du niveau de certification 286
2.1.5. Variable de la durée d'apprentissage 287
2.1.6. Variable du cadre d'apprentissage 287
2.2. Variables liées à l'usage 288
2.2.1. Variable du statut socioprofessionnel 288
2.2.2. Variable du lieu de résidence 289
2.2.3. Variable de la durée du séjour 290
2.2.4. Variable du contexte principal de conversation 290
2.2.5. Variable de la fréquence d'interaction 291
2.2.6. Variable du type d'interlocuteur 291
3. Synthèse de l'analyse quantitative et statistique des données 292

4. Discussion des résultats quantitatifs et statistiques *295*
4.1. Effet de la longueur d'exposition sur la compétence lexicopragmatique 295
4.2. Effet de l'expertise en langue cible sur la compétence lexicopragmatique 297
4.3. Effet de l'intensité d'exposition à la langue cible sur la compétence lexicopragmatique 298
4.4. Effet de la diversité d'exposition sur la compétence lexicopragmatique 299
5. Remarques finales *300*
CHAPITRE 12 - ANALYSE QUALITATIVE ET COMPARATIVE DES DONNEES 305
1. Analyse linguistique de l'activité 1 (complétion) *306*
1.1. Cadre de l'activité 1 306
1.2. Spécificités lexicales des réponses observées dans l'activité 1 307
1.3. Analyse des réponses de l'activité 1 309
1.3.1. Analyse comparative 309
1.3.2. Analyse des difficultés 312
1.4. Synthèse intermédiaire de l'activité 1 313
2. Analyse linguistique de l'activité 2 (reformulation) *314*
2.1. Cadre de l'activité 2 314
2.2. Spécificités lexicales des réponses observées dans l'activité 2 315
2.3. Analyse des réponses de l'activité 2 318
2.3.1. Analyse comparative 318
2.3.2. Analyse des difficultés 323
2.4. Synthèse intermédiaire de l'activité 2 326
3. Analyse linguistique de l'activité 3 (équivalence fonctionnelle) *327*
3.1. Cadre de l'activité 3 327
3.2. Spécificités lexicales des réponses observées dans l'activité 3 328
3.3. Analyse des réponses de l'activité 3 331
3.3.1. Analyse comparative 331
3.3.2. Analyse des difficultés observées dans l'activité 3 340
3.4. Conclusion intermédiaire de l'activité 3 343
4. Analyse linguistique de l'activité 4 (QCM lexical) *345*
4.1. Cadre de l'activité 4 345
4.2. Analyse des réponses de l'activité 4 346
4.2.1. Dialogue 1 346
4.2.2. Dialogue 2 348
4.2.3. Dialogue 3 350
4.3. Synthèse intermédiaire de l'activité 4 352
5. Analyse linguistique de l'activité 5 (QCM sémantique) *355*
5.1. Cadre de l'activité 5 355
5.2. Analyse des réponses de l'activité 5 355
5.2.1. Dialogue 1 355
5.2.2. Dialogue 2 358
5.3. Synthèse intermédiaire activité 5 362
6. Synthèse des résultats du test lexicopragmatique *362*
6.1. Typologie des difficultés des LNN observées dans le test lexicopragmatique 363
6.1.1. En production 364
6.1.2. En compréhension 368
6.2. Autres difficultés 370

TROISIEME PHASE DU PROTOCOLE EXPERIMENTAL : **375**

POINT ETAPE 2 DU PROTOCOLE EXPERIMENTAL 377
CHAPITRE 13 - METHODOLOGIE DE RECUEIL ET D'ANALYSE DES DONNEES DES ENTRETIENS SEMI-DIRECTIFS 379
1. Questionnement de départ et objectif *379*
2. Principe et structure de l'entretien d'élicitation métapragmatique semi-directif *381*
2.1. Spécificités de notre approche 381
2.2. Déroulement de l'entretien 383
2.3. Les données recueillies 387
3. Méthodologie d'analyse *389*
3.1. Dépouillage et formalisation des données 389
3.2. Codage du relevé effectué sur les entretiens 390
CHAPITRE 14 - SYNTHESE DES DISCUSSIONS METAPRAGMATIQUES 393
1. Les difficultés de discernement de sens phraséologique et littéral *393*

2. Difficulté liée à des analogies erronées avec la langue maternelle *394*
3. Difficultés liées au repérage de la fonction pragmatique *396*
4. Difficulté d'accommodation du registre au contexte de communication *396*
5. Difficultés liées aux objectifs de communication *398*
6. Difficulté en production liée à l'asymétrie entre compétence de compréhension et diversité du répertoire lexicopragmatique *399*
7. L'influence de l'éthos communicatif sur la production en LC *400*

CONCLUSION **403**

1. Synthèse des résultats *403*
2. Limites et améliorations possibles de notre approche *407*
3. Mot de la fin et ouverture *408*

BIBLIOGRAPHIE **411**

SIGLES ET ABREVIATIONS UTILISES **442**

TABLE DES ILLUSTRATIONS **444**

LISTES DES TABLEAUX **448**

ANNEXES **450**

TABLE DES ANNEXES **451**

TABLE DES MATIERES **482**