

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 26 mars 2021 par

Rémi Lamarque

Facteurs influant sur la productivité dans des processus morphophonologiques récents et bien établis

Discipline

Sciences du Langage

École doctorale

ED 356 – Cognition, Langage, Éducation

Laboratoire

Laboratoire Parole et Langage

• **Composition du jury**

- Laurence LABRUNE Rapporteuse
- Laboratoire Cognition, Langues, Langage, Ergonomie – Équipe de
- Recherche en Syntaxe et en Sémantique à Bordeaux /
- Université Bordeaux Montaigne / CNRS /
- Institut National des Langues et Civilisations Orientales
- Andrew WEDEL Rapporteur
- Department of Linguistics / University of Arizona
- Leonardo LANCIA Examineur
- Laboratoire de Phonétique et Phonologie /
- Université Sorbonne Nouvelle / CNRS
- Philippe BLACHE Président du jury
- Laboratoire Parole et Langage / Aix-Marseille Université / CNRS
- James SNEED GERMAN Directeur de thèse
- Laboratoire Parole et Langage / Aix-Marseille Université

Affidavit

Je soussigné, Rémi Lamarque, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de James Sneed German, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Aix-en-Provence, le 04/01/2021.



Cette œuvre est mise à disposition selon les termes de la [Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Résumé

En morphologie, la fréquence de types des différents patterns dans les formes existantes a une influence considérable sur leur productivité. Cependant, ce facteur s'avère parfois insuffisant pour expliquer la productivité de certains patterns. On peut alors s'interroger sur les autres facteurs jouant un rôle dans la construction des représentations qui guident la productivité morphologique. Cette thèse aborde cette question en explorant des situations linguistiques dans lesquelles l'influence des fréquences de types est limitée par un faible nombre de formes existantes et/ou parce qu'elles entrent en conflit avec les nouveaux inputs que reçoivent les locuteurs lors d'une interaction.

Nos résultats montrent que les fréquences de types jouent effectivement un rôle majeur dans la productivité. Néanmoins, il arrive que les locuteurs ne parviennent pas à abstraire un pattern lorsque le nombre de formes existantes est trop faible. Dans cette situation, certains patterns sont sous- ou sur-représentés dans les nouvelles formes, comparé à leur fréquence dans les formes existantes. L'existence de préférences qui ne sont pas motivées par les fréquences de types suggère que les locuteurs ont des biais inhérents pour certains patterns.

De plus, les locuteurs sont capables de rapidement converger vers un partenaire d'interaction, même lorsque cela va à l'encontre des tendances observées dans les formes existantes. Ainsi, les locuteurs peuvent graduellement modifier leurs préférences en raison de l'effet cumulé de la convergence morphologique lors de multiples interactions. Par conséquent, la convergence pourrait jouer un rôle important dans l'évolution des patterns morphologiques.

Mots-clés : morphologie, phonologie, productivité, fréquence de types, convergence, japonais

Abstract

In morphology, the type frequencies of the different patterns in existing forms have a considerable influence in the way speakers create new forms, i.e., on the productivity of those patterns. However, type frequencies are sometimes insufficient to fully account for the productivity, or lack thereof, of some morphological patterns. This raises the question of which other factors may play a role in the abstraction and updating of the representations that guide morphological productivity. This thesis addresses this question by exploring linguistic situations in which the influence of the type frequencies is limited by a low number of existing forms and/or may conflict with the new input speakers encounter in an interaction.

Our results show that the type frequencies indeed play a major role in the way speakers create novel forms. However, speakers sometimes fail to abstract a pattern when the number of existing forms is too low. In this situation, certain patterns are over- or under-represented in novel forms as compared to their type frequencies among existing forms. The presence of preferences which are unmotivated by type frequencies suggests that speakers have inherent biases towards certain patterns.

Besides, speakers are able to quickly converge with an interaction partner, even when this goes against the tendencies observed in existing forms. Thus, it follows that speakers can gradually modify their preferences through the accumulated effect of morphological convergence in multiple interactions. Consequently, these findings suggest an important role for convergence in the emergence and evolution of morphological patterns.

Keywords: morphology, phonology, productivity, type frequencies, convergence, Japanese

Remerciements

J'adresse mes remerciements à toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce travail.

En premier lieu je souhaite remercier mon directeur de thèse, James Sneed German, pour ses qualités professionnelles et humaines. Sa supervision m'a apporté aussi bien la sérénité que la motivation et la stimulation intellectuelle indispensables au bon déroulement d'un doctorat. Les enseignements que j'ai retiré de chacun de nos échanges m'ont enrichi, bien au-delà du cadre de cette thèse, sur le plan intellectuel et personnel.

Mes sincères remerciements vont également aux membres de mon jury pour l'honneur qu'ils m'ont accordé en évaluant ce travail.

Merci à tous les membres du Laboratoire Parole et Langage, qui contribuent à faire de ce lieu un cadre idéal pour entreprendre un doctorat. Échanger avec eux, de manière professionnelle ou informelle, m'a beaucoup apporté intellectuellement et humainement.

Je suis particulièrement reconnaissant envers Cristel Portes qui a su, en l'espace d'une conversation, me passionner pour les sciences du langage à l'époque où je cherchais encore ma voie dans les couloirs de l'université.

Merci aux doctorant·e·s et jeunes docteur·e·s du LPL. S'il devait y avoir un inconvénient à avoir achevé cette thèse, c'est de devoir laisser ma place dans notre open space. Je ne peux malheureusement pas faire la liste de toutes les belles personnes avec qui j'ai eu la chance de partager cet espace. Je souhaite néanmoins exprimer à chacune d'entre elles toute ma reconnaissance et mon affection pour les bons moments passés en leur compagnie.

Enfin, je souhaite exprimer ma gratitude à toutes les personnes de mon entourage dont la simple présence est un soutien précieux au quotidien.

Merci à tous mes Aix'Tra copains du chœur Aix'Tralala qui m'ont accompagné tout au long de ce chemin en y semant des notes de musique remplies d'émotions positives.

Merci à mes amis. Alexis et Margot sur qui j'ai toujours pu compter depuis de nombreuses années. Et Lola pour être, entre autres, une source intarissable de rire et de bienveillance.

Et merci à ma famille, ma mère en particulier, pour me motiver à toujours donner le meilleur de moi.

Table des matières

Affidavit	I
Résumé	II
Abstract	III
Remerciements	IV
Liste des Figures	VIII
Liste des tableaux	X
Chapitre 1 – Introduction	1
1.1 Questions de recherche	1
1.2 Plan de la thèse	4
1.3 Cadre théorique	6
1.3.1 Fréquences d'occurrence et productivité des patterns morphophonologiques...	6
1.3.2 Limites de l'influence des fréquences d'occurrence	8
1.3.3 Rôle de la convergence morphologique dans l'évolution des représentations	14
Chapitre 2 – Les Compound Abbreviated Loanwords (CAL) japonais	20
2.1 Le rôle de la more en japonais	20
2.2 Le principe général de formation des CAL et ses exceptions	24
2.3 Emergence de patterns dans les CAL	31
Chapitre 3 – Étude du lexique de CAL japonais	34
3.1 Méthode	34
3.2 Résultats	35
3.3 Discussion	41
3.3.1 Remarques générales	41
3.3.2 Description en OT standard	42
3.3.3 Description en OT stochastique	44
3.3.4 Prédications concernant la productivité des patterns en fonction des fréquences de types	47
Chapitre 4 – Expérience : tâche de création de nouveaux CAL	51
4.1 Méthode	51
4.1.1 Matériel	51
4.1.2 Procédure	54
4.1.3 Participants	56
4.2 Résultats	56
4.3 Discussion	59

4.3.1	Comparaison entre les formes existantes et les nouvelles formes créées.....	59
4.3.2	Reproductibilité des résultats par un modèle d'apprentissage <i>constraint-based</i>	63
4.3.3	Validité des prédictions formulées sur la base des fréquences de types.....	69
4.3.4	Conclusion.....	77
Chapitre 5 – Expérience ESP - CAL japonais : convergence morphophonologique dans un processus morphologique récent		78
5.1	Méthode	79
5.1.1	Matériel.....	79
5.1.2	Procédure.....	81
5.1.3	Partenaire artificiel.....	85
5.1.4	Participants.....	92
5.2	Résultats	93
5.3	Discussion	100
Chapitre 6 – Les flexions verbales japonaises		107
6.1	Les verbes japonais et leurs flexions	108
6.2	Fréquences de types dans les formes existantes	110
6.3	Productivité des différents patterns dans les formes existantes	112
6.4	Productivité des patterns dans la création de nouvelles formes	114
Chapitre 7 – Expérience baseline - flexions verbales : flexion de nouvelles formes verbales japonaises		122
7.1	Méthode	123
7.1.1	Matériel.....	123
7.1.2	Procédure.....	129
7.1.3	Participants.....	130
7.2	Résultats	131
7.3	Discussion	133
Chapitre 8 – Expérience ESP - Flexions verbales : convergence morphophonologique dans un processus morphologique non récent.....		137
8.1	Méthode	137
8.1.1	Matériel.....	137
8.1.2	Procédure.....	139
8.1.3	Partenaire artificiel.....	140
8.1.4	Participants.....	142
8.2	Résultats	142
8.3	Discussion	145
Chapitre 9 – Discussion générale.....		147

9.1	Réponses aux questions de recherche.....	147
9.2	Modèle bayésien de l'apprentissage de la morphologie	149
9.3	Perspectives et conclusion	162
Références :		167
Annexes		177
	Annexe 1 : instructions de l'expérience baseline-CAL	177
	Annexe 2 : instructions de l'expérience ESP-CAL	178
	Annexe 3 : instructions de l'expérience baseline-flexions verbales.....	182
	Annexe 4 : instructions de l'expérience ESP-CAL	183
	Annexe 5 : liste des premiers constituants utilisés pour les items des expériences sur les CAL	187
	Annexe 6 : liste n°1 de deuxièmes constituants utilisés pour les items de l'expérience baseline-CAL.....	192
	Annexe 7 : liste n°2 de deuxièmes constituants utilisés pour les items de l'expérience baseline-CAL.....	198
	Annexe 8 : liste de deuxièmes constituants utilisés pour les items de l'expérience ESP-CAL.....	204
	Annexe 9 : liste de pseudo-verbès utilisés pour les items des expériences sur les flexions verbales	211
	Annexe 10 : liste des dialogues utilisés pour les items des expériences sur les flexions verbales.	217

Liste des Figures

Fig.1 : structure syllabique de syllabes lourdes japonaises incluant les segments spéciaux /R/, /Q/, et /N/.....	24
Fig.2 : dérivation de <i>pokemon</i> à partir de <i>poketto</i> et <i>monsutaa</i>	26
Fig.3 : proportion d'utilisation (en %) de chaque pattern dans le lexique, en fonction du type de deuxième more du C2.	36
Fig.4 : traitement de la deuxième more du second constituant en fonction du nombre de more du constituant.	38
Fig.5 : traitement de la deuxième more du second constituant en fonction de la nature de la consonne initiale du constituant.	39
Fig.6 : traitement de la more /R/ en tant que deuxième more du second constituant en fonction de la hauteur de la voyelle allongée.	40
Fig.7 : position des contraintes impliquées dans la formation des CAL sur une échelle de <i>strictness</i> , selon les données du lexique.	46
Fig.8 : affichage d'un item dans la tâche de création de nouvelles formes.....	55
Fig.9 : proportion d'utilisation (en %) de chaque pattern dans l'expérience en fonction de la deuxième more du C2.	57
Fig.10 : position des contraintes impliquées dans la formation des CAL sur une échelle de <i>strictness</i> , selon les données de la tâche de création de nouvelles formes.....	62
Fig.11 : état initial de la grammaire présentée au modèle GLA.....	65
Fig.12 : état de la grammaire après un apprentissage de 10 répétitions avec un bruit de 2.	67
Fig.13 : exemple d'un item tel que présenté en phase 1 de l'expérience ESP.	82
Fig.14 : exemple d'un item tel que présenté en phase 2 de l'expérience ESP, pendant le moment de réflexion du robot.	84
Fig.15 : évolution de la proportion (en %) de formes régulières au cours des phases de jeu, en fonction du comportement du robot, pour les items <i>sans conflit</i>	95
Fig.16 : évolution de la proportion (en %) de formes régulières au cours des phases de jeu, en fonction du comportement du robot, pour les items /R/.....	97
Fig.17 : évolution de la proportion (en %) de formes régulières au cours des phases de jeu, en fonction du comportement du robot, pour les items /Q/.	98

Fig.18 : proportions d'utilisation (en %) des différents patterns dans les phases 1 et 3, en fonction du comportement du robot.	99
Fig.19 : fréquence d'apparition des voyelles fournies au générateur de pseudo-verbos.....	125
Fig.20 : fréquences d'apparition des consonnes fournies au générateur de pseudo-verbos. Le symbole « ' » représente une attaque vide.	125
Fig.21 : exemple d'un item tel que présenté dans l'expérience baseline-flexions verbales...	130
Fig.22 : influence de la forme demandée et de la longueur des verbos sur la proportion de formes <i>yodan</i> produites.	132
Fig.23 : exemple d'un item tel que présenté dans la phase 1 de l'expérience ESP-flexions verbales.....	140
Fig.24 : proportions (en %) de formes <i>yodan</i> produites en fonction de la phase et du comportement du robot.	143
Fig.25 : apprentissage bayésien avec un a priori plat ($f(\theta) = 1$) et des données définies par une loi normale ($\mu=0.9 ; \sigma=0.3$).	152
Fig.26 : apprentissage bayésien avec un a priori défini par une loi normale ($\mu=0.5 ; \sigma=0.3$) et des données définies par une loi normale ($\mu=0.9 ; \sigma=0.3$).	153
Fig.27 : apprentissage bayésien avec un a priori défini par une loi normale ($\mu=0 ; \sigma=0.01$) et des données définies par une loi normale ($\mu=0.75 ; \sigma=0.3$).....	155
Fig.28 : apprentissage bayésien avec un a priori défini par une loi normale ($\mu=0.35 ; \sigma=0.3$) et des données définies par une loi normale ($\mu=0.5 ; \sigma=0.3$).	156
Fig.29 : apprentissage bayésien avec un a priori défini par une loi normale ($\mu=1 ; \sigma=0.05$) et des données définies par une loi normale ($\mu=0.4 ; \sigma=0.07$).....	157
Fig.30 : apprentissage bayésien avec un a priori défini par une loi normale ($\mu=0.8 ; \sigma=0.3$) et des données définies par une loi normale ($\mu=0.55 ; \sigma=0.3$).	160

Liste des tableaux

Tableau 1 : *V:# >> *PIED<2 ; exemple ‘family mart’	42
Tableau 2 : CONTIG >> *V:# ; exemple ‘cheese burger’	42
Tableau 3 : *PIED>2 >> *V:# ; exemple ‘cheese burger’	43
Tableau 4 : LICENCE/Q/ >> CONTIG ; exemple ‘Harry Potter’	43
Tableau 5 : LICENCE/Q/ >> *PIED<2 ; exemple ‘plastic pack’	43
Tableau 6 : exemple d’un CAL irrégulier (‘cheese burger’) pour lequel la hiérarchie de contraintes sélectionnerait un output incorrect.....	45
Tableau 7: sélection de l’output pour ‘cheese burger’ dans le cas où la hiérarchie est ponctuellement : LICENCE/Q/ >> CONTIG >> *PIED>2 >> *PIED<2 >> *V:#.....	46
Tableau 8 : valeurs z et valeurs p du modèle logistique multinomial.	59
Tableau 9 : production de 40 outputs pour chaque type d’input par le modèle GLA avec un bruit de 2.	68
Tableau 10 : proportion d’utilisation (en %) de chaque pattern pour les C2 n° 1, 2, et 3 de la liste 1 dans l’expérience baseline.	80
Tableau 11 : exemple de distribution des formes produites pour chacun des comportements du robot avec un participant hypothétique.	89
Tableau 12 : ratios z et valeurs p des comparaisons post hoc par paires avec correction de Tukey pour les items <i>sans conflit</i>	96
Tableau 13 : ratios z et valeurs p des comparaisons post hoc par paires avec correction de Tukey pour les items /R/.	98
Tableau 14 : les différentes flexions verbales en japonais.	109
Tableau 15 : fréquences de types des différents paradigmes verbaux dans les formes existantes répertoriées dans le dictionnaire JMdict.....	111
Tableau 16 : les flexions verbales du japonais entraînant la création d’un radical <i>ichidan</i> se finissant par <i>e</i>	128
Tableau 17 : ratios z et valeurs p des comparaisons post hoc par paires avec correction de Tukey pour les données de l’expérience ESP – flexions verbales.	144

Chapitre 1 – Introduction

1.1 Questions de recherche

Dans les langues naturelles, il est rare de trouver des processus morphologiques parfaitement réguliers et ne présentant aucune variabilité. Lorsqu'ils créent de nouvelles formes, les locuteurs ont dans la plupart des cas une certaine liberté dans le choix de différentes formes en compétition (Fehringer, 2004 ; Haber, 1976 ; Säily, 2011 ; Thornton, 2012). Les fréquences de types dans les formes existantes des différentes options disponibles pour les locuteurs a une influence considérable sur la manière dont les locuteurs créent de nouvelles formes (Albright & Hayes, 2003 ; Bybee, 1995, 2001 ; Pierrehumbert, 2001b, 2001c). Cependant, quelques études montrent que les fréquences de types sont parfois insuffisantes pour rendre compte de la productivité, ou de l'improductivité, de certains patterns¹ morphologiques (Becker, Ketrez, & Nevins, 2011 ; Dawdy-Hesterberg, 2014 ; Kapatsinski, 2010). Cela soulève une série de questions sur les différents facteurs qui entrent en jeu dans l'abstraction et l'évolution des représentations qui guident la productivité morphologique. Nous nous intéresserons en particulier à deux questions de recherche dont les réponses nous permettront de comprendre la nature de ces représentations :

- 1) Les locuteurs reproduisent-ils simplement les probabilités de l'input dans leur output (c.-à-d. ont-ils un comportement de type « probability-matching ») ?
- 2) Sinon, quels autres facteurs entrent en jeu ?

Dans cette thèse, nous abordons ces questions en explorant des situations linguistiques dans lesquelles l'influence des fréquences de types est soit limitée par un nombre peu élevé de formes

¹ Dans cette thèse, nous utiliserons « pattern » en tant que terme théoriquement neutre renvoyant à toute forme de généralisation abstraite, non entièrement analogique, régissant la création de nouvelles formes. Cela peut inclure aussi bien des généralisations source-oriented et product-oriented. D'une part les généralisations source-oriented, telles que les règles utilisées dans l'approche dérivationnelle, partent des inputs et définissent quelles catégories d'inputs doivent aboutir à quelles catégories d'outputs. D'autres part les généralisations product-oriented, telles que les hiérarchies de contraintes utilisées dans les approches basées sur les contraintes, sont orientées sur les caractéristiques qui doivent être évitées ou favorisées dans les outputs de manière générale.

existantes, soit en conflit avec un nouvel input auquel les locuteurs sont exposés lors d'une interaction.

Répondre à la première question de manière expérimentale s'avère une tâche assez aisée théoriquement. Dans la mesure où l'on peut connaître les probabilités présentes dans l'input auquel sont exposés les locuteurs, il suffit alors d'éliciter dans un cadre expérimental la création de nouvelles formes et de comparer les probabilités dans l'input et celles dans les nouvelles formes créées. La question 1) peut donc être abordée de cette manière. Notons toutefois que s'il est possible d'avoir une estimation assez juste des probabilités dans l'input commun à l'ensemble des locuteurs pour des processus morphologiques bien établis dans la langue et auxquels les locuteurs sont exposés de manière très régulière, les choses s'avèrent plus complexes pour des processus devenus productifs récemment et dont l'emploi est très variable d'un locuteur à l'autre. Ce premier constat nous invite d'ores et déjà à étudier et comparer des processus morphologiques de nature différente au moins en termes d'importance de l'input disponible pour les locuteurs, ce qui nous emmène vers la question 2).

L'idée d'un comportement de type *probability-matching*, dans lequel les locuteurs reproduiraient exactement les fréquences de types d'un input dans leur output, semble a priori incompatible avec l'idée de l'émergence d'un nouveau processus morphologique ou encore de l'évolution d'un processus existant. Si d'autres facteurs jouent un rôle dans ces phénomènes, on peut s'attendre à ce que leur effet soit d'autant plus apparent dans des contextes où l'input auquel sont exposés les locuteurs est limité. En effet, si l'on considère le cas d'un processus morphologique dans les premiers temps de son apparition dans la langue, il n'existe pratiquement aucune forme, donc aucune fréquence de types, sur laquelle les locuteurs peuvent se baser pour créer de nouvelles formes. Dans une telle situation, on peut s'interroger sur les facteurs qui influent sur le comportement des locuteurs, si ce ne sont les fréquences de types.

Notons que d'après une description *bayésienne* du processus d'apprentissage de la langue, il serait attendu que les locuteurs abordent cette situation avec un ensemble de biais a priori guidant leur comportement. Ces biais a priori étant légèrement réajustés avec chaque nouvel input auquel les locuteurs sont exposés, l'ampleur de leur rôle serait d'autant plus important que cet input est réduit (Pierrehumbert, à paraître). Ce fonctionnement n'exclut pas pour autant que les adultes d'une même communauté, possédant un système mature et nourri d'un input très important et assez similaire d'une personne à l'autre, aient des comportements

similaires et réguliers. L'étude de processus morphologiques devenus récemment productifs avec peu d'input disponible pour les locuteurs, en comparaison avec des processus morphologiques mieux établis disposant d'un input important, permettrait donc d'explorer l'effet des biais a priori des locuteurs et de déterminer si l'apprentissage et la productivité de patterns morphologiques se basent sur un comportement de probability-matching ou davantage bayésien.

Nous avons donc choisi comme cas d'étude pour nos travaux deux processus morphologiques du japonais : le premier est un processus d'abréviation de mots d'emprunts composés devenu productif assez récemment dans la langue puisqu'il utilise comme base des mots empruntés récemment aux langues occidentales (essentiellement l'anglais) et qui sont entrés massivement dans la langue après la Seconde Guerre mondiale (Kay, 1995) ; le second est un processus de flexion verbale utilisé au quotidien dans la langue et auquel les locuteurs sont donc fortement exposés.

Par ailleurs, si les biais a priori des locuteurs peuvent jouer un rôle dans leur comportement à un niveau individuel, et que ces biais sont ajustés avec chaque nouvel input, on peut également s'interroger sur l'effet des interactions entre locuteurs sur l'évolution de leurs biais respectifs. Notamment, une série d'études sur des modèles de populations d'agents bayésiens, ou sur l'apprentissage de langues artificielles par des locuteurs en milieu expérimental, ont montré que des phénomènes de niveau communautaire, de l'ordre de la transmission culturelle, pouvaient fortement influencer sur l'amplification et la propagation de ces biais dans la communauté. Dans un contexte d'apprentissage itératif, où chaque locuteur apprend la langue de locuteurs qui l'ont eux-mêmes appris d'autres locuteurs (etc.), des biais légèrement majoritaires chez les individus, voire non majoritaires, peuvent émerger comme règles au niveau communautaire en quelques générations (Beckner, Pierrehumbert, & Hay, 2017 ; Griffiths & Kalish, 2007 ; Kirby, 2001 ; Kirby, Cornish, & Smith, 2008 ; Kirby, Dowman, & Griffiths, 2007 ; Kirby & Hurford, 2002 ; Smith, 2009 ; Smith & Wonnacott, 2010). Dans nos travaux, nous avons donc souhaité explorer également l'effet de l'interaction avec un partenaire sur le comportement des locuteurs dans leur emploi de patterns morphologiques issus d'une langue naturelle. En particulier, nous nous sommes intéressés à la convergence morphologique dans une série d'expériences portant sur les deux processus morphologiques sus-cités en plaçant les participants dans une situation d'interaction avec un partenaire artificiel.

En résumé, afin de répondre à nos questions de recherche, nous avons choisi de baser nos études sur deux processus morphologiques (l'un récent, l'autre non) employés dans une langue naturelle (c.-à-d. le japonais). Pour chacun de ces processus morphologiques, la comparaison entre les formes existantes et les nouvelles formes produites par les locuteurs dans un cadre expérimental nous permettra d'évaluer l'étendue mais également les limites de l'influence des fréquences de types des patterns sur les productions des locuteurs. Puis la comparaison d'un processus morphologique n'offrant que peu d'input aux locuteurs à un processus morphologique auquel ils sont fortement exposés quotidiennement nous permettra de mesurer l'importance des biais a priori des locuteurs dans leur comportement. Enfin, la comparaison entre un cadre expérimental dans lequel les participants sont isolés avec un cadre expérimental dans lequel ils sont en interaction avec un partenaire artificiel nous éclairera sur le rôle de la convergence morphologique dans l'évolution des comportements des locuteurs.

1.2 Plan de la thèse

Dans le reste de cette introduction, nous reviendrons tout d'abord, en section 1.3.1, sur l'importance de la fréquence de types des patterns dans leur abstraction et leur productivité. Puis, en section 1.3.2, nous reviendrons sur différentes études illustrant les limites du rôle des fréquences de types sur la productivité. Nous décrirons le fonctionnement d'un modèle bayésien de l'apprentissage et expliquerons de quelle manière cette approche permet d'aller au-delà des approximations du probability-matching, en cela qu'elle permet non seulement de prendre en compte l'importante influence des fréquences de types mais également celle des biais a priori des locuteurs. Enfin, en section 1.3.3, nous verrons comment l'accumulation d'interactions entre les locuteurs peut jouer sur l'évolution de leurs comportements, notamment via l'effet de la convergence morphologique.

Les chapitres 2 à 5 de cette thèse seront consacrés à l'étude du processus de formation des Compound Abbreviated Loanwords (désormais « CAL ») japonais, notre cas d'étude de processus morphologique devenu productif récemment. Le chapitre 2 sera consacré à la description de ce processus morphologique produisant des abréviations à partir de paires de mots d'emprunt. Nous examinerons ensuite les fréquences de types de différents patterns de formation dans les formes existantes via l'étude d'un lexique de CAL (chapitre 3).

Puis, dans le chapitre 4, nous présenterons une expérience dans laquelle il a été demandé à des participants japonais de créer de nouvelles abréviations à partir de paires de mots anglais. Dans cette première expérience les participants réalisaient la tâche seuls, sans aucune forme d'interaction. Cette étude se focalise sur l'influence des biais a priori des participants, notamment dans des situations linguistiques où l'input qui leur est disponible est limité.

Dans le chapitre 5, nous présenterons une autre expérience de création de nouvelles formes dans laquelle, cette fois, les participants doivent interagir avec un partenaire artificiel. Cette étude aborde la question de la convergence morphologique et de l'influence du nombre de formes existantes sur la propension des participants à converger vers leur partenaire.

Dans les chapitres 6 à 8 de cette thèse, nous présenterons les études relatives au processus de flexion verbale en japonais, notre cas d'étude de processus morphologique bien établi dans la langue. Le chapitre 6 est consacré à une description du processus morphologique. Après avoir décrit le fonctionnement des flexions verbales (section 6.1), nous présenterons les fréquences de types des différents patterns que peuvent suivre ces flexions (section 6.2). Puis nous verrons ce que la littérature nous apprend sur la productivité des différents patterns, que ce soit dans les formes existantes (section 6.3) ou dans les nouvelles formes créées dans un cadre expérimental (section 6.4).

Puis, dans le chapitre 7, nous présenterons une expérience dans laquelle des locuteurs japonais ont été amenés à employer les patterns de flexions verbales sur des pseudo-verbos. Cette étude possède deux objectifs. Le premier est de vérifier si les locuteurs ont un comportement plus proche du probability-matching lorsqu'ils sont exposés à un input important. Cela serait en effet attendu aussi bien dans une conception bayésienne de l'apprentissage de la morphologie que dans une approche entièrement basée sur le probability-matching. Le second est de collecter des données sur le traitement de pseudo-verbos en dehors de toute interaction. Ces données pouvant servir de baseline pour de futures expériences impliquant des interactions.

Enfin, dans le chapitre 8, nous présenterons une autre expérience dont la tâche consiste à employer les patterns de flexions verbales à des pseudo-verbos, mais cette fois en incluant une interaction avec un partenaire artificiel. Nous nous intéresserons ici à la convergence dans un processus morphologique pour lequel les locuteurs ont un input considérable.

Le chapitre 9 de cette thèse sera consacré à une discussion générale dans laquelle nous répondrons à nos deux questions de recherche énoncées ci-dessus (section 9.1). Nous y verrons

également plus en détails de quelle manière les résultats obtenus dans nos diverses expériences peuvent s'expliquer par des raisonnements bayésiens (section 9.2). Enfin, nous présenterons les pistes de réflexion à explorer dans de futures recherches pour confirmer ou infirmer les conclusions issues de nos travaux. Nous réfléchissons également aux implications de nos résultats sur notre compréhension des processus d'émergence, d'acquisition et d'évolution des patterns morphologiques (section 9.3).

1.3 Cadre théorique

1.3.1 Fréquences d'occurrence et productivité des patterns morphophonologiques

La nature des généralisations, si toutefois elles existent, qui régissent la création de formes complexes est au cœur d'un long débat dans la littérature. Une grande diversité de phénomènes d'ordre phonologique ou morphologique peuvent être décrits à l'aide de mécanismes purement analogiques, entièrement basés sur des similarités phonologiques entre des formes attestées dans un jeu de données sans aucune forme de généralisation (Prasada & Pinker, 1993 ; Rumelhart & McClelland, 1986 ; Skousen, 1989, 1993). Cependant, il a été montré que des modèles d'apprenant basés uniquement sur des mécanismes analogique (Nosofsky, 1990) affichent de moins bonnes performances que des modèles basés sur des règles dans des tâches de dérivation de pseudo-mots. Concrètement, en permettant à des formes d'être similaires de manières diverses et variées, les modèles analogiques échouent à identifier les caractéristiques structurelles précises qui sont à la fois nécessaires et suffisantes pour déclencher une alternation de forme, et sont ainsi surpassés par les modèles basés sur des règles sur ce point précis (Albright & Hayes, 2003). Ces résultats indiquent qu'une certaine forme d'abstraction est nécessaire afin de modéliser le comportement des locuteurs dans la dérivation de formes complexes. On peut alors s'interroger sur la manière dont les locuteurs peuvent abstraire ces patterns morphologiques pour ensuite les réemployer.

Dans la création de nouvelles formes, lorsqu'une langue autorise une certaine variabilité, les locuteurs se basent principalement sur leurs connaissances des formes existantes dans leurs décisions concernant les formes à produire. On trouve dans la littérature des arguments en faveur du rôle aussi bien des fréquences de types que des fréquences de tokens dans la

productivité d'un pattern, c.-à-d. sa propension à être généralisé et employé sur de nouvelles formes. Pierrehumbert (2001a) avance que s'il existe un effet des fréquences de tokens, comme soutenu par Moreton (1997) entre autres, il est difficile dans les langues naturelles de distinguer leur effet de celui des fréquences de types, puisque les deux y sont fortement corrélés. De plus, il existe des indices de la limite de l'effet des fréquences de tokens en elles-mêmes. Notamment, les patterns qui ne sont représentés que par quelques types ne sont pas généralisés à de nouvelles formes même lorsque ces types ont une très haute fréquence de tokens (Bybee, 2001 ; Pierrehumbert, 2001b). Par ailleurs, bien que les mots avec une haute fréquence de tokens possèdent une représentation lexicale forte, ils ont tendance à être mémorisés comme des unités autonomes et indivisibles, ce qui a un effet négatif sur la productivité de leur pattern de formation (Bybee, 1985, 1995 ; Guillaume, 1927[1973]).

Au contraire, il y a plusieurs arguments dans la littérature soulignant le rôle crucial des fréquences de types en elles-mêmes, puisque les patterns phonologiques et morphologiques les plus productifs dans les langues du monde sont ceux possédant les plus hautes fréquences de types (Albright & Hayes, 2003 ; Bybee, 1995, 2001 ; Pierrehumbert, 2001b, 2001c). En outre, les modèles d'apprentissage appliqués à la morphologie présentés dans Albright & Hayes (2003), qu'ils soient basés sur des règles ou des mécanismes d'analogie, présentent de meilleures performances lorsqu'ils sont entraînés sur des fréquences de types plutôt que des fréquences de tokens.

Notons toutefois que Hay & Baayen (2002) soutiennent l'idée que ce n'est pas la fréquence de types seule mais le passage, qui impliquent à la fois les fréquences de types et de tokens, qui joue un rôle dans la productivité des affixes. Le passage semble en effet crucial en ce qui concerne la productivité des affixes puisqu'accéder aux formes complexes sans passage rend ces morphèmes invisibles, puisqu'ils n'apparaissent jamais de manière isolée. Par conséquent, des types possédant une grande fréquence de tokens ne jouent qu'un rôle mineur dans la productivité d'un pattern car l'accès à ces formes se fait généralement directement dans leur entièreté, c.-à-d. sans passage. De l'autre côté, les types avec une très faible fréquence de tokens n'augmentent également que très peu la productivité des affixes. Bien qu'ils activent effectivement la représentation lexicale de l'affixe, cette activation sera aussi rare que l'emploi de ces formes. Par conséquent, les types qui influent le plus sur la productivité d'un affixe sont ceux dont la fréquence de tokens ne se situe à aucun de ces extrêmes, c.-à-d. des formes complexes qui sont assez fréquentes pour activer régulièrement la représentation lexicale de

l'affixe, mais sans être assez fréquentes pour que les locuteurs y accèdent dans leur entièreté, sans passage. Néanmoins, le passage n'est peut-être pas aussi crucial pour des processus morphologiques autre que l'affixation, comme dans notre cas d'étude d'abréviations de mots d'emprunt composés, où chaque constituant de la forme complexe possède sa propre représentation lexicale autonome en tant que forme simple. Par ailleurs, Hay & Baayen (2002) observent qu'une haute fréquence de types augmente la probabilité de trouver des types possédant une fréquence de tokens idéale pour le passage, établissant ainsi une corrélation naturelle entre les fréquences de types, le passage, et par conséquent la productivité.

1.3.2 Limites de l'influence des fréquences d'occurrence

Les différents travaux sur le lien entre fréquence et productivité des patterns suggèrent que l'abstraction de généralisations phonologiques ou morphologiques est en grande partie basée sur les fréquences de types des différents patterns existants dans la langue. Cependant, tous les patterns n'ont pas une fréquence d'utilisation dans la création de nouvelles formes qui corresponde à leur fréquence d'apparition dans les formes existantes.

Une première série d'études montrant un écart entre fréquence et productivité est celle menée par Kapatsinski (2010). Il y présente, entre autres, le cas des suffixes d'extension de radical en russe, *-i* et *-a*, servant à la dérivation de verbes à partir de noms avec consonne finale. Il existe une règle sans exception dans les formes existantes (dans les dictionnaires de russe moderne) imposant que lorsque *-i* est attaché à un radical avec une consonne finale vélaire, cette consonne vélaire se palatalise. La palatalisation ne se produit jamais pour les consonnes non vélares, ni pour les consonnes vélares suivies de *-a*, mais est systématique pour les consonnes vélares suivies de *-i*. Cependant, les radicaux avec une finale vélaire sont très souvent suivis de *-a*, tandis que les radicaux avec une finale non vélaire sont très souvent suivis de *-i*. Par conséquent, ce phénomène de palatalisation est assez rare et il est largement plus fréquent d'observer un radical avec une finale vélaire ne subissant pas la palatalisation, puisqu'il est suivi de *-a* ; ou d'observer *-i* n'entraînant pas la palatalisation puisqu'il n'est pas attaché à un radical avec finale vélaire.

Afin de tester si la règle de palatalisation présenterait une productivité parfaite dans la création de nouvelles formes, Kapatsinski (2010) a lancé des simulations sur différents modèles computationnels d'apprenant basés sur des règles ou sur des contraintes (respectivement le

Minimal Generalization Model [Albright & Hayes, 2003] et le Gradual Learning Algorithm [Boersma, 1997]). Après avoir été entraînés sur un lexique de russe, aucun de ces modèles n'utilisait la palatalisation de manière systématique dans les contextes pertinents lorsqu'il leur est demandé de créer de nouveaux verbes à partir de mots d'emprunt à l'anglais, c.-à-d. lorsqu'ils créent de nouvelles formes. Kapatsinski (2010) soutient donc que la productivité d'une règle morphologique décroît à mesure que l'affixe déclenchant cette règle est attaché à des radicaux qui ne peuvent suivre cette règle. En d'autres termes, dans le cas du russe, la règle de palatalisation n'est pas très productive car le suffixe *-i* qui la déclenche est souvent attaché à des radicaux ne pouvant pas être palatalisés (c.-à-d. des radicaux avec finale non vélaire), tandis qu'il est rarement attaché à des radicaux qui peuvent être palatalisés (c.-à-d. des radicaux avec finale vélaire). La règle de palatalisation, même si elle s'applique dans 100% des cas concernés dans les formes existantes, est éclipsée par la règle plus générale consistant à attacher *-i* à n'importe quelle consonne sans la palataliser.

Kapatsinski (2010) apporte d'autres arguments en faveur de cette idée dans une expérience d'apprentissage d'une grammaire artificielle dans laquelle des locuteurs natifs de l'anglais sans aucune connaissance du russe ont été exposés à une langue artificielle possédant deux suffixes de pluriel : *-i* et *-a*. Comme en russe, *-i* provoque systématiquement la palatalisation des consonnes vélares en position finale de radical, et *-a* ne la provoque jamais. Deux langues artificielles légèrement différentes ont été créées et les participants étaient exposés à une seule de ces deux langues. Le suffixe *-i* était attaché à des radicaux avec finale vélaire aussi souvent dans les deux langues, mais tandis que dans la première langue il était rarement attaché à des radicaux avec finale non vélaire, dans la deuxième langue, il était souvent attaché à ce type de radicaux. Les participants exposés à la seconde langue ont palatalisé les consonnes vélares en position finale suivies de *-i* moins souvent que les participants exposés à la première langue, ce qui soutient l'hypothèse de Kapatsinski.

Étant donné que l'hypothèse de Kapatsinski est profondément ancrée dans le cas d'étude d'un processus d'affixation, il n'est pas nécessairement évident d'en traduire les tenants et les aboutissants pour d'autres processus morphologiques. En particulier, transposer cela aux abréviations de mots d'emprunt composés en japonais, l'un de nos cas d'étude, nécessite quelques ajustements. Dans ce qui suit, nous envisagerons donc la possibilité que des patterns liés à des contextes spécifiques, même lorsqu'ils sont très réguliers dans les formes existantes, peuvent être éclipsés par des patterns plus généraux dans la création de nouvelles formes, et

que la productivité des patterns contextuels et généraux peut être influencée par leur fréquence de types respective que ce soit à l'intérieur et en-dehors du contexte concerné. Nous verrons que cela peut s'avérer particulièrement pertinent dans notre cas d'étude des CAL japonais puisque ce processus morphologique présente des patterns généraux et des patterns spécifiques à certains contextes phonologiques, comme c'est le cas pour les processus d'extension de radical en russe.

Notons que les résultats de Kapatsinski (2010) apportent déjà des éléments de preuve que pour certains processus morphologiques, les locuteurs ne semblent pas suivre un comportement de probability-matching entre l'input qu'ils reçoivent et l'output qu'ils produisent. Dans une récente revue de littérature sur la nature probabiliste de la phonologie, Pierrehumbert (à paraître) soutient que s'il existe de forts arguments en faveur de l'importance des probabilités en phonologie, l'idée du probability-matching en production est une extrapolation. Cette idée trouve ses racines dans les travaux de Labov (1989) dans lesquels l'acquisition de la langue par l'enfant est décrit comme un apprentissage et une reproduction des probabilités dans l'input auquel il est exposé. Pierrehumbert remarque cependant que cette conception de l'acquisition du langage semble quelque peu incohérente avec l'idée de variation et de changement linguistique, en particulier lorsqu'ils se produisent sur plusieurs générations de locuteurs et souligne que les sociolinguistes n'ont jamais véritablement soutenu l'idée que les enfants apprenaient et reproduisaient toutes les probabilités de l'input avec une parfaite précision. Hayes, Zuraw, Siptár, & Londe (2009) décrivent également le probability-matching comme une loi générale qui se limite malgré tout à une approximation et soulignent l'intérêt d'étudier les cas qui dévient de cette loi.

Afin de concilier l'existence de la variation et de changement linguistique avec la nature probabiliste de l'acquisition de la langue, un facteur à prendre en compte serait les différences intrinsèques avec lesquelles les locuteurs abordent chaque situation d'apprentissage. En effet, tous les locuteurs ont un vécu et une expérience de la langue qui leur sont propres et qui peuvent biaiser leur comportement face à une situation linguistique nouvelle. Ces biais sont en constante réévaluation avec chaque exposition à un nouvel input et plus l'exposition est longue, moins les biais a priori ont d'importance dans le système mature. Comme souligné par Pierrehumbert (à paraître), ceci est la position adoptée par les modèles bayésiens de l'apprentissage, dont les applications récentes en linguistique permettent de modéliser l'émergence de régularités dans les systèmes phonologiques des langues du monde (Moulin-Frier, Diard, Schwartz, & Bessière,

2015) ; l'acquisition d'une stratégie de segmentation lexicale à partir de biphones chez l'enfant (Daland & Pierrehumbert, 2011) ; et le traitement de clusters consonantiques illicites dans des mots d'emprunt (Wilson & Davidson, 2013).

Les modèles bayésiens de l'apprentissage consistent à évaluer la vraisemblance d'hypothèses concernant l'état du monde. S'ils permettent d'efficacement modéliser ces processus linguistiques, et de nombreux autres processus cognitifs par ailleurs (Griffiths, Kemp, & Tenenbaum, 2008), c'est précisément parce qu'ils se basent sur un théorème qui permet d'évaluer la vraisemblance d'hypothèses, après l'observation de données, en fonction de deux éléments cruciaux : la vraisemblance des données observées étant donné l'hypothèse, et la probabilité a priori de l'hypothèse. Sous sa forme la plus simple, le théorème de Bayes suit la formule ci-dessous :

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Dans cette formule, A représente l'hypothèse, tandis que B représente les données. $P(A|B)$ correspond à la vraisemblance de l'hypothèse après observation des données, c'est sa probabilité a posteriori. $P(B|A)$ représente la vraisemblance des données en admettant l'hypothèse A . Dans le numérateur, $P(A)$ représente la probabilité a priori de l'hypothèse A . Enfin, $P(B)$ est la probabilité, toutes hypothèses confondues, d'observer les données B . Ce dénominateur est une constante de normalisation permettant d'obtenir, lorsque l'on évalue les probabilités a posteriori de toutes les hypothèses, une probabilité totale de 1. On peut en conclure que la probabilité a posteriori $P(A|B)$ est proportionnelle à la vraisemblance des données $P(B|A)$ et la probabilité a priori $P(A)$. Ainsi, même avec une probabilité a priori $P(A)$ très élevée, la probabilité a posteriori $P(A|B)$ peut être faible si la vraisemblance des données $P(B|A)$ est extrêmement faible avec cette hypothèse, et vice-versa. C'est pourquoi les biais a priori des locuteurs jouent un rôle plus important lorsqu'ils sont exposés à un input peu important, c.-à-d. dont la vraisemblance (ou invraisemblance) ne peut être que modérée.

Même en suivant un raisonnement bayésien, il est donc attendu que les locuteurs adultes puissent adopter des comportements similaires à du probability-matching. Celui-ci reste donc cohérent avec les résultats d'études décrivant des comportements proches du probability-matching, en particulier dans des processus morphologiques pour lesquels les locuteurs ont été

exposé à un input important (Hayes & Londe, 2006 ; Hayes et al., 2009). Par conséquent, même s'il est tentant de conclure que les locuteurs adultes reproduisent les probabilités de l'input dans leurs productions, souvent, il s'agit en réalité non seulement d'une approximation, mais également d'une extrapolation à partir des données disponibles basées sur des systèmes devenus matures et stables après une exposition à un input important ayant estompé les biais a priori des locuteurs.

En plus des raisons pratiques et cognitives pour lesquelles les adultes dévient du probability-matching, Pierrehumbert (à paraître) mentionne également deux études qui traitent de cas linguistiques spécifiques présentant des patterns dont la productivité ne correspond pas à leur fréquence dans l'input. La première étude citée est celle de Becker et al. (2011) portant sur l'alternance laryngale en turc pour laquelle les locuteurs ne reproduisent pas les probabilités d'un pattern phonotactique lorsqu'ils créent de nouvelles formes. Dans le lexique turc, l'alternance laryngale peut être prédite par trois facteurs concordants : le lieu d'articulation de la consonne occlusive finale du radical ; la qualité de la voyelle qui précède ; et la longueur du mot. Cependant, les locuteurs turcs ne produisent pas d'alternance laryngale pour des pseudo-mots qui présentent la voyelle pénultième qui devrait la déclencher mais qui ne présentent pas les deux autres facteurs prédictifs. En revanche, les deux autres facteurs, quand ils apparaissent seuls dans un pseudo-mot, sont suffisant à déclencher l'alternance laryngale. Ce résultat indique que certaines régularités de l'input sont accidentelles tandis que d'autres sont motivées et que les locuteurs ne reproduisent pas les régularités accidentelles dans leurs productions (Becker et al., 2011).

Les deuxièmes études mentionnées sont celles menées par Dawdy-Hesterberg (2014) portant sur deux processus morphologiques de l'arabe. Le premier processus morphologique étudié est celui de la formation du pluriel. Dawdy-Hesterberg, remarque que dans la formation du pluriel de pseudo-noms, les locuteurs ont tendance à sur- ou sous-employer certains patterns en raison de la forte variabilité qui existe dans l'input. Les résultats montrent que le comportement des locuteurs ressemble grossièrement à du probability-matching. Cela suggère une fois de plus que les probabilités dans l'input permettent de prédire la productivité morphologique de manière approximative, malgré certaines limites évidentes. En revanche, le deuxième processus morphologique étudié présente des comportements clairement divergents du probability-matching. Il s'agit du *masdar*, un processus permettant de dériver un nom référant à une action à partir d'un verbe. Dawdy-Hesterberg (2014) étudie la formation du

masdar de pseudo-verbos possédant l'une des trois structures suivantes : [CaCaCa] ; [CaCiCa] ; ou [CaCuCa]. Chacune de ces structures favorise un pattern de *masdar* : les verbes de structure [CaCaCa] privilégient un *masdar* de structure [CaCC] ; les verbes de structure [CaCiCa] privilégient un *masdar* de structure [CaCaC] ; et les verbes de structure [CaCuCa] privilégient un *masdar* de forme [CaCaaCaT]². Pourtant, dans la formation de *masdar* de pseudo-verbos, les locuteurs privilégient le pattern [CaCC] dans des proportions similaires pour les trois structures de verbes. Cela peut s'expliquer par le fait que la fréquence de types des verbes de structure [CaCaCa] (N=1143) est bien supérieure à celle des verbes de structures [CaCiCa] (N=282) ou [CaCuCa] (N=94) dans le lexique arabe. Par conséquent, le pattern de *masdar* [CaCC] est le plus fréquent dans le lexique et est généralisé à toutes les structures de verbes. Dawdy-Hesterberg (2014) remarque que les locuteurs ne semblent pas prendre en compte les voyelles de la structure des pseudo-verbos dans la formation du *masdar*. Ce qui peut s'expliquer, entre autres, par le fait que dans le système orthographique de l'arabe, les voyelles sont notées par des diacritiques, et de manière optionnelle. Quoi qu'il en soit, les locuteurs semblent ne reconnaître qu'une seule structure avec des voyelles non spécifiées [CVCVCV] à laquelle ils appliquent le pattern le plus fréquent dans le lexique. Cela rappelle le cas présenté par Kapatsinski (2010) dans lequel un pattern général très fréquent éclipse des patterns récurrents mais spécifiques à des contextes phonologiques particuliers.

Bien entendu, le cas présenté par Kapatsinski (2010) peut être ajouté à cette liste de situations linguistiques illustrant une claire divergence du probability-matching. Dans le chapitre 4, nous présenterons dans le cas d'étude des CAL japonais un autre exemple de pattern contextuel régulier dans les formes existantes qui ne montre pas de productivité dans la création de nouvelles formes. Cependant, nous verrons que dans ce cas, la faible productivité du pattern contextuel ne peut être expliquée par la compétition avec un pattern général. En accord avec le principe des modèles bayésiens de l'apprentissage, elle pourrait plutôt être causée par l'exposition relativement faible des locuteurs à ce contexte phonologique précis dans un processus morphologique devenu productif récemment, et à l'importance que jouent les biais a priori des locuteurs dans une telle situation.

² Dans les templates représentant les patterns morphologiques non concaténatifs de l'arabe, les lettres « C » et « V » représentent n'importe quelle consonne ou voyelle respectivement, tandis que « T » représente le suffixe grammatical de féminin. Le segment « T » est considéré à part pour diverses raisons d'ordre morphologique et prosodique (Dawdy-Hesterberg, 2014).

1.3.3 Rôle de la convergence morphologique dans l'évolution des représentations

Que l'on envisage que les locuteurs reproduisent les probabilités de l'input lorsqu'ils produisent de nouvelles formes, ou que l'on considère qu'ils adoptent un comportement plus bayésien, il n'en demeure pas moins que les probabilités de l'input ont une influence considérable sur les probabilités dans l'output des locuteurs. Pourtant, nous savons par ailleurs que les locuteurs sont capables de converger très rapidement, au cours d'une simple conversation, avec leur interlocuteur. Or, ces deux observations peuvent paraître contradictoires de prime abord. S'il se basait simplement sur les probabilités de l'input auquel il a été exposé, un locuteur adulte, avec un système mature, ne devrait pas produire de formes qui diffèrent drastiquement de ces probabilités lors d'une unique interaction avec un interlocuteur. Néanmoins, il a été observé que lorsque des locuteurs communiquent entre eux, ils ont tendance à se conformer à la manière de parler des uns et des autres, c.-à-d. à converger les uns vers les autres (Garrod & Pickering, 2004 ; Giles & Coupland, 1991 ; Giles, Coupland, & Coupland, 1991 ; Kim, Horton, & Bradlow, 2011 ; Nguyen & Delvaux, 2015 ; Pardo, 2006 ; Pickering & Garrod, 2004, 2006 ; Soliz & Giles, 2014 ; inter alia, voir ci-dessous pour des références abordant la convergence à différents niveaux de la production linguistique).

Rácz, Beckner, Hay, & Pierrehumbert (2020), présentant une des rares études sur la convergence en morphologie, remarquent que ce phénomène de convergence a été largement étudié et observé dans différents domaines de la production linguistique tels que : la syntaxe (Bock, 1986 ; Estival, 1985 ; Gries, 2005 ; Hall, Ferreira, & Mayberry, 2015) ; le choix de noms (Brennan & Clark, 1996 ; Roberts, 2010) ; le débit de parole (Webb, 1972) ; la fréquence fondamentale (Babel & Bulatov, 2012 ; Gregory, Webster, & Huang, 1993) ; la réalisation des voyelles (Babel, 2010, 2012) ; etc. Bien que l'on trouve également de nombreuses recherches sur la convergence dans les choix lexicaux (Boulis & Ostendorf, 2005 ; Brandstetter et al., 2017 ; Horton, 2007 ; Horton & Brennan, 2016 ; Ibarra & Tanenhaus, 2016), rares sont les études portant sur la convergence dans le choix du processus de formation de formes complexes.

Parmi les études sur la convergence morphologiques citées, les premières sont les travaux de Szmrecsanyi (2005, 2006) portant sur les différentes marques de futur en anglais ('will see' vs 'gonna see'). Les résultats de Szmrecsanyi montrent notamment que l'occurrence d'une de ces formes est un prédicteur de son réemploi plus tard dans la conversation, même chez un autre locuteur.

Un autre exemple de convergence morphologique se trouve dans l'étude de Beckner, Rácz, Hay, Brandstetter, & Bartneck (2016) qui applique le protocole de Asch (1951, 1955, 1956), visant à étudier l'effet de la pression sociale sur le jugement des participants dans la réalisation de tâches simples (mais non linguistiques, dans le cas des études de Asch), à l'étude de la formation du passé des verbes anglais (régulier vs irrégulier). Les résultats confirment que la convergence morphologique est possible et que ce phénomène est intrinsèquement social, en cela que les participants ne convergent pas vers leurs partenaires d'interactions lorsque ce sont des robots, mais uniquement lorsque ce sont des êtres humains. De plus les résultats montrent que certains participants sont naturellement plus enclins que d'autres à imiter leurs partenaires que ce soit dans des tâches linguistiques ou autres. Enfin, le degré de convergence peut aussi être influencé par la nature des items présentés (en l'occurrence, des verbes anglais), en cela que les participants convergent plus facilement pour des verbes présentant des ambiguïtés, de la variation dans la formation de leur passé.

Bien entendu, à cette courte liste vient s'ajouter l'étude de Rácz et al. (2020) portant également sur la convergence morphologique dans la formation du passé des verbes anglais. Dans cette expérience, les participants devaient produire la forme accomplie de pseudo-verbes dans une expérience en trois phases : une première phase dépourvue d'interaction sociale ; une deuxième phase dans laquelle ils interagissaient avec un partenaire artificiel ; et une troisième phase de nouveau dépourvue d'interaction sociale. L'expérience se déroulant entièrement sur ordinateur, les participants n'avaient pas conscience du fait que leur partenaire était un robot. En effet, aucune information à ce sujet ne leur était transmise, et le robot était simplement présenté comme un partenaire de jeu connecté en ligne depuis un autre ordinateur. Les résultats montrent que le partenaire artificiel a une influence significative sur la proportion de formes régulières produites par les participants, même après son départ. De plus, l'atypicalité du comportement du robot concernant les formes qu'il crée (c.-à-d. utiliser des formes régulières pour des pseudo-verbes qui sont souvent perçus comme irréguliers par des locuteurs humains, et vice-versa) peut avoir un effet négatif sur le degré de convergence des participants.

Un premier pas vers la conciliation de l'idée de cette convergence omniprésente dans le langage et le comportement probabiliste des locuteurs est de noter que la convergence observée lors d'une brève interaction, qui plus est dans un contexte expérimental, bien qu'elle prenne en compte les probabilités présente dans l'input des locuteurs, fait également entrer en jeu d'autres facteurs d'ordre cognitif. En particulier, Beckner et al. (2016) et Rácz et al. (2020) soulignent

l'importance que peut avoir l'effet d'amorçage dans des protocoles expérimentaux tels que ceux qu'ils présentent. Ne serait-ce que dans le domaine de la morphologie, on trouve de nombreuses études rendant compte de l'effet d'amorçage (Crepaldi, Rastle, Coltheart, & Nickels, 2010 ; Grainger, Colé, & Segui, 1991 ; Meunier & Segui, 1999 ; Moder, 1992 ; Veríssimo & Clahsen, 2009). En résumé, le simple fait d'être exposé à une forme présentant un certain pattern morphologique peut conditionner les participants à appliquer ce pattern à d'autres formes structurellement similaires, particulièrement si l'occasion se présente au sein de la même tâche expérimentale, ou la même conversation, dans un laps de temps assez court. Auquel cas, ce comportement peut être assimilé à un simple réflexe d'imitation et il est alors difficile d'évaluer l'impact réel qu'aura le nouvel input sur la grammaire du locuteur, notamment sur le long terme.

Il serait erroné, en revanche, de considérer que la convergence ne serait qu'un effet de court terme, entièrement assimilable à un effet d'amorçage, et qu'il n'en découle aucune modification des représentations des locuteurs. En effet, nous savons que l'accumulation d'interactions avec divers interlocuteurs peut entraîner un changement linguistique progressif, non seulement au niveau individuel mais également au niveau communautaire (voir Foulkes & Hay [2015] pour une revue de la littérature sur le sujet).

Une série de travaux visant à reproduire l'apprentissage itératif d'une langue dans des conditions expérimentales arrivent à des conclusions similaires. Cette idée a tout d'abord été explorée à l'aide d'une modélisation d'une population d'agents bayésiens apprenant une langue artificielle dans une chaîne de transmission où l'input fourni à un agent est l'output produit par l'agent de la génération précédente (Griffiths & Kalish, 2007 ; Kirby, 2001 ; Kirby et al., 2007 ; Kirby & Hurford, 2002). Ces études montrent que sous l'effet de phénomènes de transmission culturelle, des structures émergent rapidement dans les langues artificielles transmises, bien que ces structures ne soient pas présentes dans l'input fourni à la première génération d'agents.

Kirby et al. (2007), en particulier, montrent que de légers biais dans les préférences des agents au début de la chaîne de transmission sont rapidement amplifiés par les phénomènes de transmission culturelle jusqu'à devenir des règles bien établies pour les agents en fin de chaîne de transmission. Le principe sous-jacent à ce phénomène est que la transmission de la langue d'un agent à un autre agit comme un goulot d'étranglement. L'output produit par un agent, qui servira d'input à la génération suivante, ne peut être qu'un échantillon de l'input auquel il a été lui-même exposé. Les formes retenues dans cet échantillon peuvent ainsi présenter une certaine

régularité, indépendamment de la volonté de l'agent qui les produit. Le caractère itératif de l'apprentissage permet ensuite à ces régularités de se confirmer au fil des générations. D'autres expériences d'apprentissage itératif de langues artificielles font appel à de véritables locuteurs humains plutôt qu'à des modèles agent-based et viennent confirmer ces observations en obtenant des résultats similaires (Beckner et al., 2017 ; Kirby et al., 2008 ; Smith & Wonnacott, 2010).

Notons que dans la plupart de ces études, les structures observées dans les langues en fin de chaîne de transmission semblent être une simple amplification des préférences des agents/locuteurs en début de chaîne de transmission, sans modification de l'ordre hiérarchique de ces préférences. Pour simplifier, il est observé qu'une légère préférence en entrée de chaîne de transmission devient largement prédominante en sortie. Cependant, Smith (2009) montre qu'en complexifiant le fonctionnement de la transmission de la langue dans un modèle de population d'agents bayésiens, les structures observées en fin de chaîne de transmission ne sont alors plus nécessairement le simple reflet des préférences des agents en début de chaîne. Smith (2009) observe que la transmission de la langue dans une véritable communauté de locuteur ne se fait pas uniquement d'un seul locuteur d'une génération vers un seul locuteur de la génération suivante, et que les locuteurs d'une même génération interagissent et s'influencent également les uns les autres. Les modèles et les protocoles expérimentaux utilisés, se basant sur une transmission « un à un », simplifient fortement la dimension verticale de la transmission (c.-à-d. d'une génération à la suivante) et effacent entièrement sa dimension horizontale (c.-à-d. entre locuteurs d'une même génération). Afin de tester si cela peut avoir une influence sur les résultats obtenus, Smith (2009) propose d'étudier des modèles intégrant un peu plus de complexité dans la verticalité de la transmission, en cela que les agents reçoivent leur input non pas d'un seul agent de la génération précédente, mais de plusieurs. Ses résultats montrent que dans ce cas, le lien entre les préférences dans la population d'agents en entrée de chaîne de transmission et les structures observées en sortie n'est pas transparent. Parfois, les préférences les plus répandues parmi les agents en entrée de chaîne de transmission n'émergent pas comme structures dans les langues en sortie de chaîne, et à leur place émergent des structures qui ne se retrouvent que dans des proportions faibles ou moyennes dans les a priori des agents en entrée de chaîne. Sans remettre en question le rôle crucial que peuvent jouer les interactions répétées entre locuteurs dans la réduction de la variation et l'émergence de structures, ces résultats soulignent la nature parfois imprévisible de ce phénomène d'émergence dans un système aussi

complexe qu'une véritable communauté de locuteurs. Par ailleurs, ces résultats expérimentaux pourraient apporter des éléments d'explication concernant les cas évoqués en section 1.3.2 de patterns dominants dans les formes existantes qui ne sont pourtant pas productifs dans la création de nouvelles formes. Il n'est pas à exclure que la nature des interactions entre les locuteurs permette à ces patterns d'être fréquemment employés sans pour autant être la préférence majoritaire dans la communauté, et donc sans être particulièrement productif dans la création de nouvelles formes.

L'intuition que les structures linguistiques émergent et évoluent avec l'accumulation d'interactions entre les locuteurs se vérifie donc de manière expérimentale. Au cœur de ce phénomène se trouve l'idée que les locuteurs convergent peu à peu les uns vers les autres lors de chaque interaction. Cependant, il reste difficile de mesurer de quelle manière la convergence à court terme, lors d'une tâche expérimentale ou d'une conversation, influence la grammaire des locuteurs sur le long terme. Comme souligné dans plusieurs des études sur la convergence citées dans cette section, comprendre le lien entre la convergence et les changements linguistiques de long terme reste à ce jour un défi à relever afin de comprendre les processus sous-jacents à l'acquisition et l'évolution des langues. La réalisation d'études longitudinales, comme celle de Sonderegger, Bane, & Graff (2017), est probablement une piste à explorer pour comprendre la nature de ce lien. Cette étude s'intéresse à l'évolution de l'accent chez des participants à une émission de télé-réalité vivant dans une maison isolée et constamment sous enregistrement pendant trois mois. Les résultats montrent que la prononciation des participants fluctue largement au quotidien mais que les changements sur le long-terme (c.-à-d. plusieurs semaines à quelques mois) n'apparaissent que de manière sporadique. Cela suggère qu'il existe effectivement un lien entre les deux phénomènes et les auteurs identifient au moins deux facteurs influant sur la force de ce lien : d'une part, l'importance du contraste entre les sons produits par les uns et les autres ; et d'autre part, les différences intrinsèques chez les individus dans leur inclinaison à modifier leur accent. Ces résultats ne sont pas sans rappeler ceux de Beckner et al. (2016) évoqués précédemment.

Le défi consistant à éclaircir le lien entre la convergence à court terme et à long terme dépasse malheureusement le cadre de cette thèse. Dans les chapitres qui suivent (notamment chapitres 5 et 8), nous présenterons toutefois deux expériences sur la convergence morphologique qui adaptent le protocole expérimental de Rácz et al. (2020) à nos deux cas d'étude de processus morphologiques japonais. Nous verrons que sur le court terme les

locuteurs sont capables de converger rapidement avec leur interlocuteur dans la manière de créer de nouvelles formes, et ce même si les patterns utilisés par l'interlocuteur sont rares ou quasi-absents dans les formes existantes. Cela indique que la convergence est un moteur potentiel de l'apparition et de la généralisation de nouveaux patterns morphologiques, bien que son effet réel sur le long terme soit également dépendant de facteurs propres à la nature de la communauté de locuteurs et de leurs interactions. Toutefois, l'un des intérêts du protocole expérimental employé est de proposer une phase de création de nouvelles formes avant et après la phase d'interaction avec un partenaire artificiel. Cela permet d'observer que l'effet de la convergence n'est pas limité à la seule interaction avec l'interlocuteur mais perdure au contraire, si ce n'est que quelques instants, après le départ de ce dernier.

Chapitre 2 – Les Compound Abbreviated Loanwords (CAL) japonais

Le premier processus morphologique que nous avons étudié dans le cadre de nos travaux est celui permettant de construire des Compound Abbreviated Loanwords en japonais. Dans ce chapitre, nous présenterons le fonctionnement de ce processus morphologique récemment devenu productif en japonais. La more jouant un rôle important dans la manière dont ces abréviations sont formées, nous décrirons tout d’abord le rôle de cette unité prosodique dans la langue japonaise en général (section 2.1). Puis, nous présenterons le principe général de formation des CAL (section 2.2). Enfin, nous illustrerons les différents patterns que peuvent suivre ces abréviations, notamment lorsqu’elles ne respectent pas le principe général de formation, en fonction du matériel phonologique leur servant de base (section 2.3).

2.1 Le rôle de la more en japonais

Les Compounds Abbreviated Loanwords (CAL) japonais sont des abréviations de deux mots empruntés à une langue étrangère, la plupart du temps à l’anglais. Ils sont formés en combinant les deux mots d’emprunt en une seule abréviation. Un exemple bien connu est *pokemon*, qui est l’abréviation des mots anglais ‘pocket’ et ‘monster’³ (adaptés en *poketto* et *monsutaa* en japonais). Ce processus morphologique est devenu productif récemment en japonais avec les emprunts lexicaux massifs à l’anglais qui sont arrivés après la fin de la Seconde Guerre mondiale. En théorie, n’importe quelle paire de mots d’emprunt peut aboutir à

³ Nous utiliserons dans cette thèse différents types de transcriptions que nous distinguerons comme suit :

- L’italique est réservé aux transcriptions orthographiques en japonais selon la méthode de romanisation Hepburn (Hepburn, 1903), avec voyelles longues retranscrites par un dédoublement de la voyelle. Ex : [a:] > aa ; [i:] > ii ; etc.
- Les transcriptions phonologiques /exemple/, phonétiques [exemple], et orthographiques < れい > sont toujours en langue japonaise.
- Les guillemets anglais simples sont utilisés pour la traduction de ces exemples en anglais ou français. Ex : ree [ree] ‘exemple’.
- Tous les types de transcriptions du japonais pourront être enrichies par un découpage en more (et non en syllabe) par l’ajout de points « . ». Ex : re.e ; /re.:/ ; [re.:] ‘exemple’

Notons également un autre détail typographique : le point, plutôt que la virgule, a été utilisé comme séparateur des décimales pour tous les nombres présentés dans cette thèse.

la création d'un CAL. Cela arrive fréquemment lorsqu'il s'agit de nommer des nouvelles technologies, des marques et enseignes et également avec les noms propres, notamment de personnes célèbres.

La formation des CAL est fortement influencée par la structure moraïque des formes de base qui sont combinées dans l'abréviation. Par conséquent, pour étudier ce processus morphologique il est nécessaire de comprendre précisément le rôle de la more en japonais. La more est une unité prosodique dont la taille se situe entre le phonème et la syllabe. Dans les langues où la more est pertinente, elle est liée au poids des syllabes : les syllabes légères (c.-à-d. sans coda) pèsent une more ; tandis que les syllabes lourdes (c.-à-d. avec coda) pèsent deux mores (Davis, 2011).

Le statut de la more en tant qu'unité rythmique a longtemps fait débat dans la littérature, et différentes analyses de son rôle en japonais ont servi comme argument principal pour les différents points de vue à ce sujet. Les premières mentions du rythme particulier du japonais date de Bloch (1950), qui décrit la prononciation du japonais comme une succession rapide de fractions de même longueur dans un rythme *staccato*. Bloch affirme que les locuteurs sont capables de compter ces fractions, en cela qu'ils perçoivent les mots et phrases contenant le même nombre de fractions comme étant de même durée. Ladefoged (1975) a par la suite affiné cette description du japonais en suggérant que la more serait l'unité pertinente de cette langue et que les mores avaient approximativement la même durée en production. Cette observation fait écho aux travaux sur la typologie rythmique des langues basée sur l'isochronie de certaines unités rythmiques et distinguant d'un côté les langues *syllable-timed*, telle que le français, et de l'autre les langues *stress-timed*, telle que l'anglais (Abercrombie, 1967 ; Pike, 1945). Viendrait alors s'ajouter un troisième type, les langues *mora-timed*, telle que le japonais. Cependant, cette typologie rythmique est maintenant largement démentie par manque de preuve concrète d'une quelconque isochronie des différentes unités rythmiques (voir Arvaniti [2009] pour une revue de littérature).

Pour le cas du japonais, cela a mené à des questionnements sur l'existence même de la more (voir Beckman [1982] et Warner & Arai [2001] pour une revue de la littérature sur le sujet). Beckman (1982) suggère que l'existence même de la more pourrait n'être que le résultat de l'influence du système d'écriture qui, dans les syllabaires japonais, présente une correspondance quasi parfaite entre caractères et mores. Cependant, bien que Warner & Arai

(2001) reconnaissent que les études sur la réalité phonétique de la more, notamment en production, ont échoué à apporter des preuves convaincantes, ils soulignent que les études du rôle de la more en perception affichent des résultats prometteurs. Ramus & Mehler (1999) ont montré à l'aide d'une expérience impliquant la perception de parole dégradée que le rythme syllabique (qui comprend des caractéristiques telles que la longueur des consonnes, la diversité des types et des poids des syllabes, l'organisation temporelle des consonnes et voyelles, etc.) était un indice suffisant pour des locuteurs français afin de différencier des phrases anglaises et japonaises. De plus, dans une série d'expérience sur la segmentation de mots et de détection de phonèmes, les locuteurs japonais affichent des patterns de réponses suggérant qu'ils utilisent une stratégie de segmentation basée sur la more (Cutler & Otake, 1994 ; Otake, Hatano, Cutler, & Mehler, 1993). Ce pattern diffère à la fois des stratégies des locuteurs de l'anglais et du français, dont la différence avait déjà été établie par Cutler et al. (1986). Ces études attestent donc de l'existence de la more en tant qu'unité prosodique.

Dans une analyse récente du japonais de Tokyo (c.-à-d. le japonais standard), Labruno (2012a) questionne la pertinence de la syllabe pour la description de cette langue, en passant en revue les différents arguments récurrents des descriptions du japonais basées sur la syllabe. Elle présente une description de la langue basée uniquement sur la more et le pied qui s'avère aussi concise et explicative que l'alternative employant la syllabe. Bien que cette dernière étude remette sérieusement en cause le rôle de la syllabe en japonais, le débat sur l'existence des unités prosodiques dans cette langue dépasse le cadre de nos travaux. Par conséquent, dans cette thèse, nous adopterons un point de vue plus précautionneux concernant l'existence de la syllabe et de la more en japonais. La more sera considérée comme une unité de mesure du poids des syllabes qui joue un rôle structurel non seulement en japonais mais également dans divers phénomènes linguistiques à travers les langues du monde. En ce sens, les définitions que nous adoptons pour ces termes se rapprochent de celles proposées dans les travaux de Davis (2011) sur les phénomènes linguistiques liés à la quantité. Même en adoptant ce positionnement, nous considérons que la more est l'unité prosodique qui joue un rôle de premier ordre dans de nombreux et divers aspects de la langue japonaise tels que les patterns d'accentuation, les erreurs de parole, l'allongement compensatoire, la troncation et formation de mots-valises, la régularisation temporelle dans la poésie et les jeux de langue, etc. (Kubozono, 1989, 1999).

La plupart des mores japonaises correspondent à une syllabe légère de structure (C)(j)V. Néanmoins, il existe trois segments moraiques spéciaux pouvant apparaître en position de coda

et qui sont traditionnellement transcrits /R/, /Q/ et /N/. Une partie au moins de la réalisation de ces trois segments est sous-spécifiée :

- /R/ correspond à l’allongement de la voyelle précédente. Exemple : /to.R.kjo.R/ contient quatre mores et se prononce [to:kjo:] ‘Tokyo’.
- /Q/ correspond à la première partie d’une consonne géminée. Exemple : /sa.Q.po.ro/ contient quatre mores et se prononce [sapporo] ‘Sapporo’.
- /N/ est une consonne moraïque nasale homorganique, c’est-à-dire, qui assimile le lieu d’articulation de la consonne qui suit. Exemple : /bo.N.sa.i/ contient quatre mores et se prononce [bonsai] ‘bonsai’. Notons que /N/ possède des réalisations diverses lorsqu’il n’est pas suivi d’une consonne, notamment, il se réalise [N] avant une pause (Labrune, 2006a).

Les trois types de syllabes lourdes du japonais peuvent être décrites dans une approche autosegmentale de la phonologie (J. A. Goldsmith, 1976 ; Gussenhoven & Jacobs, 2011). La différence entre /R/ et /Q/ se situe au niveau de la tier CV, où /R/ est spécifié en tant que voyelle (V) tandis que /Q/ est spécifié en tant que consonne (C). Leur réalisation étant entièrement sous-spécifiée par ailleurs, ils assimilent tous les traits du segment disponible le plus proche ayant la même spécification sur la tier CV. Ainsi, /R/ assimile tous les traits de la voyelle qui précède, alors que /Q/ assimile tous les traits de la consonne qui suit. /N/ est similaire à /Q/ en cela qu’il est spécifié comme consonne sur la tier CV. En revanche, contrairement à /Q/, son mode d’articulation est spécifié comme nasal, ne laissant la possibilité que pour l’assimilation du lieu d’articulation de la consonne suivante. Par conséquent, l’ajout de tiers dédiées au lieu et mode articulatoire est nécessaire afin d’illustrer la diffusion d’un seul de ces traits. Notez que seuls les traits pertinents sont affichés dans les schémas ci-dessous.

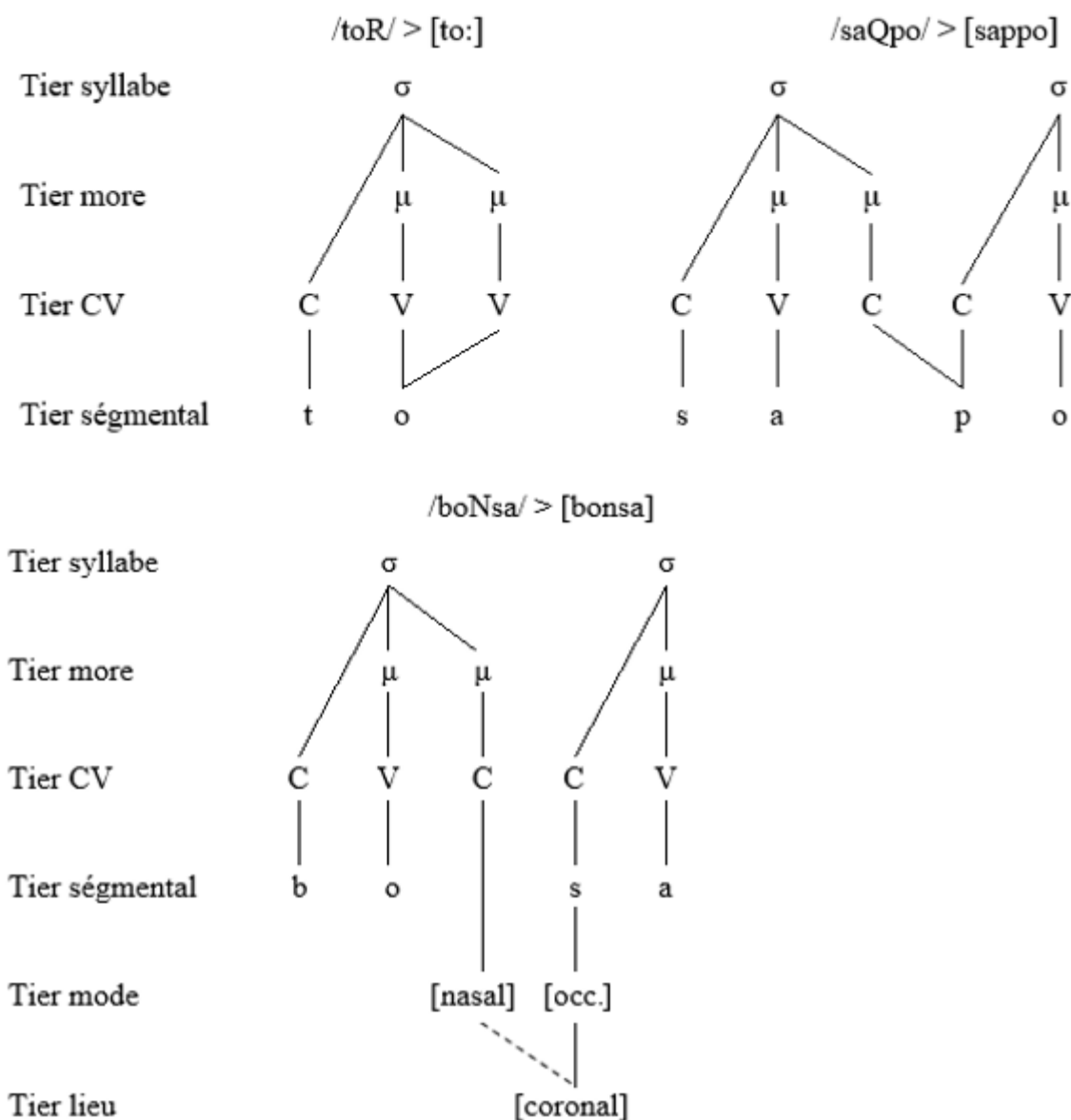


Fig.1 : structure syllabique de syllabes lourdes japonaises incluant les segments spéciaux /R/, /Q/, et /N/. Pour /Q/ et /N/, la syllabe suivante est également représentée afin d'illustrer l'assimilation des traits de l'attaque de la seconde syllabe par la coda de la première.

2.2 Le principe général de formation des CAL et ses exceptions

Comme indiqué précédemment, les CAL sont des abréviations créées en combinant deux mots d'emprunt en un seul mot court. Plus précisément, les mots qui peuvent servir de base pour la formation des CAL sont ceux qui appartiennent à la strate lexicale spécifique des emprunts récents aux langues occidentales, en opposition aux anciens emprunts au chinois.

McCawley (1968) identifie plusieurs strates dans le lexique japonais qui peuvent être décrites comme des domaines dans lesquelles certaines règles de la langue s'appliquent, et qui peuvent s'exprimer sous la forme de quatre traits morphémiques : [+ Native], [+ Sino-Japanese], [+ Onomatopoeia] et [+ Foreign].

Notons que l'appartenance d'un morphème à l'une ou l'autre de ces strates dépend fortement de la perception et du traitement que les locuteurs ont de ce mot, indépendamment de sa véritable étymologie. Cependant, les strates peuvent être grossièrement décrites comme suit : la strate 'Native' (ou *Yamato*, ou *Wago*) correspond aux mots autochtones de la langue japonaise ; la strate 'Sino-Japanese' (ou *kango*) correspond aux emprunts massifs au chinois ancien datant du IV^{ème} siècle ; la strate 'Onomatopoeia' (ou *giseigo* et *gitaigo*), qui n'est pas considérée comme une strate indépendante par certains auteurs puisque la plupart des morphèmes onomatopéiques ont une origine autochtone, correspond aux mots mimétiques, ce qui inclue non seulement les onomatopées (les *giseigo*) mais aussi d'autres types d'idéophones (les *gitaigo*) ; et la strate 'Foreign' (ou *gairaigo*) correspond aux emprunts récents aux langues occidentales dont les premiers datent du XVI^{ème} siècle environ, mais dont la quantité a fortement augmenté au cours du XX^{ème} siècle (Labrune, 2006a, 2012b).

Itô & Mester (1995b, 1995a, 1999, 2003, 2008) ont décrit la structure du lexique japonais comme une série de cercles concentriques représentant les différentes strates lexicales de la langue. Dans cette représentation, la strate au cœur de la structure est celle sur laquelle le plus de contraintes de marque s'appliquent (c.-à-d. la strate 'Native'), et la strate située le plus en périphérie de la structure est celle sur laquelle le moins de contraintes de marque s'appliquent (c.-à-d. la strate 'Foreign'). En raison des différences phonologiques entre ces différentes strates lexicales, nous définissons dans cette thèse les CAL comme étant des abréviations basées sur des paires de morphèmes appartenant uniquement à la strate périphérique 'Foreign', de la même manière que Labrune (2006b, 2007, 2008). Cela exclue donc les abréviations basées sur plusieurs morphèmes provenant d'autres strates, ou les abréviations mixtes basées sur un morphème 'Foreign' et un autre issu d'une autre strate.

La plupart des CAL japonais sont créés selon un principe général consistant à conserver les deux mores initiales des deux mots d'emprunt. Cela crée une abréviation d'une longueur de quatre mores (Labrune, 2006b, 2007, 2008). *Pokemon*, l'abréviation des deux mots anglais 'pocket' et 'monster', est un exemple de CAL respectant ce principe général de formation. Cette abréviation contient, *po.ke*, les deux mores initiales de *po.ke.t.to* (prononciation japonaise de 'pocket'), le premier constituant de l'abréviation (désormais « C1 ») ; et *mo.n*, les deux mores initiales de *mo.n.su.ta.a* (prononciation japonaise de 'monster'), le second constituant de l'abréviation (désormais « C2 »). Ceci est illustré dans la figure 2 ci-dessous (reprise de Labrune, 2007).

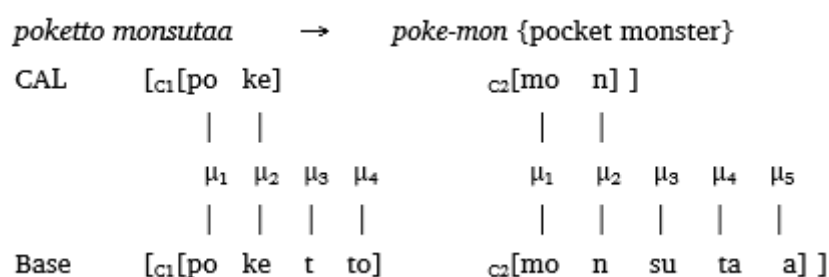


Fig.2 : dérivation de *pokemon* à partir de *poketto* et *monsutaa*.

Toutefois, tous les CAL ne suivent pas strictement ce principe général de formation. Il existe diverses manières de transgresser ce principe et on peut observer une certaine variation dans la manière dont les locuteurs japonais créent ces abréviations. Il semble, cependant, que la création d'exceptions à ce principe ne soit pas totalement arbitraire mais au contraire motivée sur le plan phonologique. Grâce à l'analyse d'un lexique d'environ 800 CAL⁴, Labrune (2006b, 2007, 2008) a pu observer que la plupart des exceptions étaient causées par la présence d'un segment spécial /R/ ou /Q/ en tant que deuxième more de C2. Cela est lié au fait que ces segments sont évités en position finale de mot dans la phonologie du japonais. Le problème est d'autant plus important dans le cas de /Q/ qui correspond à la première partie d'une consonne géminée et dont la prononciation demeure totalement non spécifiée s'il n'est pas correctement suivi d'une consonne. Lorsque /R/ ou /Q/ est la deuxième more du deuxième constituant d'un

⁴ Ce lexique a été constitué à l'aide de dictionnaires spécialisés, journaux, conversations personnelles, etc. En d'autres termes, il s'agit de tous les CAL dont Labrune a pu attester au moins une occurrence.

CAL, le principe général de formation amènerait cette more à apparaître en position finale de l'abréviation, produisant ainsi une forme illicite.

La théorie de l'Optimalité (désormais « OT », Prince & Smolensky [1993]) s'avère être une approche efficace pour la description d'un tel conflit, dans la mesure où nous pouvons redéfinir ces conditions et restrictions comme des contraintes imposées sur l'output du processus de formation des CAL. En s'appuyant sur une précédente analyse en OT proposée par Labrune (2007), Lamarque & German (2017) montrent que la variabilité dans les exceptions au principe général de formation peut être décrite en hiérarchisant la sévérité (ou *strictness*) d'un nombre limité de contraintes et, sur cette base, propose une analyse de la formation des CAL en OT Stochastique (Anttila, 1997b, 1997a ; Boersma, 1997 ; Boersma & Hayes, 2001). Une version simplifiée de cette analyse, incluant dix contraintes, est présentée en (1) ci-dessous. Sept contraintes sont spécifiques à la formation des CAL (1a-g), tandis que trois contraintes correspondent à des principes généraux de la phonologie du japonais (1h-j).

(1) Les dix contraintes impliquées dans la formation des CAL :

- a. MOTPROSODIQUEBINAIRE : le mot prosodique (c.-à-d. l'abréviation) doit être binaire en pieds⁵.
- b. ANCRAGEGAUCHE : tous les pieds de l'abréviation doivent être ancrés au bord gauche des lexèmes de la base dont ils proviennent.
- c. DEP : tous les segments de l'abréviation doivent être présents dans la base (pas d'épenthèse).
- d. IDENTTRAITS : les traits distinctifs sont identiques entre les segments correspondants dans la base et l'abréviation (pas de modification de traits distinctifs).
- e. *PIED>2 : les pieds de l'abréviation ne doivent pas contenir plus de deux mores.
- f. *PIED<2 : les pieds de l'abréviation ne doivent pas contenir moins de deux mores.
- g. CONTIG : les mores conservées dans l'abréviation doivent être contiguës dans la base.
- h. PARSE μ : toutes les mores doivent être parsées dans un pied.
- i. *V:# : pas de voyelle longue (/R/) en position finale de mot.

⁵ Le pied est une unité prosodique supérieure à la more et la syllabe. Le pied canonique est un regroupement de deux mores ou deux syllabes. Cette unité prosodique joue fréquemment un rôle dans les patterns accentuels des langues en cela qu'il est courant que l'un des constituants du pied soit accentué, et l'autre non. En japonais, le pied canonique est binaire en mores (Labrune, 2005).

j. LICENCE/Q/ : le segment spécial /Q/ ne doit pas être sous-spécifié.

Les quatre premières contraintes favorisent la création de CAL de deux pieds (MOTPROSODIQUEBINAIRE) correspondant aux mores initiales (ANCRAGEGAUCHE) des constituants de base, sans qu'aucun matériel phonologique ne soit ajouté (DEP) ou modifié (IDENTTRAITS). MOTPROSODIQUEBINAIRE n'est jamais enfreinte dans les CAL existants, et s'il existe quelques cas marginaux d'infractions des trois autres contraintes, ils ne sont jamais provoqués par le conflit sur lequel porte notre travail entre ces contraintes et les contraintes phonologiques générales de la langue. Par ailleurs, PARSE μ est une contrainte très générale et jamais enfreinte en japonais qui impose que toutes les mores soient incluses dans un pied, mais elle ne joue pas d'autre rôle dans ce conflit que d'exclure systématiquement les candidats contenant des mores non parsées. Par conséquent, nous nous concentrerons dans la suite sur les interactions entre cinq contraintes restantes : *PIED>2, *PIED<2 et CONTIG d'une part ; et *V:# et LICENCE/Q/, d'autre part.

Comme nous l'avons vu précédemment, ces contraintes seront en conflit lorsque /R/ ou /Q/ est la deuxième more du deuxième constituant d'un CAL. Dans cette situation, les locuteurs japonais utilisent une de ces quatre stratégies pour résoudre le conflit :

1- Le pattern de Maintien : cela consiste à maintenir la more problématique et former une abréviation de quatre mores. En d'autres termes, cela revient à appliquer le principe général de formation malgré la présence de la more problématique en position finale de mot. Ceci n'est réellement possible que dans le cas de /R/, car il n'existe pas de prononciation possible pour /Q/ lorsqu'il n'est pas suivi d'une consonne. Nous trouvons cependant deux exemples de CAL pour lesquels nous pourrions considérer que /Q/ y a été maintenu avec une prononciation non conventionnelle. Dans le premier cas, /Q/ est réalisé comme un allongement de la voyelle précédente (c.-à-d. comme /R/) au lieu d'une gémiation de consonne ('no good' > /no.R.gu.Q.do/ > [no.o.gu.u]) ; dans le second cas, /Q/ est prononcé *tsu* ('world cup' > /wa.R.ru.do.ka.Q.pu/ > [wa.a.ka.t̃su]). Cette dernière prononciation provient de la proximité entre /Q/ et *tsu* dans le système d'écriture, où le caractère pour /Q/ <ツ> est le même que celui pour *tsu* <ツ> réduit à la moitié de sa taille, et souvent appelé *chiisai tsu* 'le petit *tsu*'. Selon que le pattern de Maintien est appliqué à /R/ ou /Q/, cela entraînera respectivement une infraction aux contraintes *V:# et LICENCE/Q/.

Ex : - /R/ est maintenu dans *chi.R.ba.R*⁶

	C1	C2
En anglais :	cheese	burger
En japonais :	チーズ[<i>t̃ei.R.zu</i>]	バーガー[<i>ba.R.ga.R</i>]
CAL :	チーバー [<i>t̃ei.R.ba.R</i>] 4 mores	

- /Q/ est maintenu dans *wa.R.ka.tsu*

	C1	C2
En anglais :	world	cup
En japonais :	ワールド[<i>wa.R.ru.do</i>]	カップ[<i>ka.Q.pu</i>]
CAL :	ワーカツ [<i>wa.R.ka.t̃su</i>] 4 mores	

2- Le pattern de Suivi : dans ces abréviations, la more problématique est maintenue mais est suivie d'une autre more afin qu'elle n'apparaisse pas en position finale de mot, empêchant ainsi l'infraction aux contraintes phonologiques correspondantes. Cependant, le pattern de Suivi est une infraction à la contrainte *PIED>2. Ce pattern crée des abréviations de cinq mores de longueur.

Ex : - /R/ est suivi dans *pu.ra.ke.R.su*

	C1	C2
En anglais :	plastic	case
En japonais :	プラスチック[<i>pu.ra.su.t̃ei.Q.ku</i>]	ケース[<i>ke.R.su</i>]
CAL :	プラケース [<i>pu.ra.ke.R.su</i>] 5 mores	

- /Q/ est suivi dans *po.su.do.Q.ku*

	C1	C2
En anglais :	post	doc
En japonais :	ポスト[<i>po.su.to</i>]	ドック[<i>do.Q.ku</i>]
CAL :	ポストック [<i>po.su.do.Q.ku</i>] 5 mores	

⁶ Par souci de clarté, /R/, /Q/ et /N/ seront désormais systématiquement transcrits par « R », « Q » et « N » dans toutes les transcriptions du japonais portant sur les CAL (chapitres 2 à 5), y compris phonétiques. Leurs réalisations effectives sont déductibles du contexte et sont d'une moindre importance dans notre propos que de clairement repérer quel segment apparaît dans quelle position du mot.

3- Le pattern de Suppression : dans ce cas, la more problématique est simplement supprimée de l'abréviation. Ainsi elle n'apparaît plus en position finale de mot, mais ceci est une infraction à *PIED<2. Ce pattern crée des abréviations d'une longueur de trois mores.

Ex : - /R/ est supprimé dans *mi.su.do*

	C1	C2
En anglais :	mister	donuts
En japonais :	ミスター[mi.su.ta.R]	ドーナツ [do.R.na.t̃su]
CAL :	ミストド[mi.su.do] 3 mores	

- /Q/ est supprimé dans *bu.ra.pi*

	C1	C2
En anglais :	Brad	Pitt
En japonais :	ブラッド[bu.ra.Q.do]	ピット [pi.Q.to]
CAL :	ブラピ[bu.ra.pi] 3 mores	

4- Le pattern de Remplacement : cela consiste à supprimer la more problématique mais à la remplacer par la more qui suit dans le constituant de base. Cela n'enfreint pas les contraintes sur les mores apparaissant en position finale de mot et crée un CAL de quatre mores, mais contrairement au principe général de formation, les mores préservées dans l'abréviation ne sont pas contiguës dans la base. Cela constitue donc une infraction à CONTIG.

Ex : - /R/ est remplacé dans *re.N.sa.ba*

	C1	C2
En anglais :	rental	server
En japonais :	レンタル[re.N.ta.ru]	サーバー[sa.R.ba.R]
CAL :	レンサバ[re.N.sa.ba] 4 mores	

- /Q/ est remplacé dans *a.me.fu.to*

	C1	C2
En anglais :	American	football
En japonais :	アメリカン[a.me.ri.ka.N]	フットボール[φu.Q.to.bo.R.ru]
CAL :	アメフト[a.me.φu.to] 4 mores	

Nous verrons dans les sections qui suivent dans quelles proportions ces différents patterns sont présents dans les formes existantes. Cela nous permettra, entre autres, d'établir une hiérarchie entre les différentes contraintes. Nous laissons donc la question de la hiérarchisation en suspens, celle-ci sera abordée en section 3.3.

2.3 Emergence de patterns dans les CAL

À la suite de son analyse d'environ 800 CAL, Labrune (2007) remarque que les abréviations de quatre mores représentent 77.6% de son lexique. Ceci s'explique en grande partie par le fait que le principe général de formation entraîne la création d'abréviation de quatre mores. Par ailleurs, les CAL de trois mores représentent 16.6% du lexique, et sont bien plus fréquents que les CAL de cinq mores (4.6%), de six mores (0.4%) ou de deux mores (0.4%). La plupart des CAL de trois mores sont provoqués par l'application du pattern de Suppression sur /R/ ou /Q/ en tant que deuxième more du deuxième constituant. Il est à noter toutefois que tous les CAL de quatre mores ne respectent pas nécessairement le principe général de formation, et tous les CAL de trois mores ne sont pas causés par la Suppression de /R/ ou /Q/.

Labrune identifie également des contextes phonologiques plus précis que la simple présence de /R/ ou /Q/ en tant que deuxième more du deuxième constituant qui vont accentuer cette tendance à employer des patterns irréguliers. Par exemple : dans le cas de /R/ cette more aurait tendance à être davantage supprimée lorsque la voyelle allongée est une voyelle non haute. Cela est probablement dû au fait que les voyelles hautes sont phonétiquement plus courtes en japonais (Shimizu Han, 1962), rendant leur allongement moins problématique. Aussi, de manière générale, le pattern de Suppression serait davantage employé lorsque le second constituant débute par une consonne bilabiale ; et le pattern de Suivi serait plus fréquemment employé lorsque que le second constituant est trimoraïque. Nous reviendrons sur ces cas particuliers lorsque nous étudierons la fréquence de types des différents patterns présents dans les lexiques en section 3.2. Retenons toutefois que les travaux de Labrune ont permis de mettre en lumière un ensemble de facteurs morpho-phonologiques à l'origine de la création de la plupart des CAL réguliers et irréguliers.

Il est intéressant de noter que les CAL de trois mores sont assez fréquents pour que leur existence soient remarquées par les locuteurs. Lors de nos précédents travaux sur les CAL,

certaines locuteurs nous avaient indiqué que de telles exceptions étaient plus fréquentes dans les abréviations les plus récentes. Dans nos études précédentes, nous avons donc testé cette intuition de locuteur en formulant l'hypothèse que le processus de formation des CAL était encore en cours de régularisation, notamment concernant le traitement de /R/ et /Q/ en tant que deuxième more du deuxième constituant. Nous avons tout d'abord créé un lexique de 300 CAL employés dans le contexte des jeux vidéo (désormais le « lexique jeux-vidéo ») puisqu'il s'agit d'un média devenu populaire récemment et dans lequel de nouvelles abréviations sont créées régulièrement. L'analyse du lexique jeux-vidéo montre que le pattern de Suppression y est plus souvent employé sur /R/ et /Q/ que dans le lexique de Labrune. En effet, il est employé dans plus de 80% des cas dans le lexique jeux-vidéo, contre un peu moins de 50% des cas dans le lexique de Labrune. Par conséquent, le lexique jeux vidéo contient une plus grande proportion de CAL de trois mors (Lamarque, 2015).

Dans une étude suivante, nous avons demandé à des locuteurs japonais de différents groupes d'âge de créer de nouvelles abréviations à partir de paires de mots anglais. L'expérience se déroulait sous forme d'un questionnaire en ligne avec questions ouvertes. Comme attendu, les jeunes locuteurs avaient tendance à utiliser le pattern de Suppression significativement plus souvent sur /R/ et /Q/ que les participants plus âgés, créant ainsi davantage de CAL de trois mors. Néanmoins, les résultats de cette expérience montraient un niveau étonnamment élevé de variation à la fois inter- et intra-individuelle. Notamment, le pattern de Suppression n'était employé sur /R/ et /Q/ au mieux (c.-à-d. par les jeunes locuteurs) que dans moins de 30% des cas (Lamarque, 2016 ; Lamarque & German, 2017). Les résultats de ces études soulèvent la question de l'acquisition par les locuteurs japonais des patterns gouvernant la création des CAL, puisque l'on remarque un fort décalage entre les probabilités observées dans les formes attestées dans les lexiques, qui tendent vers la régularité, et celles observées dans les nouvelles formes créées par les locuteurs, qui montrent une forte variabilité. Si la productivité d'un pattern est étroitement liée à sa fréquence de types, comme cela est suggéré dans la littérature, un tel décalage est difficile à expliquer. Par ailleurs, en partant de ce même principe, on peut se demander ce qui se produit au tout début de l'apparition d'un processus morphologique, pour lequel, par définition, les formes existantes sont rares. Dans les sections qui suivent, nous aborderons cette question en réanalysant le lexique de Labrune et le lexique jeux-vidéo afin d'obtenir des données précises sur la fréquence de types des différents patterns présents dans les CAL. Puis nous comparerons les résultats de cette réanalyse à ceux

d'une tâche de création de nouvelles formes focalisée sur ces patterns. Nous cherchons en particulier à évaluer la validité des prédictions basées sur les fréquences de types dans les lexiques quant aux résultats de l'expérience de création de nouvelles formes.

Chapitre 3 – Étude du lexique de CAL japonais

Dans notre étude des CAL, nous nous intéressons à la productivité des différents patterns dans la création de nouvelles abréviations. Comme expliqué en section 1.3.1, la productivité d'un pattern dépend fortement de sa fréquence de types dans le lexique. Par conséquent, la première étape de nos recherches consiste à réanalyser les deux lexiques de CAL à notre disposition en se focalisant sur les différents patterns de traitement de la seconde more du second constituant. Cela nous permettra d'obtenir des données précises sur la fréquence de types de ces patterns. Nous utiliserons ces données sur les fréquences de types pour formuler des prédictions quant à leur productivité dans la création de nouvelles formes, c'est-à-dire, à quel point ils devraient être réemployés par les locuteurs lorsqu'ils créent de nouveaux CAL.

3.1 Méthode

Bien que les deux lexiques de CAL à notre disposition se soient avérés particulièrement utiles pour comprendre les phénomènes morphophonologiques qui entraînent en jeu dans la création des CAL, ils n'ont pas été créés dans le but initial de fournir des données précises sur la fréquence de types des différents patterns de traitement de la deuxième more du deuxième constituant. En effet, puisqu'il s'agit de listes exhaustives de tous les CAL rencontrés dans diverses sources, certains C2 y apparaissent plusieurs fois, avec différents C1, et vice-versa. En l'état, ils ne permettaient donc pas d'obtenir des données précises sur les fréquences de types concernant la deuxième more du C2. C'est pour cette raison que nous avons dû entreprendre une réanalyse de ces lexiques. Pour ce faire, nous avons combinés le lexique de Labrune et le lexique jeux-vidéo avant d'en exclure tous les C2 qui apparaissaient plusieurs fois. Par exemple, les CAL pour 'post card' (*po.su.to ka.R.do* > *po.su.ka*) et pour 'trading card' (*to.re.R.di.N.gu ka.R.do* > *to.re.ka*) ont tous les deux le même C2 « *ka.R.do* », dans lequel /R/ est la deuxième more. Dans les deux cas, le pattern de Suppression est appliqué. De tels exemples devraient donc être considérés comme un seul type de C2. Dans de rares cas cependant, le même mot de base servant de C2 peut être traité différemment dans deux CAL différents. Par exemple, les CAL pour 'skate board' (*su.ke.R.to bo.R.do* > *su.ke.bo.R*) et 'snow board' (*su.no.R bo.R.do* > *su.no.bo*) ont tous les deux le même C2 « *bo.R.do* », dans lequel /R/ est la deuxième more.

Toutefois, dans un cas, /R/ est maintenu, tandis que dans l'autre, il est supprimé. Dans ce genre de cas, nous avons considéré chacun des deux formes comme un type différent.

En supprimant tous les doublons de la sorte, le nouveau lexique combiné contient 722 C2 uniques (sur environ 1100 abréviations). Parmi ces 722 C2, 21 présentent un comportement atypique ne correspondant à aucun des quatre patterns employés de manière récurrentes dans le traitement de la deuxième more de C2. Cela inclue, par exemple, des CAL dans lesquels les mores finales du C2 sont préservées, plutôt que les mores initiales (infraction à ANCRAGEGAUCHE), ou dans lesquels une more de la base est modifiée dans l'abréviation (infraction à IDENTTRAITS), ou encore dans lesquels une more n'étant pas présente dans la base est ajoutée dans l'abréviation (infraction à DEP). Ces CAL ont été exclus de notre analyse, ce qui amène le nombre total de C2 considérés pour mesurer la fréquence de types des différents patterns à 701.

3.2 Résultats

Les principaux résultats de notre réanalyse du lexique sont présentés dans la figure 3 ci-dessous. Sur ce graphique, on peut voir la proportion d'utilisation de chacun des quatre patterns communs pour le traitement de la deuxième more du C2, en fonction du type de more dont il s'agit. Nous avons considéré trois types de mores :

- /R/ : la more d'allongement vocalique, problématique en position finale de mot.
- /Q/ : la more correspondant à la première partie d'une consonne géminée, fortement problématique en position finale de mot.
- *Sans conflit* : toute more non problématique en position finale de mots. Cela inclut des mores de structure /(C)(j)V/ et la more nasale homorganique /N/. Autrement dit, tout type de mores à l'exclusion de /R/ et /Q/.

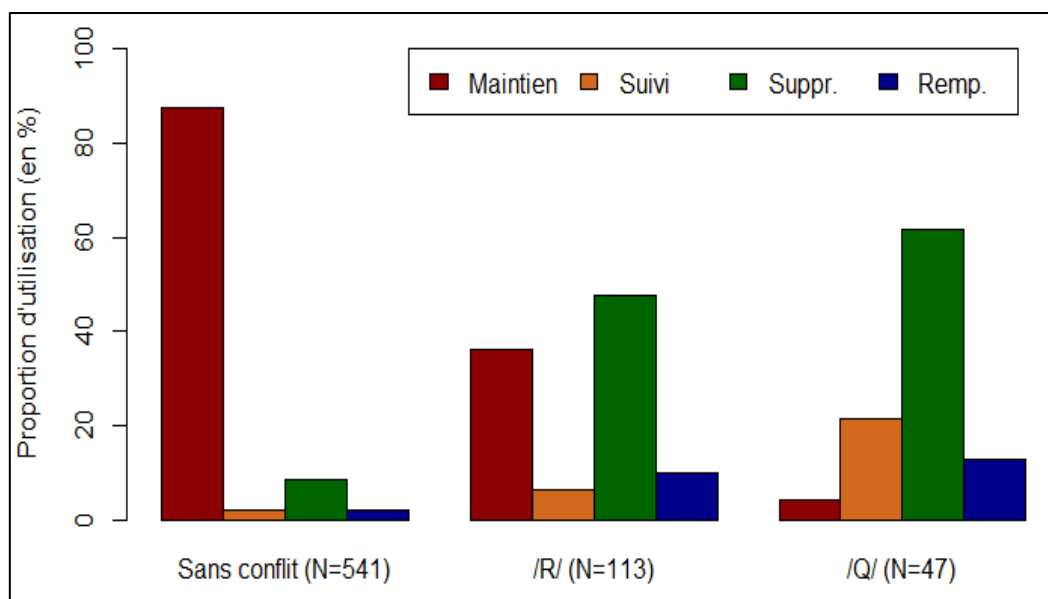


Fig.3 : proportion d'utilisation (en %) de chaque pattern dans le lexique, en fonction du type de deuxième more du C2.

Il est important de noter en premier lieu que l'ensemble des mores *sans conflit* en tant que deuxième more du C2 est naturellement bien plus fréquent que /R/ et /Q/. L'allongement vocalique /R/ n'apparaît dans cette position que dans 16% des cas, et /Q/ est encore plus rare, n'apparaissant que dans 7% des cas.

Pour les mores *sans conflit*, c'est-à-dire lorsque le principe général de formation des CAL n'enfreint aucune contrainte phonologique concernant les mores apparaissant en position finale de mot, le traitement le plus employé est sans surprise le pattern de Maintien (dans 88% des cas). Lorsque le pattern de Maintien n'est pas utilisé, le pattern de Suppression (8% des cas) est légèrement plus utilisé que les deux autres (2% des cas chacun).

Concernant la more /R/, pour laquelle le principe général de formation engendre une infraction à la contrainte *V:#, les patterns de Maintien et de Suppression sont tous les deux fortement représentés, avec une légère prédominance du pattern de Suppression. En revanche, les patterns de Suivi et de Remplacement sont peu employés.

Enfin, dans le cas de la more /Q/, pour laquelle le principe général de formation engendre une infraction à la contrainte pratiquement inviolable LICENCE/Q/, le pattern de

Maintien est pratiquement absent⁷. Le pattern de Suppression est le plus employé dans ce cas, mais le pattern de Suivi est également proportionnellement plus présent dans le cas de /Q/ que dans celui des autres types de mores. Il s'avère que dans tous les cas d'emploi du pattern de Suivi sur la more /Q/, le mot de base est, dans la langue d'emprunt, un monosyllabique correspondant plus précisément à une seule syllabe lourde. En japonais, la plupart des mots anglais de structure /C₁V₁C₂/ sont adaptés en emprunts de structure /C₁V₁.Q.C₂V_e/ (la deuxième voyelle, V_e, étant épenthétique). Par exemple : 'shop' > *sho.Q.pu* ; 'pack' > *pa.Q.ku* ; 'book' > *bu.Q.ku*. On peut constater que les mots ayant cette structure en anglais deviennent quasi-systématiquement des emprunts trimoraïques en japonais. Lorsque ces mots sont utilisés comme deuxième constituant d'un CAL, le pattern de Suivi leur est souvent appliqué, ce qui revient à les conserver entièrement dans l'abréviation, comme dans les exemples ci-dessous.

'television shop'	>	<i>te.re.bi.jo.N sho.Q.pu</i>	>	<i>te.re.sho.Q.pu</i>
'plastic pack'	>	<i>pu.ra.su.chi.Q.ku pa.Q.ku</i>	>	<i>pu.ra.pa.Q.ku</i>
'magazine book'	>	<i>ma.ga.ji.N bu.Q.ku</i>	>	<i>ma.ga.bu.Q.ku</i>

Nous pourrions envisager que l'abréviation qui en résulte n'est pas un CAL, mais plutôt que le premier constituant est une abréviation de mot d'emprunt simple (c.-à-d. non composé) suivi d'un autre mot d'emprunt non abrégé. Toutefois, les abréviations de mots d'emprunt simples ne sont pas dérivées selon les mêmes principes que celles des mots d'emprunt composés. Le fait que le premier constituant soit abrégé selon le processus de formation des CAL encourage à interpréter le tout comme un CAL. Par ailleurs, ces abréviations de cinq mores sont accentuées comme l'ensemble des mots composés en japonais, c.-à-d. sur la première more du second constituant (Labrune, 2007).

Comme nous l'avons évoqué précédemment, Labrune (2007) avait déjà identifiés ce traitement particulier des constituants trimoraïques. Cela nous invite à réexaminer également les différents contextes phonologiques identifiés par Labrune qui, bien que n'ayant pas un effet aussi crucial que le présence de /R/ ou /Q/ en deuxième more de C₂, pourrait avoir une influence significative sur la formation des CAL. Revoyons tout d'abord plus en détails le traitement des

⁷ Les deux cas considérés comme des Maintiens de la more /Q/ correspondent ici aux deux exemples discutables (présentés en section 2.2) dans lesquels nous pourrions considérer que /Q/ adopte une prononciation non conventionnelle : 'no good' > *no.R gu.Q.do* > *no.R gu.u* ; et 'world cup' > *wa.R.ru.do ka.Q.pu* > *wa.R ka.tsu*.

C2 trimoraïques, quelle que soit la nature de leur seconde more. Labrune remarquait que pour ceux-ci, le pattern de Suivi était employé plus fréquemment. Grâce à notre réanalyse des lexiques de CAL, nous pouvons vérifier si cela s'avère correct en termes de fréquences de types dans les abréviations existantes. La figure 4 ci-dessous présente de quelle manière la deuxième more du deuxième constituant est traitée en fonction de si ce dernier est trimoraïque ou non. Notons que les patterns de Suivi et de Remplacement ne peuvent pas être utilisés sur des constituants de deux mores. Ces constituants, au nombre de 36 (soit 5% du total), ont donc été exclus de cette analyse. On peut observer que le nombre d'utilisations du pattern de Maintien est bien inférieur pour les deuxièmes constituants trimoraïques. En revanche, l'utilisation des patterns de Suivi et de Suppression est plus importante pour ces constituants.

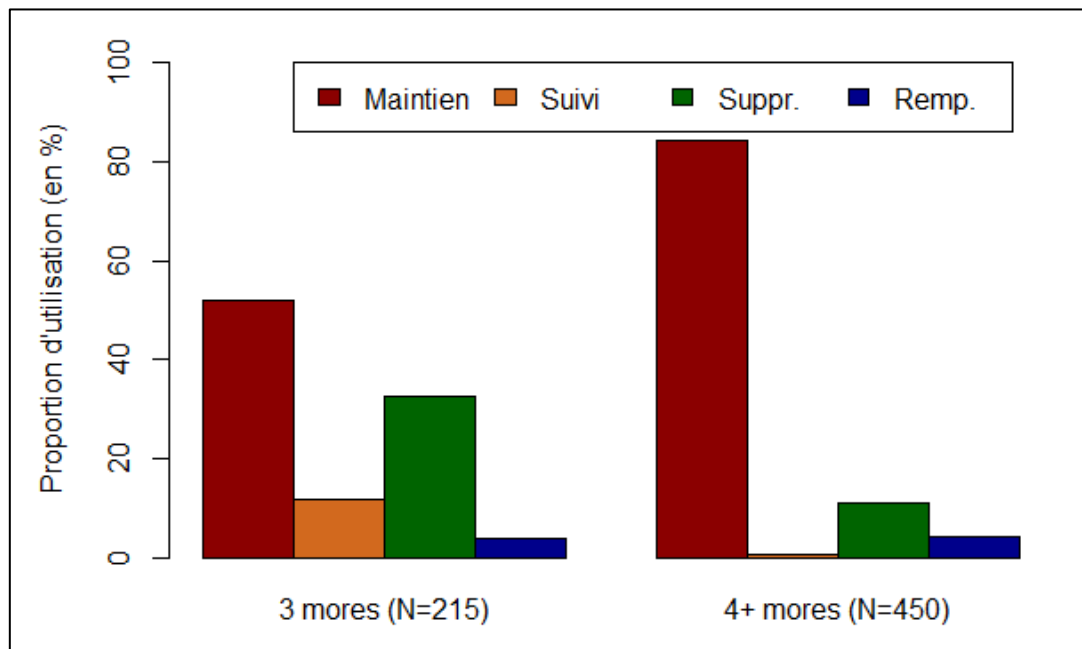


Fig.4 : traitement de la deuxième more du second constituant en fonction du nombre de more du constituant.

Un autre facteur que Labrune (2007) identifie comme influant sur la formation des CAL est la présence d'une consonne bilabiale à l'initiale du deuxième constituant. La figure 5 ci-dessous illustre de quelle manière est traitée la seconde more du C2 en fonction de la nature de sa consonne initiale (bilabiale vs non bilabiale). On constate que le nombre d'utilisations du pattern de Maintien est plus faible pour les C2 avec une consonne initiale bilabiale. Ceci est compensé par un emploi plus important du pattern de Suppression.

Un modèle linéaire généralisé a été appliqué aux données des 665 C2 de plus de deux mores avec : le type de pattern utilisé (régulier vs irrégulier) comme variable dépendante binaire ; et la longueur du constituant (3 mores vs 4+ mores) et le lieu d'articulation de sa consonne initiale (bilabiale vs non bilabiale) comme prédicteurs⁸. L'analyse de ce modèle a révélé un effet significatif de la longueur du constituant, en cela que les C2 trimoraïques sont plus souvent irréguliers ($z=-5.496$; $p<0.001$) ; ainsi que du lieu d'articulation de la consonne initiale, en cela que les C2 débutant par une consonne bilabiale sont plus souvent irréguliers ($z=-3.87$; $p<0.001$) ; mais pas d'effet significatif de l'interaction entre ces deux prédicteurs ($z=-0.159$; $p=0.87$).

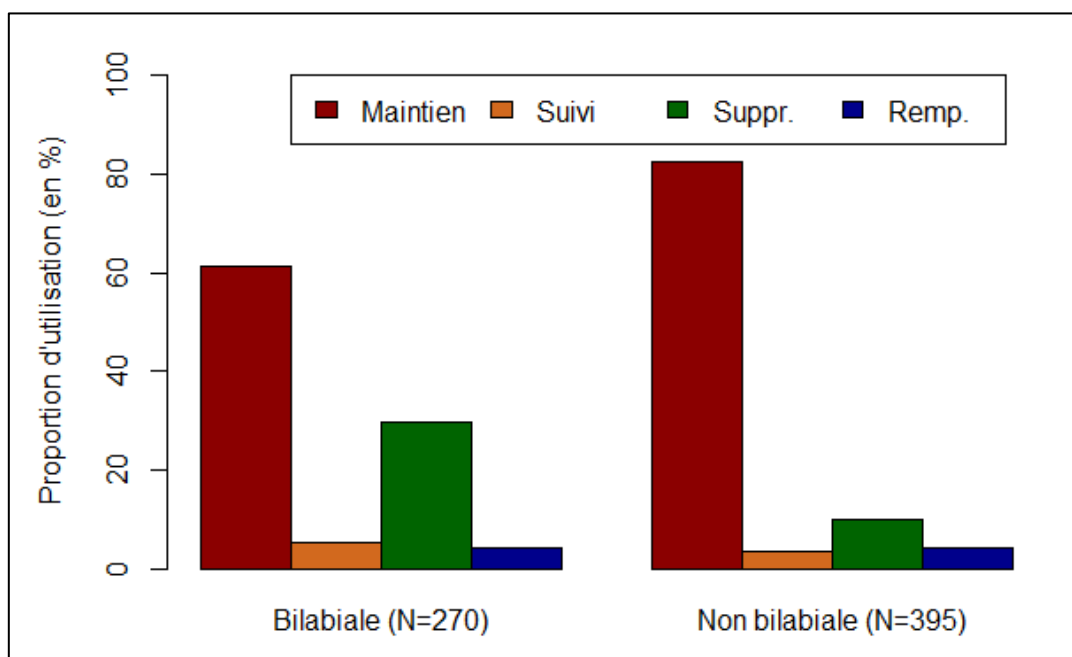


Fig.5 : traitement de la deuxième more du second constituant en fonction de la nature de la consonne initiale du constituant.

Enfin, le troisième facteur de variabilité identifié par Labrune (2007) est celui de la hauteur de la voyelle pour les deuxièmes constituants dont /R/ est la seconde more. La figure 6 ci-dessous présente de quelle manière est traitée la more /R/ en tant que deuxième more de C2, en fonction de la hauteur de la voyelle allongée dans les 113 CAL concernés. On observe que le pattern de Maintien est moins souvent employé sur /R/ quand il représente l'allongement

⁸ Toutes les analyses statistiques de cette thèse ont été réalisées sur R (version 3.5.2, R Core Team [2018]).

d'une voyelle non haute. Au contraire, le nombre d'utilisations du pattern de Suppression est plus important dans ces cas-là.

Étant donné que le facteur de la hauteur de la voyelle allongée n'est pertinent que pour les CAL /R/, nous avons testé son effet dans une analyse statistique séparée sur ce type de CAL uniquement. Pour les mêmes raisons que précédemment, nous avons retiré 9 CAL de deux mores (soit 8% du total) de notre analyse. Un modèle linéaire généralisé a été appliqué aux données des 104 C2 dont la deuxième more est /R/ avec : le type de pattern utilisé (régulier vs irrégulier) comme variable dépendante binaire ; et la hauteur de la voyelle allongée (haute vs non haute), la longueur du constituant (3 mores vs 4+ mores) et le lieu d'articulation de sa consonne initiale (bilabiale vs non bilabiale) comme prédicteurs. Des tests du rapport de vraisemblance (log likelihood ratio tests) entre différents modèles ont été effectués pour évaluer la significativité des effets des différents facteurs. Des comparaisons de modèles ont révélé des effets marginalement significatifs de la hauteur de la voyelle ($\chi^2=-3.4947$; $p=0.06$) et du lieu d'articulation de la consonne initiale ($\chi^2=-3.5395$; $p=0.06$), mais pas d'effet significatif de l'interaction de ces deux prédicteurs ($\chi^2=0.14328$; $p=0.705$) ou de la longueur du constituant ($\chi^2=0.87732$; $p=0.3489$).

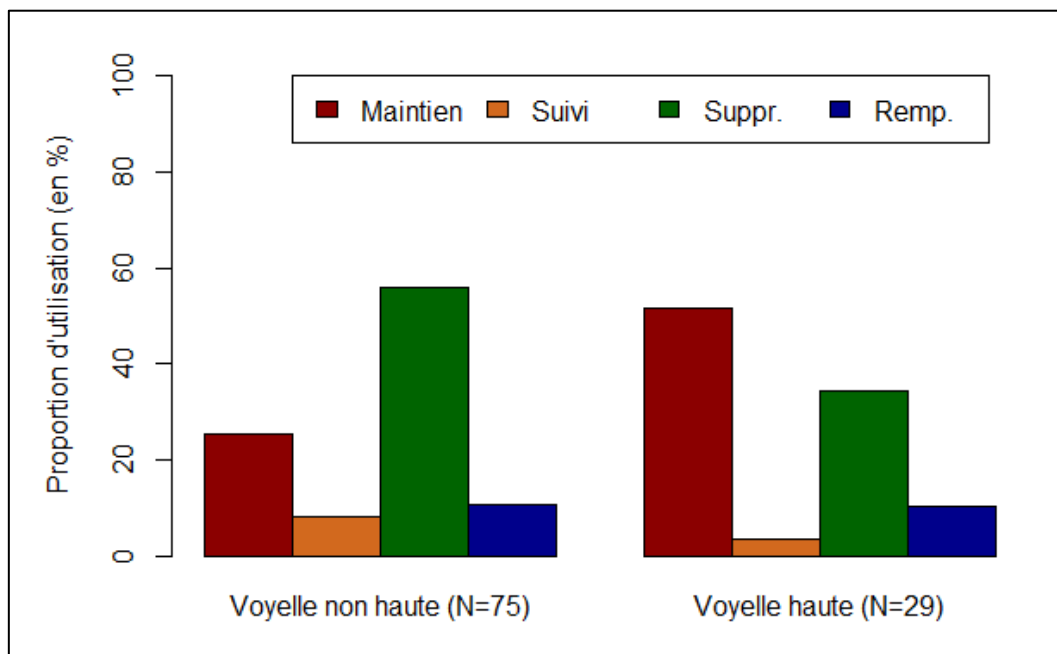


Fig.6 : traitement de la more /R/ en tant que deuxième more du second constituant en fonction de la hauteur de la voyelle allongée.

3.3 Discussion

3.3.1 Remarques générales

Tout en fournissant des données précises sur la fréquence de types des différents patterns d'abréviation, la réanalyse du lexique de CAL présente des résultats qui viennent confirmer les observations des précédentes études sur le sujet (Labrune, 2006b, 2007, 2008 ; Lamarque, 2015, 2016 ; Lamarque & German, 2017). Ces résultats montrent que la formation des CAL est gouvernée par un principe général qui entraîne la création d'abréviations d'une longueur de quatre mores qui correspondent aux deux mores initiales de chacun des constituants. Ce principe général correspond au pattern de Maintien dans les figures présentées dans la section précédente. On observe, notamment dans la figure 3, que lorsqu'il n'entre pas en conflit avec les contraintes phonologiques de la langue (c.-à-d. pour les C2 dont la deuxième more est *sans conflit*), le principe général de formation est utilisé dans la grande majorité des cas (88%). Même si un peu de variation subsiste en l'absence de conflit avec des contraintes phonologiques, les CAL canoniques suivant le principe général de formation sans enfreindre de contraintes phonologiques sont les formes les plus communes (68% de toutes les abréviations).

Une seconde observation confirmée par cette réanalyse est que les contraintes phonologiques portant sur l'apparition de /R/ ou /Q/ en position finale de mot jouent un rôle important dans la création de CAL irréguliers. Concernant les C2 dont la deuxième more est /R/, il semble que les patterns de Maintien et Suppression soient en compétition. Cela peut s'expliquer par le fait que la contrainte défavorisant l'apparition de /R/ en position finale est relativement faible. En effet, la more /R/ n'apparaît jamais en position finale dans la strate Native, mais elle apparaît fréquemment en position finale dans les mots des strates Sino-Japanese, Onomatopoeia, et Foreign (voir section 2.2). Cela rend son conflit avec le principe général de formation plus difficile à résoudre pour les locuteurs. Cependant, la balance penche légèrement en faveur du pattern de Suppression. En ce qui concerne les C2 dont la deuxième more est /Q/, le pattern de Maintien est pratiquement inapplicable, ce qui entraîne inévitablement la création de CAL irréguliers.

La troisième observation que la réanalyse du lexique vient confirmer est celle portant sur les facteurs secondaires pouvant influencer sur la formation de CAL irréguliers. En effet, les

C2 trimoraïques, la présence d'une consonne bilabiale à l'initiale du C2, ou les C2 dont la deuxième more est l'allongement (/R/) d'une voyelle non haute sont autant de facteurs qui augmentent la probabilité qu'une abréviation irrégulière soit créée.

Notons que le pattern utilisé pour former la grande majorité de ces CAL irréguliers est le pattern de Suppression, et ce dans tous les contextes phonologiques, même dans les cas où la deuxième more du C2 est *sans conflit*. On remarque tout de même que les C2 trimoraïques semblent encourager l'utilisation du pattern de Suivi, bien que le pattern de Suppression reste dominant parmi les formes irrégulières dans ce contexte.

3.3.2 Description en OT standard

Concernant l'analyse en OT du conflit entre le principe général de formation des CAL et les contraintes phonologiques générales du japonais, les résultats de cette réanalyse nous permettent d'établir une hiérarchie entre les différentes contraintes impliquées. Rappelons que nous avons effectué dans de précédents travaux une analyse en OT détaillée de la formation des CAL (Lamarque & German, 2017) que nous synthétisons et adaptons ici aux données de la réanalyse du lexique. Les tableaux 1 à 3 ci-dessous illustrent de quelle manière les contraintes impliquées dans le traitement de /R/ peuvent être hiérarchisées.

Tableau 1 : *V:# >> *PIED<2 ; exemple 'family mart'.

/ɸamiri maRto/	*V:#	*PIED<2
☐ (ɸa.mi)(ma)		*
(ɸa.mi)(ma.R)	*!	

Tableau 2 : CONTIG >> *V:# ; exemple 'cheese burger'.

/t̃eiRzu baRgaR/	CONTIG	*V:#
(t̃ei.R)(ba.ga)	*!	
☐ (t̃ei.R)(ba.R)		*

Tableau 3 : *PIED>2 >> *V:# ; exemple ‘cheese burger’.

/t̄eiRzu baRgaR/	*PIED>2	*V:#
(t̄ei.R)(ba.R.ga)	* !	
▣ (t̄ei.R)(ba.R)		*

Dans le cas des CAL dont la deuxième more du C2 est /R/, bien qu’il y ait une certaine compétition entre le pattern de Suppression et de Maintien, le fait que le pattern de Suppression soit plus souvent employé nous permet de conclure que la contrainte *V:# domine *PIED<2 (cf. tableau 1). Puis le fait que le pattern de Maintien soit plus fréquent que le pattern de Remplacement et de Suivi nous permet de placer *V:# en dessous de CONTIG (cf. tableau 2) et *PIED>2 (cf. tableau 3) respectivement. Les données ne permettent de déterminer clairement s’il existe une différence hiérarchique entre CONTIG et *PIED>2. En revanche, au vu des données, et par transitivité des observations formulées sur la base des tableaux 1 à 3, on peut conclure que *PIED<2 est dominée par CONTIG et *PIED>2 (soit {CONTIG, *PIED>2} >> *PIED<2). Concernant le traitement de /Q/, les tableaux 4 et 5 ci-dessous nous renseignent sur la position de LICENCE/Q/ dans la hiérarchie.

Tableau 4 : LICENCE/Q/ >> CONTIG ; exemple ‘Harry Potter’.

/hariR poQtaR/	LICENCE/Q/	*CONTIG
(ha.ri)(po.Q)	* !	
▣ (ha.ri)(po.ta)		*

Tableau 5 : LICENCE/Q/ >> *PIED<2 ; exemple ‘plastic pack’.

/purasut̄eiQku paQku/	LICENCE/Q/	*PIED<2
(pu.ra)(pa.Q)	* !	
▣ (pu.ra)(pa.Q.ku)		*

Le fait que le pattern de Maintien soit employé dans pratiquement aucun CAL dont la deuxième more du C2 est /Q/ alors que l’on voit parfois apparaître le pattern de Remplacement et de Suivi nous permet de conclure que LICENCE/Q/ domine aussi bien CONTIG (cf. tableau 4)

que *PIED>2 (cf. tableau 5). Comme la hiérarchie entre ces deux dernières contraintes est indéterminée, on peut noter que LICENCE/Q/ >> {CONTIG, *PIED>2}. Au vu des données et par transitivité, on sait également que LICENCE/Q/ domine toutes les autres contraintes. La hiérarchie de toutes les contraintes impliquées dans le conflit entre le principe général de formation des CAL et la phonologie du japonais est donc la suivante :

$$\text{LICENCE/Q/} \gg \{\text{CONTIG, *PIED>2}\} \gg *V:\# \gg *PIED<2$$

Si l'on observe l'ensemble des données, cette hiérarchie reflète assez bien la fréquence de types des différents patterns. La contrainte LICENCE/Q/ étant pratiquement inviolable, avec seulement deux cas discutables de Maintien de /Q/, elle se situe au sommet de la hiérarchie. Puis avec seulement 27 et 28 cas d'infractions, viennent les contraintes CONTIG et *PIED>2 qui correspondent respectivement aux patterns de Remplacement et de Suivi. La contrainte phonologique *V:# prévenant l'apparition de /R/ en position finale est plutôt faible, puisqu'elle est enfreinte dans 41 des 113 CAL concernés. Enfin, en bas de la hiérarchie se trouve *PIED<2 qui correspond au pattern de Suppression, représentant le type de CAL irréguliers de loin le plus commun.

3.3.3 Description en OT stochastique

Afin de rendre compte de la variabilité dans le processus de formation des CAL, la hiérarchie que nous venons d'établir ne devrait pas être perçue comme absolue, conformément à l'OT standard, mais plutôt comme une indication de la position et de l'étendue de chaque contrainte sur une échelle de *strictness*, comme c'est le cas en OT stochastique (Smolensky & Legendre, 2006). En effet, en l'état, cette hiérarchie de contraintes ne peut s'appliquer à la formation d'un certain nombre de CAL irréguliers tels que ceux suivant le pattern de Remplacement ou de Suivi, ou les cas dans lesquels /R/ est Maintenu. Le tableau 6 ci-dessous illustre de quelle manière cette hiérarchie échoue à sélectionner l'output attesté dans le lexique pour le cas d'un maintien de /R/ final. Le symbole « ■ » indique l'output sélectionné par la hiérarchie de contraintes, tandis que le candidat en gras est l'output effectivement attesté dans le lexique.

Tableau 6 : exemple d'un CAL irrégulier ('cheese burger') pour lequel la hiérarchie de contraintes sélectionnerait un output incorrect.

$\widehat{t\acute{e}iRzu\ baRgaR/}$	LICENCE/Q/	CONTIG	*PIED>2	*V:#	*PIED<2
$(\widehat{t\acute{e}i.R})(\mathbf{ba.R})$				*!	
$(\widehat{t\acute{e}i.R})(\mathbf{ba.R.g\acute{a}})$			*!		
$\blacksquare (\widehat{t\acute{e}i.R})(\mathbf{ba})$					*
$(\widehat{t\acute{e}i.R})(\mathbf{ba.g\acute{a}})$		*!			

Dans cet exemple, on peut constater que la hiérarchie de contraintes sélectionne le troisième candidat qui n'enfreint que *PIED<2, la contrainte la plus faible. Elle rejette ainsi l'output attesté en raison de son infraction à *V:#, une contrainte de rang supérieur. L'OT stochastique propose une solution pour éviter ce type d'incohérences en représentant chaque contrainte comme une distribution gaussienne sur une échelle de *strictness*, plutôt que par un rang absolu dans une hiérarchie. À chaque évaluation d'un candidat, la *strictness* effective de chaque contrainte est déterminée de manière stochastique, la probabilité d'obtenir n'importe quelle valeur pour chaque contrainte étant représentée par sa distribution gaussienne. La variabilité apparaît lorsque la distribution de plusieurs contraintes se chevauchent, permettant des inversions ponctuelles dans l'ordre hiérarchique des contraintes correspondantes, en vertu de la nature stochastique de la valeur de leur *strictness* à chaque évaluation. Les contraintes impliquées dans la formation des CAL peuvent donc être représentées sur une échelle de *strictness* comme sur la figure 7 ci-dessous. Il s'agit ici d'une simple représentation schématique visant à reproduire de manière approximative les données du lexique, nous étudierons en section 4.3.2 comment modéliser plus concrètement une grammaire en OT stochastique. Notons que la distribution de la contrainte quasi-inviolable LICENCE/Q/ est principalement située en dehors du bord gauche de la figure, avec très peu de chevauchement avec les autres contraintes.

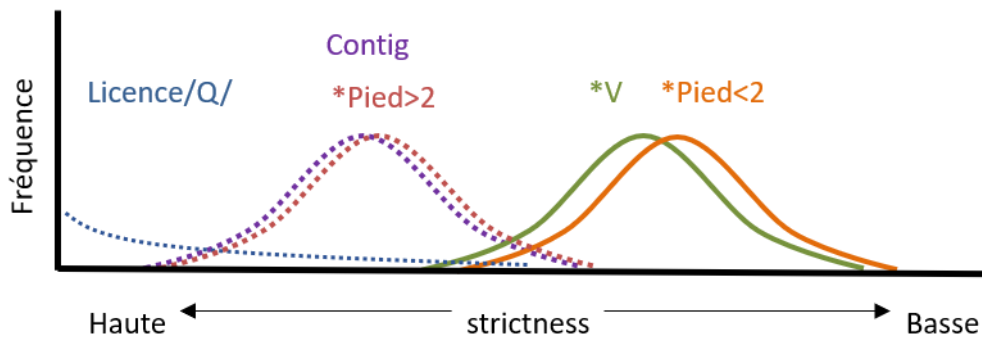


Fig.7 : position des contraintes impliquées dans la formation des CAL sur une échelle de *strictness*, selon les données du lexique.

Avec cette représentation de l'organisation des contraintes impliquées dans la formation des CAL, on peut désormais rendre compte de cas tels que 'cheese burger' présenté ci-dessus. Étant donné le fort chevauchement entre *V:# et *PIED<2, si une valeur aléatoire est sélectionnée sous la courbe qui les représente pour déterminer ponctuellement leur *strictness*, il est très probable que lors d'une évaluation la *strictness* effective de *PIED<2 soit supérieure à celle de *V:#. Auquel cas, l'ordre hiérarchique de ces deux contraintes sera ponctuellement inversé et un output présentant un Maintien de la more /R/ pourra être sélectionné. Le tableau 7 illustre la sélection de l'output de 'cheese burger' lorsque la *strictness* effective des contraintes pour son évaluation les place dans l'ordre hiérarchique ponctuel suivant :

LICENCE/Q/ >> CONTIG >> *PIED>2 >> *PIED<2 >> *V:#.

Tableau 7: sélection de l'output pour 'cheese burger' dans le cas où la hiérarchie est ponctuellement : LICENCE/Q/ >> CONTIG >> *PIED>2 >> *PIED<2 >> *V:#.

$\widehat{t\epsilon i}Rzu\ baRgaR/$	LICENCE/Q/	CONTIG	*PIED>2	*PIED<	*V:#
$\widehat{t\epsilon i.R}(ba.R)$					*
$\widehat{t\epsilon i.R}(ba.R.ga)$			*!		
$\widehat{t\epsilon i.R}(ba)$				*!	
$\widehat{t\epsilon i.R}(ba.ga)$		*!			

3.3.4 Prédications concernant la productivité des patterns en fonction des fréquences de types

Même si ces résultats nous éclairent sur l'utilisation récurrente de différents patterns dans les CAL existants, ils soulignent également la variabilité à laquelle sont sujettes ces abréviations. Ce processus morphologique étant très productif en japonais, dans un contexte où la culture populaire emploie une grande quantité de mots d'emprunt à l'anglais, certains locuteurs au moins doivent régulièrement faire face au problème de la création de nouvelles formes. Il reste cependant à déterminer si les locuteurs ont assimilé l'utilisation de ces différents patterns, et à quel point leur comportement est régulier dans ce processus de création. En théorie, la fréquence de types des différents patterns devrait être une forte indication sur leur productivité (voir section 1.3.1). Puisque nous connaissons désormais la fréquence de types des différents patterns, voyons quelles sont les différentes prédictions que l'on peut formuler sur la base de ces données.

Si l'on regarde les résultats dans leur globalité, il apparaît que le pattern de Maintien est celui qui a la plus haute fréquence de types, largement au-dessus de tous les autres. En toute logique, la productivité de ce pattern devrait également être la plus élevée. Cependant, il existe un contexte phonologique dans lequel ce pattern n'est pas acceptable : lorsque la deuxième more du deuxième constituant est /Q/. En d'autres termes, il n'existe aucun type présentant ce pattern dans ce contexte, à l'exception de deux exemples discutables. Par conséquent, on peut s'attendre à ce que le pattern de Maintien n'ait aucune productivité dans ce contexte phonologique. Néanmoins, nous devons également considérer que Kapatsinski (2010) présente un cas dans lequel des nouvelles formes sans le moindre type structurellement équivalent parmi les formes existantes (en l'occurrence, une consonne vélaire finale non palatalisée suivie de *-i*) peuvent apparaître malgré tout, si un pattern plus général structurellement similaire était largement plus fréquent (en l'occurrence, une consonne finale, peu importe son lieu d'articulation, non palatalisée suivie de *-i*). Cependant, nous pourrions considérer que la contrainte limitant la présence de /Q/ en position finale d'un CAL est bien plus sévère que celle qui pèse sur les consonnes vélaire finales non palatalisées suivies de *-i* en russe. En effet, la première empêche l'apparition d'un segment entièrement non spécifié d'émerger dans une forme de surface, tandis que la seconde favorise l'altération du lieu d'articulation d'une consonne. Par conséquent, il serait fortement inattendu, malgré la sur-représentation du pattern

de Maintien dans les fréquences de types totales, que le segment non spécifié /Q/ soit rendu acceptable en position finale. En d'autres termes, notre prédiction est que le pattern de Maintien ne peut pas éclipser les patterns irréguliers dans ce contexte phonologique en raison de la quasi-inviolabilité de la contrainte qui lui est associé. Puisque parmi ceux-ci le pattern de Suppression est celui qui possède la plus haute fréquence de types dans le lexique, nous prédisons également qu'il sera productif dans la création de nouvelles formes dans ce contexte.

La situation est plus complexe pour le cas de la more /R/ en tant que deuxième more du deuxième constituant. On pourrait envisager que les patterns de Maintien et de Suppression seraient tous les deux productifs, puisque tous les deux ont une haute fréquence de types dans ce contexte phonologique. Cependant, cela présuppose que les locuteurs japonais traitent les CAL présentant ce contexte différemment de ceux présentant une more *sans conflit* en deuxième more de C2. Si on peut affirmer cela dans le cas de la more /Q/, puisqu'une contrainte quasiment inviolable de la langue empêche les locuteurs de la traiter comme n'importe quelle autre more, cela n'est pas aussi clair dans le cas de /R/. Dans le contexte de /R/ en deuxième more du C2, la situation pourrait être davantage similaire à celle présentée par Kapatsinski (2010), où un pattern général éclipse un pattern spécifique à un contexte phonologique. En effet, le pattern de Maintien, correspondant au principe général de formation utilisé pour 74% de toutes les abréviations (517 types), pourrait éclipser le pattern de Suppression appliqué à un contexte phonologique particulier.

Notons que dans ce processus morphologique où il existe deux niveaux d'inadéquation du pattern général, il est difficile d'évaluer si les deux contextes problématiques (c.-à-d. la more /R/ ou /Q/ en deuxième more de C2) contribuent à accroître la fréquence du même pattern contextuel. Est-ce que les hautes fréquences de types du pattern de Suppression dans le traitement de la more /R/ d'une part, et /Q/ d'autre part contribuent ensemble à faire de ce pattern celui par défaut lorsque le pattern général est inadéquat ? Dans ce cas, ce pattern aurait une fréquence de types assez importante puisqu'il est utilisé dans 18% des cas (129 types, si on compte également les mores *sans conflit* qui sont supprimées). Ou bien, est-ce que la probabilité d'occurrence des patterns contextuels varie en fonction du contexte phonologique ? Dans ce cas, le pattern de Suppression de la more /R/ compte pour seulement 8% des abréviations (54 types).

Quoi qu'il en soit, il est difficile de prédire l'issue de la compétition entre les patterns de Maintien et de Suppression dans le traitement de la more /R/ dans la création de nouvelles formes. Par ailleurs, il est également probable que la compétition entre ces deux patterns soit gérée différemment d'un locuteur à l'autre. Par conséquent, nos prédictions vis-à-vis du traitement de /R/ resteront aussi générales que possibles : les patterns de Maintien et/ou de Suppression pourraient être productifs pour le traitement de la more /R/. On peut s'attendre à ce qu'un de ces patterns domine l'autre, ou bien qu'ils soient proches d'une probabilité de 50-50. Dans tous les cas, les patterns de Suivi et de Remplacement ne devraient être que peu productifs.

Avant de nous tourner vers la tâche de création de nouvelles formes nous permettant de tester si les locuteurs japonais ont assimilé les patterns de formation des CAL, voyons quelles seraient les implications de la confirmation ou de la contradiction des prédictions formulées dans cette section. La comparaison des résultats de cette tâche avec les prédictions basées sur les fréquences de types dans le lexique nous éclairera sur les éventuelles limites des explications en termes de probability-matching de la productivité des patterns morphologiques :

- Pour le traitement des mores *sans conflit* : le pattern de Maintien devrait être le plus productif. Les autres patterns ne devraient avoir qu'une faible productivité. Puisque le pattern de Maintien a une très haute fréquence, il serait inattendu que cette prédiction soit contredite car cela remettrait en cause l'existence d'un quelconque lien entre les fréquences de types et la productivité, et d'une manière plus générale, la nature probabiliste de l'apprentissage linguistique.
- Pour le traitement de la more /Q/ : le pattern de Maintien devrait être quasiment improductif, et le pattern de Suppression devrait être le plus productif. La confirmation de l'improductivité du pattern de Maintien suggérerait que tous les patterns généraux ne peuvent pas éclipser des patterns contextuels moins fréquents et que la *strictness* des contraintes qui s'appliquent sur chaque pattern joue un rôle primordial dans l'issue de leur compétition. Un tel résultat sera tout de même cohérent avec un comportement de probability-matching puisque la probabilité de /Q/ en position finale est quasi-nulle en japonais. Cela ne serait pour autant pas trivial, car Kapatsinski (2010) présente un cas similaire qui illustre pourtant un comportement déviant du probability-matching. Concernant le pattern de Suppression, une confirmation de sa productivité indiquerait

que les locuteurs traitent les formes irrégulières également en se basant sur les fréquences de types, donc selon un mécanisme de probability-matching.

- Pour le traitement de la more /R/ : les patterns de Maintien et de Suppression pourraient tous les deux être productifs. Qu'un des deux patterns domine l'autre, ou qu'ils aient une productivité équivalente, ils devraient dominer les patterns de Suivi et de Remplacement. Si le pattern de Maintien s'avère dominant, cela soutiendrait l'idée que les patterns généraux très fréquents peuvent éclipser des patterns contextuels, comme le montrait Kapatsinski (2010), et cela irait à l'encontre du probability-matching. Au contraire, si le pattern de Suppression domine, on conclura que cette conception du rapport entre pattern général et patterns contextuels ne s'applique pas au cas des CAL. Puisque la productivité du pattern de Suppression serait le reflet de sa fréquence de types, cela soutiendrait l'idée d'un mécanisme de probability-matching. On pourrait également conclure que plus il est dominant, plus convaincante est l'idée que le nombre d'utilisations du pattern de Suppression en général, pas uniquement pour la more /R/, renforce la probabilité qu'il soit utilisé comme pattern irrégulier par défaut.

Chapitre 4 – Expérience : tâche de création de nouveaux CAL

Afin de tester les hypothèses présentées dans la section précédente, nous avons mené une expérience dans laquelle il a été demandé à des locuteurs de créer de nouvelles abréviations. Les données collectées dans cette expérience pourront être comparées avec celles sur les fréquences de types obtenues grâce à la réanalyse du lexique de CAL. Ainsi, il est possible d'évaluer à quel point les fréquences de types permettent d'expliquer le comportement des locuteurs dans la création de nouvelles formes, et de manière plus intéressante, d'observer dans quelles conditions elles sont insuffisantes pour expliquer ces comportements.

4.1 Méthode

4.1.1 Matériel

Dans cette expérience, nous avons présenté aux participants des paires de mots d'emprunt (c-à-d. un C1 et un C2) pouvant servir de base à la formation d'un CAL. Un item est donc défini comme la combinaison d'un C1 et d'un C2. Une liste de 120 C1 (disponible en annexe 5) et deux listes de 120 C2 (annexes 6 et 7) ont été utilisées comme base pour créer les items de cette expérience. L'existence de tous les mots d'emprunt dans ces listes a été vérifiée en les cherchant dans le corpus *chūnagon* (alias le BCCWJ : Balanced Corpus of Contemporary Written Japanese). Ce corpus créé par le National Institute for Japanese Language and Linguistics contient approximativement 104.3 millions de mots issus de sources écrites diverses datant de 1976 à 2006 : livres, magazines, journaux, rapports d'activité, forum internet, blogs, manuels, documents légaux, etc. (Maekawa et al., 2014). Les mots d'emprunts servant de C2 ont été sélectionnés pour répondre à un certain nombre de critères phonologiques nous permettant de tester, entre autres, l'influence de la nature de la deuxième du C2 sur la formation des CAL. Les deux listes de C2 étaient donc structurées de la même manière et contenaient :

- 48 C2 *sans conflit* : des mots dont la deuxième more est soit de structure $/(C)(j)V/$, soit la nasale homorganique $/N/$.

- 48 C2 /R/ : des mots dont la deuxième more est /R/.
- 24 C2 /Q/ : des mots dont la deuxième more est /Q/.

Au sein de chaque type de C2, des sets de C2 avec des différences phonologiques minimales ont été créés afin de contrôler l'influence des autres caractéristiques phonologiques pouvant jouer sur la formation des CAL : la présence d'une consonne bilabiale en initiale ; la longueur du C2 ; et la hauteur de la voyelle pour les C2 /R/.

Pour les C2 /R/, il y a donc trois caractéristiques phonologiques qui peuvent influencer sur la manière de créer de nouvelles formes. Cependant, il n'était pas possible de trouver dans le corpus *chūnagon* assez de sets de huit mots avec des différences phonologiques minimales couvrant toutes les combinaisons possibles de présence ou non de ces différentes caractéristiques. Par ailleurs, cela aurait créé un déséquilibre important au niveau du nombre d'items de chaque type dans les listes. Nous avons donc choisi de nous intéresser principalement à l'effet de la présence d'une consonne bilabiale en initiale et de la hauteur de la voyelle allongée pour le cas des C2 /R/. Les sets de C2 /R/ contiennent donc :

- 1) Un C2 débutant par une **consonne non bilabiale**, suivie d'une **voyelle non haute**.
- 2) Un C2 débutant par une **consonne non bilabiale**, suivie d'une **voyelle haute**.
- 3) Un C2 débutant par une **consonne bilabiale**, suivie d'une **voyelle non haute**.
- 4) Un C2 débutant par une **consonne bilabiale**, suivie d'une **voyelle haute**.

Voici un exemple de set de C2 /R/ :

Liste 1, set 38 :

- 1) *ka.R.to* 'kart'
- 2) *shi.R.to* 'seat'
- 3) *pa.R.to* 'part'
- 4) *bi.R.to* 'beat'

Pour les C2 /Q/, il y a deux caractéristiques à considérer : la longueur du C2 ; et la nature de la consonne initiale. Cependant, encore une fois le corpus *chūnagon* ne contenait que quelques sets de quatre mots d'emprunt pouvant couvrir les quatre combinaisons possibles de ces caractéristiques tout en étant identiques par ailleurs. Comme la longueur du C2 semble être particulièrement influente pour le traitement des C2 /Q/, c'est cette caractéristique que nous avons

préfééré retenir. Par conséquent, les C2 /Q/ sont regroupés en douze sets de deux mots dont l'un est trimoraïque et l'autre est plus long, tout en débutant par les trois mêmes mores.

Voici un exemple de set de C2 /Q/ :

Liste 1, set 25:

- 1) *ka.Q.pu* 'cup'
- 2) *ka.Q.pu.ru* 'couple'

Enfin, pour les C2 *sans conflit*, il y a également deux caractéristiques à contrôler : la longueur du C2 ; et la nature de la consonne initiale. Il était une fois de plus impossible de trouver dans le corpus *chūnagon* des sets complets pour satisfaire toutes les combinaisons de caractéristiques. Non avons donc construit des sets de trois C2 en excluant la combinaison la plus rare : les C2 trimoraïque débutant par une consonne bilabiale. Les sets de C2 *sans conflit* contiennent donc :

- 1) Un C2 de **plus de trois mores** avec une **consonne initiale non bilabiale**.
- 2) Un C2 de **trois mores** avec une **consonne initiale non bilabiale**.
- 3) Un C2 de **plus de trois mores** avec une **consonne initiale bilabiale**.

Voici un exemple de set de C2 *sans conflit*. Notons que pour trouver assez de sets, la more initiale ne pouvait pas être identique pour tous les mots du même set, mais elle respecte toujours les caractéristiques mentionnées ci-dessus :

Liste 1, set 13:

- 1) *su.ra.i.do* 'slide'
- 2) *do.ra.i* 'dry'
- 3) *bu.ra.i.do* 'bride'

Les douze derniers C2 *sans conflit* sont des mots ne possédant aucune des caractéristiques favorisant un traitement irrégulier. Ils sont donc structurellement similaires au premier mot des sets de trois, mais ils ne sont pas inclus dans un set de mots similaires. Ils ont été inclus afin d'obtenir des listes plus équilibrées en termes de types de deuxième more de C2, et éviter que les C2 /R/ et /Q/ y soit trop sur-représentés. On peut considérer ces douze C2 *sans conflit* comme douze sets d'un seul mot. Par exemple :

Liste 1, set 6:

1) *do.ra.go.N* ‘dragon’

Afin de former un item, chaque C2 doit être appairé à un C1. Pour limiter l’influence des C1 sur le traitement des items, seuls des mots ne contenant aucune caractéristique phonologique influant sur la formation des CAL ont été sélectionnés comme C1. Ces caractéristiques phonologiques, identifiées dans les précédentes études sur les CAL, sont très similaires à celles qui influencent le traitement des C2. Cependant, le traitement irrégulier d’un C1 est bien moins fréquent (Labrune, 2007). Par conséquent, tous les C1 sont des mots d’emprunt d’une longueur de plus de trois mores, ne débutant pas par une consonne bilabiale, et dont la deuxième more n’est ni /R/, ni /Q/. De plus, les C1 ont également été regroupés en sets de tailles égales aux sets de C2 dans le but de limiter les paires possibles aux constituants possédant le même numéro de set. Le détail de la formation des paires de C1 et C2 est donné dans la section suivante.

4.1.2 Procédure

Dans une expérience menée en ligne, nous avons demandé à des locuteurs japonais de créer de nouvelles abréviations à partir des items décrits dans la section précédente. Pour chaque participant, une seule liste de C2 était sélectionnée aléatoirement pour être appairée avec la liste de C1 et former les items. Ainsi, chaque participant a créé 120 nouveaux CAL. Pour former les items, les C1 et C2 de sets correspondants était appairés de manière aléatoire pour chaque participant. Par exemple, pour tous les participants ayant été exposés à la première liste de C2, des paires aléatoires ont été créées à partir des C1 du set 38, en (1) ci-dessous, et des C2 du set 38, en (2) ci-dessous. Chaque constituant n’apparaissait qu’une seule fois par participant. La même procédure était appliquée à tous les sets afin de créer les 120 items présentés au participant.

(1) Liste de C1, set 38 :

ku.ra.i.a.N.to ‘client’

ji.re.N.ma ‘dilemma’

re.se.pu.sho.N ‘reception’

shi.N.bo.ru ‘symbol’

(2) Liste 1 de C2, set 38 :

ka.R.to ‘kart’

shi.R.to ‘seat’

pa.R.to ‘part’

bi.R.to ‘beat’

Nous avons vérifié que toutes les paires de mots possibles n'étaient pas des paires servant de base à un CAL existant dans les lexiques à notre disposition. Néanmoins, même s'ils n'étaient pas appairés ensemble dans un CAL existant, il est possible que les constituants apparaissent séparément dans différents CAL existants. Ceci était inévitable pour obtenir un nombre suffisant d'items avec les spécificités phonologiques mentionnées précédemment, puisqu'un grand nombre de mots d'emprunt reconnaissables apparaissent dans des CAL. Enfin, pour chaque participant, une fois la procédure de création des 120 items achevée, l'ordre des items était établi de manière aléatoire.

Lors de la présentation de chaque item aux participants, les C1 et C2 étaient écrits à la fois en *katakana* (un des deux syllabaires japonais, principalement utilisé pour transcrire les mots d'emprunt ou les idéophones) et en alphabet latin. Quatre possibilités d'abréviations pour ces deux mots, correspondant aux quatre patterns communs utilisés dans la formation des CAL, étaient également présentées, seulement en *katakana* (comme dans la figure ci-dessous).

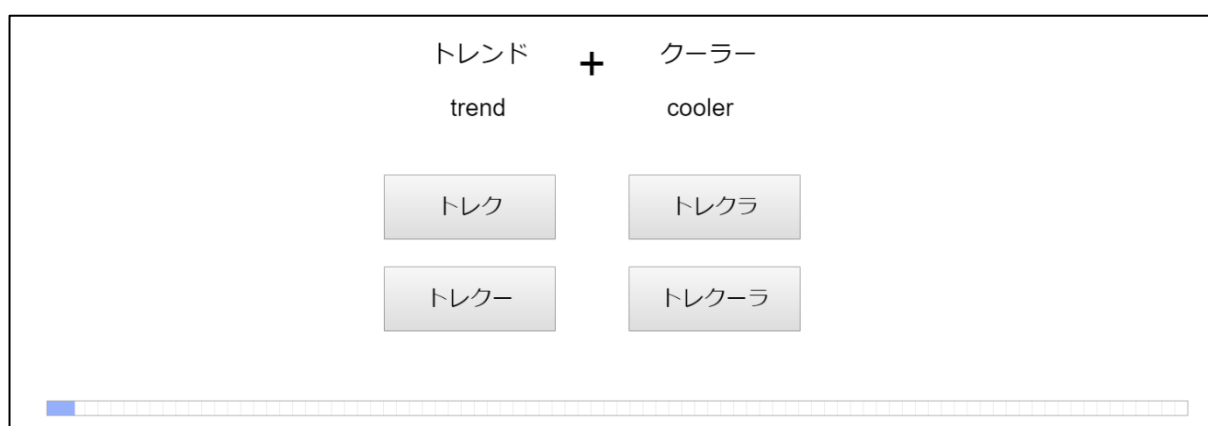


Fig.8 : affichage d'un item dans la tâche de création de nouvelles formes.

Le choix de quel pattern parmi les quatre afficherait chaque bouton était déterminé de manière aléatoire pour chaque item. Afin de repérer d'éventuels participants non coopératifs, qui pourraient appuyer sans cesse sur le même bouton, ou cliquer au hasard sur les boutons sans se concentrer sur la tâche, les temps de réponse ainsi que la position du bouton choisi étaient enregistrés pour chaque item. Les instructions invitaient les participants à ne pas se préoccuper du sens de l'abréviation mais plutôt de choisir l'abréviation qui leur semblait la plus naturelle selon leur intuition (voir annexe 1 pour les instructions complètes). Cette expérience a été programmée sur Construct 2, un logiciel de création de jeux (multi-joueurs) pour navigateur web en 2D (Scirra, 2020 ; release 279).

4.1.3 Participants

Les participants, 136 au total, ont été recrutés via une plateforme de crowdsourcing japonaise : Lancers (lancers.jp). Avec ce mode de recrutement, les participants restent anonymes. Cependant, ils sont identifiables grâce à leur pseudonyme sur le site de Lancers. Une série de questions d'ordre démographique nous a permis de collecter diverses informations sur les participants : leur âge ; leur sexe ; leur profession ; leur lieu de résidence actuel ; leur région ou leur pays d'origine ; leurs précédentes régions ou pays de résidence ; l'âge auquel le japonais est devenu une langue employée dans leur quotidien ; leurs problèmes de vue ou de lecture ; leur niveau de maîtrise des langues étrangères ; et leur niveau d'éducation. L'âge moyen des participants est de 38.7 ans (écart type = 9.56). Quant au sexe : 3 participants n'ont pas souhaité l'indiquer ; 67 étaient des femmes ; et 66 étaient des hommes.

4.2 Résultats

Étant donné que les temps de réponses très élevés ou très bas de certains participants sont probablement dus à un manque d'attention ou d'implication dans la tâche, la première étape de notre analyse des résultats a été d'exclure les participants visiblement non coopératifs. Le critère d'exclusion qui a été retenu est celui de la transformation logarithmique du temps médian de réponse, car il présentait la distribution la plus normale sur l'ensemble des participants parmi tous les critères considérés. Ainsi, nous avons exclu les participants dont le log du temps de réponse médian était éloigné de plus de deux écarts types par rapport à la moyenne du log du temps de réponse médian de l'ensemble des participants. Cela a mené à l'exclusion de neuf participants, dont huit ayant des temps de réponse trop courts, et un ayant des temps de réponse trop longs. Les résultats présentés ci-après sont donc ceux des 127 participants restants. Parmi ces 127 participants, 61 ont été exposés à la liste 1 de C2, tandis que 66 ont vu la liste 2.

La figure 9 ci-dessous présente la proportion d'utilisation de chaque pattern dans l'expérience, par l'ensemble des participants, en fonction de la deuxième more du deuxième constituant de la base. La même distinction que précédemment a été effectuée entre les mores *sans conflit* et les mores problématiques en position finale /R/ et /Q/. Notons que nous désignerons désormais par les termes « items *sans conflit* » et « CAL *sans conflit* » les items et

irrégulières dans cette situation. Parmi les trois patterns irréguliers, il ne semble pas y avoir de forte préférence, même si le pattern de Remplacement est légèrement plus présent.

Nous avons tout d'abord testé si le type d'item avait une influence significative sur la proportion de formes régulières et vérifié la validité de cette distinction en trois types de deuxième more du C2. Les autres caractéristiques phonologiques pouvant influencer sur la proportion de formes régulières (c.-à-d. la longueur du C2 ; le lieu d'articulation de sa consonne initiale ; et la hauteur de la voyelle allongée pour le cas des items /R/) ont également été prises en compte dans cette analyse. Un modèle linéaire mixte généralisé a été appliqué à l'ensemble des données avec : le type de pattern utilisé (régulier vs irrégulier) comme variable dépendante binaire ; le type d'items (/R/ ; *sans conflit* ; /Q/), la longueur du C2 (3 mores vs 4+ mores), le lieu d'articulation de la consonne initiale du C2 (bilabiale vs non bilabiale) comme facteur fixe ; la hauteur de la voyelle allongée comme facteur imbriqué dans le type d'item /R/ ; et le participant et le numéro du C2 en effets aléatoires. Les items /R/ sont le niveau de référence pour le type d'items, puisque leur ratio de formes régulières/formes irrégulières se situe entre les deux autres types d'items. Des tests du rapport de vraisemblance entre différents modèles ont été effectués pour évaluer la significativité des effets des différents facteurs. Les comparaisons de modèles ont révélé un effet significatif du type d'item ($\chi^2=242.02$; $p<0.001$) ; de la longueur du constituant ($\chi^2=9.9222$; $p<0.01$) ; et de la hauteur de la voyelle allongée pour les items /R/ ($\chi^2=6.5101$; $p<0.05$). Mais pas d'effet significatif du lieu d'articulation de la consonne initiale du C2 ($\chi^2=0.516$; $p=0.4726$). Un modèle intégrant l'interaction entre le type d'item et la longueur du constituant ne pouvant pas converger avec nos données, ce facteur n'a pas été intégré dans nos analyses.

Tandis que les données du lexique présentaient le pattern de Suppression comme étant largement majoritaire parmi les formes irrégulières, les résultats de l'expérience montrent que les différents patterns irréguliers sont employés dans des proportions assez similaires. Afin de tester s'il existe des différences significatives dans l'utilisation de ces patterns, un modèle logistique multinomial¹⁰ a été appliqué à l'ensemble des données avec : le pattern utilisé comme variable dépendante à quatre niveaux (Suppression ; Maintien ; Suivi ; Remplacement) ; et le

¹⁰ Toutes les analyses statistiques basées sur des modèles logistiques multinomiaux de cette thèse ont été effectuées sur R à l'aide du package *nnet* (Venables & Ripley, 2002).

type d'items comme prédicteur (*sans conflit* ; /R/ ; /Q/). Le pattern de Suppression sert de niveau de référence pour le modèle, car nous sommes principalement intéressés par les différences d'utilisation entre celui-ci et les autres patterns irréguliers. Le tableau 8 ci-dessous présente les valeurs z et valeurs p fournies par le modèle. Les valeurs dans chaque ligne sont obtenues en choisissant le type d'items correspondant comme niveau de référence pour le prédicteur.

Tableau 8 : valeurs z et valeurs p du modèle logistique multinomial.

	Maintien		Suivi		Remplacement	
	Valeur z	Valeur p	Valeur z	Valeur p	Valeur z	Valeur p
Items <i>sans conflit</i>	43.8	<0.001	0.91	0.36	-3.39	<0.001
Items /R/	26.94	<0.001	5.14	<0.001	4.53	<0.001
Items /Q/	-19.51	<0.001	3.19	<0.01	7.85	<0.001

Ces résultats confirment qu'il existe une différence significative entre l'utilisation du pattern de Maintien et du pattern de Suppression pour tous les types d'items, que ce dernier soit moins utilisé comme dans le cas de items *sans conflit* et /R/, ou plus utilisé comme dans le cas des items /Q/. Dans les items *sans conflit*, il n'y a pas différence significative entre l'utilisation du pattern de Suivi et de Suppression. Toutefois, le pattern de Suppression est utilisé plus souvent que le pattern de Remplacement. Concernant les items /R/ et /Q/, le pattern de Suppression est toujours utilisé de manière significativement moins fréquente que les deux autres patterns irréguliers.

4.3 Discussion

4.3.1 Comparaison entre les formes existantes et les nouvelles formes créées

L'expérience de création de nouvelles formes nous a permis d'explorer de quelle manière les japonais créent de nouveaux CAL. L'analyse des abréviations existantes suggère que leur formation est gouvernée par un principe général qui entre parfois en conflit avec des

contraintes phonologiques de la langue. Cela a des répercussions notables sur la manière dont les locuteurs traitent les différents types de mores en deuxième position du deuxième constituant. Tout d'abord, il est important de noter que même si le principe général de formation est très productif, il existe toujours une forte variation dans la manière dont les participants créent de nouvelles formes. Même dans le cas des items *sans conflit*, le pattern de Maintien représente seulement deux tiers des réponses. Cela laisse entrevoir que certains locuteurs au moins ne sont pas encore tout à fait conscients des généralisations qui gouvernent la formation des CAL.

Un autre aspect remarquable des résultats est qu'aucun pattern irrégulier ne se démarque clairement des autres. En d'autres termes, dans tous les contextes phonologiques, lorsque les participants n'utilisent pas le pattern régulier, ils ne montrent aucune préférence claire pour l'un des patterns irréguliers et ils sont utilisés presque aussi fréquemment. Bien que l'on observe des différences significatives dans les proportions d'utilisation des patterns irréguliers dans divers contextes, ces différences sont d'une ampleur assez faible. En comparaison, dans les données du lexique, le pattern de Suppression était largement plus fréquent que les autres patterns irréguliers. Il est d'autant plus étonnant que pour les items /R/ et /Q/, où les formes irrégulières sont les plus fréquentes, le pattern de Suppression se trouve dans des proportions inférieures aux deux autres patterns irréguliers. Non seulement les différences observées sont de faible ampleur, mais elles sont particulièrement difficiles à expliquer étant donné ce que l'on sait du lexique de formes existantes.

Ces résultats mettent en lumière la variabilité dans les CAL à deux niveaux différents : tout d'abord entre le pattern régulier et les patterns irréguliers ; ensuite, dans l'utilisation des différents patterns irréguliers. Dans cette variabilité, il reste apparent que la dynamique de la création des CAL est fortement conditionnée par le conflit entre le principe général de formation et les contraintes phonologiques générales de la langue (c.-à-d. *V:# et LICENCE/Q/). En effet, dans le cas des items *sans conflit*, où les contraintes phonologiques n'entrent pas en conflit avec le principe général de formation, la proportion d'utilisation du pattern de Maintien est clairement au-dessus de celle des autres patterns, étant deux fois plus fréquent que l'ensemble des patterns irréguliers combinés. Cette productivité du principe général de formation suggère qu'il a bien été assimilé par les locuteurs et que ces derniers sont capables de l'appliquer lorsqu'ils créent de nouvelles formes, en particulier lorsqu'il n'y a aucun obstacle phonologique à son application.

Dans le cas des items /R/, où le principe général de formation entre en conflit avec une contrainte peu sévère (c.-à-d. *V:#), on peut constater que la proportion d'utilisation du pattern de Maintien tombe en dessous de 50%. Même s'il s'agit toujours du pattern dominant, une importante différence avec les items *sans conflit* est qu'il y a, en tout, plus de formes irrégulières que de formes régulières créées dans ce contexte phonologique. Il est intéressant de remarquer que même si le principe général de formation est une option valable pour les items /R/, la plupart du temps, les participants préfèrent éviter de l'utiliser, même s'il ne semble pas y avoir de consensus sur la meilleure alternative à celui-ci. Les items /R/ montrent donc la variabilité qui existe dans les CAL aux deux niveaux mentionnés précédemment. Non seulement les participants sont partagés entre la création d'une forme régulière ou irrégulière, mais ils sont également partagés sur le pattern à employer s'ils créent une forme irrégulière.

Le cas des items /Q/, même s'il partage des similitudes avec celui des items /R/, se différencie par la *strictness* de la contrainte qui entre en conflit avec le principe général de formation (c.-à-d. LICENCE/Q/), le rendant pratiquement inutilisable dans ce contexte phonologique. De manière attendue, le pattern de Maintien n'a été que très peu utilisé pour la création de nouvelles formes avec les items /Q/, ce qui signifie que les participants n'ont créé quasiment que des formes irrégulières. Cependant, cette fois encore, il n'y a pas de nette préférence dans le pattern irrégulier qu'ils utilisent. Par conséquent, la variabilité au niveau du ratio de formes régulières/irrégulières est bien plus basse que pour les items /R/, alors que la variabilité sur l'utilisation des différents patterns irréguliers reste identique. Cela confirme qu'il n'existe pas de consensus sur le pattern irrégulier à utiliser, même quand le pattern régulier semble inapplicable. Tandis que dans le cas des items /R/, les participants étaient indécis à la fois sur l'utilisation ou non du pattern régulier et sur le choix du pattern irrégulier à employer, dans le cas de /Q/, il est clair pour les participants que le principe général de formation ne devrait pas être utilisé, mais cela ne les aide pas davantage à converger vers une unique solution pour traiter ces formes irrégulières.

Les résultats confirment aussi qu'une longueur trimorique du C2 ainsi que l'allongement d'une voyelle non haute dans le cas des C2 /R/ sont des facteurs entraînant également un traitement irrégulier plus fréquent. Toutefois, l'influence de ces deux facteurs sur nos résultats ne peut qu'être minime, puisqu'ils ont été contrebalancés dans les différents types d'items. Notons que même dans les cas où ces facteurs ne pouvaient pas être parfaitement contrebalancés, ils sont tout de même représentés dans des proportions très proches dans les différents types

d'items. Cela est détaillé dans les listes de C2 utilisées dans cette expérience disponibles dans les annexes 6 et 7.

En résumé, ces résultats indiquent que les participants sont conscients du principe général de formation des CAL et qu'ils sont plutôt compétents pour l'appliquer à la création de nouvelles formes tout en étant sensibles aux conflits de différentes sévérités qu'il peut générer sur le plan phonologique. Ils sont cependant indécis quant au comportement à adopter lorsque ces conflits penchent en faveur des contraintes phonologiques de la langue. Par conséquent, la proportion d'utilisation du principe général de formation diminue à mesure que la *strictness* de la contrainte qu'il enfreint augmente, mais la distribution des formes irrégulières dans les différentes alternatives à ce principe général est plutôt équilibrée dans tous les contextes phonologiques.

En termes d'OT stochastique, ces résultats peuvent s'expliquer en replaçant CONTIG et *PIED>2 au même niveau que *PIED<2, comme sur la figure 10 ci-dessous. En effet, les différences entre les données du lexique et celle de la tâche de création de nouvelles formes pourraient venir de l'incapacité des locuteurs à hiérarchiser *PIED<2 par rapport aux deux autres contraintes dont l'infraction favorise d'autres patterns irréguliers. Le fait que ces trois contraintes occupent la même position sur l'échelle de *strictness* est en accord avec le fait qu'aucun pattern irrégulier ne semble être préféré aux autres dans les différents contextes phonologiques.

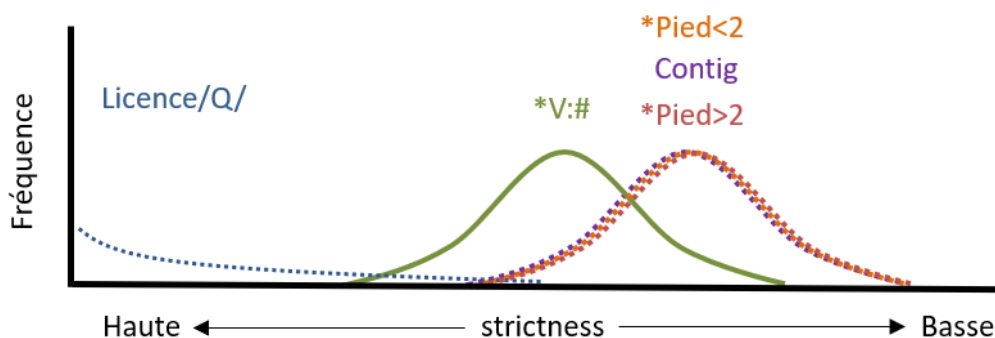


Fig.10 : position des contraintes impliquées dans la formation des CAL sur une échelle de *strictness*, selon les données de la tâche de création de nouvelles formes.

Notons que le pattern de Maintien est le plus fréquent pour le traitement de /R/ même si *V:# est placée au-dessus du groupe de trois contraintes. Un Maintien de la more /R/ émergerait seulement si, lors de l'évaluation d'un output, la *strictness* de *V:# se retrouve en-dessous de

celle des trois autres contraintes en même temps, ce qui n'arrive pas dans la majorité des cas. Cependant, à chaque fois que *V:# n'est pas la contrainte la moins sévère, chacune des trois autres contraintes aura autant de chances d'être celle avec la *strictness* la plus basse, chaque résultat aboutissant à un traitement différent de la more /R/. Ainsi, *V:# est globalement plus sévère que les trois autres contraintes, comme le montre la préférence pour les formes irrégulières plutôt que la pattern de Maintien, mais la distribution égale des formes irrégulières entre les différents patterns, causée par la compétition serrée des contraintes correspondantes, permet au pattern de Maintien d'être le plus fréquent pour le traitement de /R/.

4.3.2 Reproductibilité des résultats par un modèle d'apprentissage *constraint-based*

Maintenant que nous disposons de données concernant la création de nouveaux CAL, nous pouvons nous interroger sur la manière dont les données du lexique peuvent faire émerger chez les locuteurs des grammaires pouvant générer les outputs observés dans cette tâche de création de nouvelles formes. Pour cela, nous allons nous intéresser au modèle d'apprentissage *Gradual Learning Algorithm* (désormais GLA) développé conjointement au cadre théorique de l'OT stochastique (Boersma, 1997 ; Boersma & Hayes, 2001), et intégré dans le logiciel *Praat* (Boersma & Weenink, 2020). Ce modèle d'apprentissage *constraint-based* requiert deux éléments dans son fonctionnement. Le premier est une grammaire de contraintes non hiérarchisées indiquant sur quels types d'inputs elle est applicable et pour chacun de ces inputs, quels outputs enfreignent quelles contraintes. Le second est un lexique de paires d'input-output sur lequel s'entraîner. Après une phase d'entraînement, dont plusieurs aspects sont paramétrables, il peut être demandé au modèle GLA, entre autres, de produire une certaine quantité de nouvelles formes. Notre objectif est de déterminer si le modèle GLA peut reproduire les résultats de la tâche de création de nouvelles formes sur la base d'un apprentissage d'une grammaire et d'un lexique correspondant aux données du lexique de CAL.

La grammaire que nous avons présentée au modèle GLA est constituée des cinq contraintes impliquées dans le conflit entre le principe général de formation des CAL et la phonologie du japonais : LICENCE/Q/ ; CONTIG ; *PIED>2 ; *V:# ; et *PIED<2. Les différents types d'inputs correspondent aux CAL *sans conflit*, /R/ et /Q/. Et pour chaque type d'inputs, quatre outputs représentant les quatre patterns communs de formation des CAL sont possibles. Les infractions commises par chacun de ces outputs sont également renseignées. La figure 11

ci-dessous présente l'état initial de la grammaire. La *ranking value* de chaque contrainte représente sa position dans la hiérarchie, ou dans notre cas, sur l'échelle de *strictness*. Cette *ranking value* est la moyenne de la gaussienne représentant la contrainte sur l'échelle de *strictness*. La *disharmony* est la valeur de *strictness* effective que prend la contrainte lors d'une évaluation stochastique particulière. Elle varie de la *ranking value* en fonction du niveau de bruit (c.-à-d. une valeur représentant l'écart type de la gaussienne de la contrainte) qui a été sélectionné. Dans l'état initial, le modèle n'étant pas en cours d'évaluation, elles sont égales aux *ranking values*. La *plasticity* représente l'importance de l'ajustement que la contrainte subit à chaque fois que l'apprentissage amène le modèle à devoir l'ajuster. Ici, nous avons conservé la valeur par défaut pour la *plasticity*. En revanche, nous avons choisi une *ranking value* initiale très faible dans le but, entre autres, de pouvoir observer des formes irrégulières pour les inputs *sans conflit* qui présentent un candidat optimal n'enfreignant aucune contrainte, et ce même avec un entraînement très court. De plus, la *ranking value* initiale de la contrainte LICENCE/Q/ est largement supérieure à celle des autres contraintes pour refléter son caractère quasi-inviolable dans la phonologie du japonais. Sans cela, avec un entraînement court, le modèle n'aurait pas le temps d'assimiler cette caractéristique et proposerait systématiquement un nombre extraordinairement élevé d'output avec un /Q/ final non spécifié. Notons que l'on peut observer le niveau de *disarharmony* total de chaque input, à droite des tableaux, sachant que chaque infraction à une contrainte ajoute la valeur de *disharmony* de cette contrainte à ce score.

Concernant le lexique sur lequel nous avons entraîné cette grammaire, il correspondait aux données du lexique de CAL, c.-à-d. qu'il contenait :

- 541 paires d'input-output correspondant à des CAL *sans conflit* dont :
 - 474 possédaient un output suivant le pattern de Maintien.
 - 11 possédaient un output suivant le pattern de Suivi.
 - 46 possédaient un output suivant le pattern de Suppression.
 - 10 possédaient un output suivant le pattern de Remplacement.

- 113 paires d'input-output correspondant à des CAL /R/ dont :
 - 41 possédaient un output suivant le pattern de Maintien.
 - 7 possédaient un output suivant le pattern de Suivi.
 - 54 possédaient un output suivant le pattern de Suppression.
 - 11 possédaient un output suivant le pattern de Remplacement.

- 47 paires d'input-output correspondant à des CAL /Q/ dont :
 - o 2 possédaient un output suivant le pattern de Maintien.
 - o 10 possédaient un output suivant le pattern de Suivi.
 - o 29 possédaient un output suivant le pattern de Suppression.
 - o 6 possédaient un output suivant le pattern de Remplacement.

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
Licence/Q/	10.000	10.000	1.000000
*Pied<2	1.000	1.000	1.000000
*Pied>2	1.000	1.000	1.000000
*V:	1.000	1.000	1.000000
▲ □ Contig	1.000	1.000	1.000000

CAL sans conflit	Licence/Q/	*Pied<2	*Pied>2	*V:	Contig	
☞ maintien						0
suivi			*			-1.000
suppression		*				-1.000
remplacement					*	-1.000

CAL /R/	Licence/Q/	*Pied<2	*Pied>2	*V:	Contig	
☞ maintien				*		-1.000
☞ suivi			*			-1.000
☞ suppression		*				-1.000
☞ remplacement					*	-1.000

CAL /Q/	Licence/Q/	*Pied<2	*Pied>2	*V:	Contig	
maintien	*					-10.000
☞ suivi			*			-1.000
☞ suppression		*				-1.000
☞ remplacement					*	-1.000

Fig.11 : état initial de la grammaire présentée au modèle GLA.

Lors du processus d'apprentissage pour une évaluation, le modèle GLA sélectionne de manière aléatoire une paire d'input-output dans le lexique qui lui est présenté. Il produit alors

un output pour l'input de cette paire. Pour cela, il attribue une valeur de *disharmony* à chaque contrainte correspondant à sa *ranking value* modulée par un bruit dont l'importance peut être paramétrée. L'output sélectionné par le modèle GLA est celui dont la *disharmony* totale est la plus faible. Puis, le modèle GLA compare l'output qu'il a créé avec l'output réel, spécifié dans le lexique. Si les deux outputs sont identiques, rien ne se passe et l'apprentissage pour cette évaluation s'arrête là. En revanche, s'ils sont différents, le modèle GLA diminue la *ranking value* des contraintes qui sont enfreintes par l'output spécifié dans le lexique et non enfreintes par l'output que le modèle a généré. Inversement, il augmente la *ranking value* des contraintes non enfreintes par l'output spécifié dans le lexique et enfreinte par l'output généré par le modèle. L'importance avec laquelle les contraintes sont ajustées dépend de la valeur de *plasticity* éventuellement modulée à chaque évaluation par un certain bruit également (c.-à-d. le *plasticity spread*, soit l'écart type appliqué à la *plasticity*).

Bien entendu, l'apprentissage du modèle GLA peut être automatisé pour enchaîner un grand nombre de répétitions du processus décrit ci-dessus. Plusieurs paramètres peuvent être modifiés tels que le nombre de répétitions, le bruit déterminant la *disharmony* de chaque contrainte à chaque évaluation, ou différents aspects relatifs à la *plasticity*. Les paramètres par défaut, correspondant à un nombre important de répétitions (100000 répétitions)¹¹ avec un bruit faible (écart type de 2 par rapport à la *ranking value*), permettent d'obtenir une grammaire produisant des outputs très proches des données du lexique, donc assez éloignés des données de la tâche de création de nouvelles formes. Pour obtenir des outputs plus proches des résultats de cette tâche, nous avons donc tenté de réduire considérablement le nombre de répétitions à seulement 10, afin de simuler la faible exposition que les locuteurs japonais peuvent avoir avec le processus de formation des CAL. Notons qu'en maintenant un bruit de 2 sur des *ranking values* aussi faibles, la variabilité dans la hiérarchie des contraintes lors de chaque évaluation est particulièrement élevée. Par ailleurs, nous n'avons pas de raisons de croire que certaines contraintes doivent être ajustées de manière plus importantes que d'autres et n'avons donc pas modifié les paramètres de *plasticity* et *plasticity spread*, ce qui correspond à une *plasticity* de 1 avec un écart type de 0.1. Ce seuil de *plasticity* est particulièrement élevé par rapport aux

¹¹ Par défaut, le modèle effectue en réalité un cycle d'apprentissage à quatre niveaux de plus en plus faibles de *plasticity* permettant une granularité assez fine dans l'ajustement des différentes contraintes. Il effectue donc 100000 répétitions à une *plasticity* de 1, puis 100000 nouvelles répétitions à une *plasticity* de 0.1, puis idem pour une *plasticity* de 0.01, puis de 0.001. Par simplification, nous dirons simplement que le modèle fait ici 100000 répétitions.

ranking values initiales et, malgré la haute variabilité de leur issue, chaque évaluation peut donc avoir une influence considérable sur la grammaire, comme cela est vraisemblablement le cas quand les locuteurs ont peu de formes existantes à disposition.

En réinitialisant la grammaire et en reproduisant cet entraînement plusieurs fois, nous obtenons des grammaires assez différentes, notamment en termes de préférences pour les différents patterns irréguliers. Toutefois, dans la grande majorité des cas, le pattern de Suppression reste préféré aux deux autres patterns irréguliers. La figure 12 ci-dessous donne l'exemple des *ranking values* d'une grammaire obtenues après cet entraînement. On peut constater que *PIED<2 a une valeur plus faible que CONTIG et *PIED>2, qui elles ont des valeurs similaires. Cette grammaire favorise donc le pattern de Suppression par rapport au pattern de Suivi ou de Remplacement. Réduire drastiquement le nombre de répétitions ne permet donc pas, a priori, d'obtenir des grammaires qui pourrait produire des outputs comparables à ceux de l'expérience, où les patterns irréguliers sont utilisés dans des proportions similaires.

	<i>ranking value</i>
*Pied>2	2.149
Contig	2.248
*Pied<2	1.118
Licence/Q/	2.001
*V:	0.892

Fig.12 : état de la grammaire après un apprentissage de 10 répétitions avec un bruit de 2.

Il est néanmoins possible de tenter de s'approcher des résultats de la tâche de création de nouvelles formes à partir de la grammaire prise en exemple ci-dessus. En effet, comme mentionné précédemment, il est possible de demander au modèle GLA de produire un certain nombre d'outputs et, de manière intéressante, d'ajouter un bruit dans l'évaluation des contraintes lors de la production également. Afin de répliquer approximativement la tâche réalisée par nos participants, nous avons fait produire 40 outputs par type d'inputs au modèle GLA (contre 48 items *sans conflit* ; 48 items /R/ ; et 24 items /Q/ dans l'expérience). La production d'un faible nombre d'output couplé à l'ajout d'un bruit permet d'obtenir des résultats très variables. Toutefois, notons qu'avec le niveau de bruit par défaut (écart type de 2 par rapport à la *ranking value*), très élevé par rapport au *ranking values* des contraintes, la grammaire prise en exemple ci-dessus produit toujours plus de patterns de Suppression que de Remplacement ou Maintien. Le tableau 9 présente un exemple d'une production d'outputs par

cette grammaire avec un bruit de 2. Augmenter le bruit permettrait d'obtenir régulièrement des outputs dans lesquels le pattern de Suppression n'est pas préféré aux autres patterns irréguliers. Toutefois, dans ce cas, l'utilisation de pattern de Maintien devient alors incohérente avec les données de l'expérience. Cela est déjà en partie visible dans le tableau 9 en cela que le pattern de Maintien représente moins de la moitié de CAL *sans conflit*, qu'il est plus utilisé dans les CAL /R/ que dans les CAL *sans conflit*, et qu'il est utilisé assez fréquemment pour les CAL /Q/.

Tableau 9 : production de 40 outputs pour chaque type d'input par le modèle GLA avec un bruit de 2.

	<i>sans conflit</i>	/R/	/Q/
Maintien	17	19	6
Suivi	10	4	10
Suppression	11	13	17
Remplacement	2	4	7

Le modèle GLA ne permet donc pas de rendre compte des résultats de la tâche de création de nouvelles formes. Soit nous devons augmenter le nombre de répétitions de l'apprentissage ou réduire le bruit à différents niveaux du processus pour reproduire les proportions d'utilisation du pattern de Maintien, mais le pattern de Suppression est alors préféré aux autres patterns irréguliers. Soit nous devons diminuer les répétitions ou augmenter le bruit pour reproduire le fait qu'aucun pattern irrégulier ne soit préféré, mais ce que l'on observe au niveau de l'emploi du pattern de Maintien n'est alors plus cohérent avec les données. Les résultats produits par le modèle GLA indiquent que même en se basant sur un petit échantillon de CAL issu du lexique de formes existantes, les locuteurs devraient soit avoir assez d'informations pour reproduire les probabilités observées dans le lexique dans leurs outputs aussi bien au niveau de l'emploi du pattern régulier que de la préférence pour le pattern de Suppression parmi les irréguliers ; soit avoir trop peu d'informations et ne parvenir à reproduire aucune des probabilités observées dans le lexique.

Dans les résultats de l'expérience, les participants arrivent à reproduire les probabilités du pattern régulier dans leurs outputs tout en ne montrant pas de préférence, en moyenne, pour l'un des patterns irréguliers. Or, le modèle GLA montre que même une faible exposition au lexique de CAL devrait permettre au pattern de Suppression d'émerger en tant que sous-règle

à utiliser pour la formation de CAL irréguliers. Nous pouvons en conclure qu'un autre facteur empêche les locuteurs d'abstraire ce pattern à partir du faible nombre de types dont ils disposent. Au vu des résultats de l'expérience, ce facteur se manifeste par des préférences pour certains patterns irréguliers, ou des absences de préférence, non corrélées avec les fréquences de types de certains patterns qui, lorsqu'on observe la moyenne des participants, entraîne l'emploi des différents patterns dans des proportions similaires. Dans une vision bayésienne de l'apprentissage, ces préférences, ou absences de préférence, non motivées par les fréquences de types peuvent être intégrées dans les biais a priori avec lesquels les locuteurs abordent le traitement des CAL. Nous reviendrons en détails sur les modèles bayésiens de l'apprentissage en section 9.2.

4.3.3 Validité des prédictions formulées sur la base des fréquences de types

À présent que nous disposons de données sur les fréquences de types des différents patterns dans les formes existantes et sur leur productivité dans la création de nouvelles formes, nous pouvons passer en revue les différentes prédictions formulées en section 3.3 et déterminer si elles sont confirmées ou non par les données de l'expérience :

- Pour les items *sans conflit* : le pattern de Maintien devrait être le plus productif. Les autres patterns ne devraient avoir qu'une faible productivité.

Cette prédiction semble être confirmée par les résultats de la tâche de création de nouvelles formes. En effet, pour les items *sans conflit*, les participants ont préféré le pattern de Maintien dans la grande majorité des cas, et les autres patterns ne montrent que peu de productivité. La seule différence que l'on peut constater entre les données du lexique et celles de l'expérience est que la prévalence du pattern de Maintien est légèrement moindre dans les données de l'expérience. Par conséquent, tous les patterns irréguliers sont légèrement sur-représentés. Globalement, nous pouvons conclure que les prédictions basées sur les fréquences de types dans le lexique sont en accord avec les résultats de l'expérience pour les items *sans conflit*. Ainsi, la productivité du principe général de formation est en accord avec ce qui serait attendu si les participants suivaient un comportement de probability-matching.

- Pour les items /Q/ : le pattern de Maintien devrait être quasiment improductif, et le pattern de Suppression devrait être le plus productif.

Même si nous pouvons considérer que la première partie de cette prédiction est confirmée, en revanche, la deuxième partie ne correspond pas aux données de l'expérience. De manière attendue, le pattern de Maintien entraînant la création de formes imprononçables est pratiquement improductif pour les items /Q/. Cependant, si les données du lexique montraient une claire préférence pour le pattern de Suppression, il ne semble y avoir de préférence forte pour aucun des patterns irréguliers dans les données de l'expérience. Le fait que le pattern de Suppression ne soit pas plus productif que les autres patterns irréguliers est particulièrement inattendu si l'on considère qu'il possède la plus haute fréquence de types dans ce contexte phonologique et qu'il est de manière générale uniquement dominé par le pattern de Maintien qui, ici, n'est pas applicable pour des raisons phonologiques. Pour les items /Q/, l'improductivité du pattern général est en lien direct avec sa fréquence de types proche de zéro dans les formes existantes, mais les prédictions basées sur l'idée du probability-matching surestiment fortement la productivité du pattern contextuel de Suppression. Selon cette approche, il n'y a pas de raison pour qu'un pattern avec 129 types en tout, et 29 types parmi les CAL ayant /Q/ comme deuxième more de C2, ne présente pas de productivité supérieure aux patterns concurrents avec une fréquence de types bien plus faible. En résumé, la productivité des patterns irréguliers dans le traitement des items /Q/ dévie clairement du principe de probability-matching, mais l'improductivité du pattern de Maintien indique que cette déviation ne peut pas s'expliquer par une compétition entre un pattern général et des patterns contextuels comme dans les cas documentés par Kapatsinski (2010).

- Pour les items /R/ : les patterns de Maintien et de Suppression pourraient tous les deux être productifs. Qu'un des deux patterns domine l'autre, ou qu'ils aient une productivité équivalente, ils devraient dominer les patterns de Suivi et de Remplacement.

Les résultats de l'expérience montrent que le pattern de Maintien a effectivement une certaine productivité pour les items /R/. Cependant, il ne domine pas clairement les patterns irréguliers. En effet, même si le pattern de Maintien est le plus fréquent, il est moins fréquent que l'ensemble des formes irrégulières combinées. Cela indique que les locuteurs japonais traitent effectivement la more /R/ de manière différente des mores *sans conflit* et évitent d'utiliser le principe général de formation des CAL avec cette more dans la plupart des cas. Néanmoins, il semble une fois de plus qu'il n'y ait pas de préférence pour le pattern de Suppression sur les autres patterns irréguliers dans ce contexte phonologique. Les résultats de l'expérience ne correspondent donc ni à une situation dans laquelle le pattern de Maintien ou

de Suppression domine largement les autres, ni à une situation où ces deux patterns ont une productivité similaire, au-dessus des deux autres. La productivité du pattern de Maintien est légèrement inférieure à celle des patterns irréguliers combinés, et la productivité des différents patterns irréguliers est presque égale. On observe une fois de plus que la productivité du pattern de Maintien est bien corrélée avec sa fréquence de types dans le lexique, ce qui est cohérent avec l'idée d'un comportement de probability-matching. Cependant, même si notre prédiction concernant les items /R/ était assez générale et couvrait plusieurs possibilités concernant l'issue de la compétition entre le pattern de Maintien et de Suppression, elle échoue à rendre compte des résultats de l'expérience. Comme dans le cas des items /Q/, les fréquences de types s'avèrent donc insuffisantes pour prédire la productivité des patterns irréguliers. En particulier, le pattern de Suppression ne ressort pas comme un pattern contextuel productif, ce qui serait pourtant prédit en observant sa fréquence de types que ce soit dans les CAL ayant /R/ pour deuxième more de C2, ou dans l'ensemble des abréviations. Ceci est un argument supplémentaire indiquant que l'utilisation de patterns irréguliers par les locuteurs japonais dévie du comportement de probability-matching et que cela ne saurait s'expliquer par la compétition avec un pattern plus général, puisque le pattern de Maintien a également une productivité relativement faible.

La différence principale entre les données du lexique et de l'expérience réside dans l'absence de productivité du pattern de Suppression malgré sa fréquence relativement élevée dans le lexique. Puisque l'on observe la même dynamique dans les différents types d'items (c.-à-d. que le pattern de Maintien a la productivité prédite, mais que ce n'est pas le cas des patterns irréguliers), le processus morphologique de la création des CAL possède probablement des caractéristiques qui permettent aux locuteurs de reproduire dans leur output la fréquence du principe général dans l'input, mais les empêchent de faire de même avec les patterns contextuels. Comme dans le cas de l'affixation en russe, une telle déviation du probability-matching pour des patterns contextuels aurait pu être causée par une compétition avec un pattern plus général et plus fréquent, mais les résultats de notre étude n'encouragent pas cette interprétation. Les caractéristiques qui mènent à cette issue sont donc encore à déterminer.

Un premier élément d'explication à considérer, puisque l'on étudie un processus morphologique devenu productif récemment, est l'exposition limitée que les locuteurs ont avec ce processus. En effet, dans les premiers temps de l'existence d'un processus morphologique, comment est-ce que les formes existantes pourraient influencer sur la création de nouvelles formes

s'il n'existe pas, ou peu, de formes ? Les modèles bayésiens de l'apprentissage apportent une réponse convaincante à cette question puisqu'ils reposent non seulement sur le principe que les individus prennent chaque nouvel input en compte pour réajuster leurs croyances sur l'état du monde (en l'occurrence, leurs représentations linguistiques), mais aussi et surtout qu'ils abordent toute situation d'apprentissage avec un ensemble de biais a priori. Notons que même en adoptant ce point de vue, il est logique que des locuteurs ayant reçu une longue exposition à un input donné, noyant totalement ces biais a priori, adoptent un comportement assimilable à du probability-matching. En revanche, les biais a priori jouent un rôle proportionnellement plus important dans le comportement de locuteurs exposés à un input limité (Pierrehumbert, à paraître). Le cas des CAL japonais illustre ces deux situations par certains de ses aspects.

Tout d'abord, le cas des CAL japonais montre qu'un processus morphologique, même s'il n'est devenu productif que très récemment dans la langue, ne sera jamais vide de formes existantes pouvant influencer sur la création de nouvelles formes. Les CAL japonais ne sont pas un lexique isolé sans aucune connexion avec le reste de la langue. Les contraintes phonologiques générales de la langue doivent être prises en compte dans la création de ces abréviations. En particulier, les contraintes concernant la présence de certaines mores en position finale de mot jouent un rôle important dans la variabilité que présentent les CAL. Les locuteurs ont eu une longue exposition à ces contraintes qui s'appliquent de manière générale dans la phonologie de la langue, et nous observons que les participants adoptent un comportement assimilable à du probability-matching précisément en ce qui concerne la *strictness* de ces contraintes phonologiques. Ceci apparaît dans la manière dont ils utilisent le pattern régulier de Maintien très souvent lorsqu'il n'entre en conflit avec aucune contrainte, moins souvent s'il entre en conflit avec une contrainte dont la *strictness* est modérée, et pratiquement jamais lorsqu'il entre en conflit avec une contrainte dont la *strictness* est très élevée. De plus, parmi les CAL eux-mêmes, le pattern de Maintien a la plus haute fréquence de types en vertu du fait qu'il s'agisse du principe général de formation et qu'il soit applicable dans une grande variété de contextes phonologiques. Avec un peu plus de 500 formes existantes, les locuteurs sont assez familiers avec ce pattern pour en faire l'abstraction et le généraliser à de nouvelles formes.

Ensuite, la création des CAL implique des contraintes spécifiques à ce nouveau processus morphologique auxquelles les locuteurs ont été relativement peu exposés. Dans l'expérience, les participants échouent précisément à reproduire la probabilité des patterns

irréguliers qui correspondent à ces contraintes, en cela qu'ils ne présentent de préférence forte pour aucun des patterns irréguliers dans les situations dans lesquelles le pattern régulier est inadéquat pour des raisons phonologiques. Non seulement les locuteurs ne peuvent pas s'appuyer sur des probabilités phonologiques générales dans leurs décisions concernant l'utilisation des patterns irréguliers, mais même au sein des CAL, les patterns irréguliers sont souvent utilisés dans des contextes précis qui sont naturellement plus rares que les contextes ne présentant pas de conflit. Le nombre de formes à partir desquelles les locuteurs pourraient abstraire un pattern irrégulier est donc drastiquement plus faible que le nombre de formes présentant le pattern de Maintien. Or, il semblerait logique qu'un nombre minimum de formes existantes présentant un pattern soit nécessaire pour qu'il puisse être généralisé, comme cela est suggéré par les études montrant que des patterns récurrents avec des fréquences de types faibles ne peuvent pas être généralisés (Bybee, 1995 ; Dawdy-Hesterberg & Pierrehumbert, 2014 ; Pierrehumbert, 2001c). Rappelons toutefois que notre analyse à l'aide du modèle GLA indique que de manière purement mathématique, les locuteurs devraient être capables de reconnaître le pattern de Suppression comme étant celui à privilégier parmi les patterns irréguliers même en étant exposés à très peu de formes. Le fait que cela ne soit pas ce que l'on observe dans les résultats de l'expérience nous amène à préciser ce que l'on définit comme « nombre minimum de formes existantes nécessaires pour pouvoir généraliser un pattern ». Il ne s'agit pas ici de considérer que les locuteurs sont incapables d'appliquer un pattern qu'ils n'ont observés que dans quelques formes. Nous faisons en revanche l'hypothèse que les locuteurs sont biaisés vers une absence de préférences pour les patterns irréguliers, ou bien vers une préférence qui n'est pas guidée par les fréquences de types. Une exposition au pattern de Suppression assez importante pour surpasser ce biais serait alors nécessaire pour permettre à ce pattern d'être généralisé.

Si l'on considère chaque type de mores individuellement, on constate d'une part qu'il y a un peu moins de 50 CAL dans le lexique dont la deuxième more du C2 est /Q/ et pour lesquels le principe général de formation est inapplicable. Cela n'est peut-être pas suffisant pour que les locuteurs puissent généraliser l'un des patterns irréguliers, quelle que soit l'importance de la proportion qu'il occupe parmi ces 50 formes. D'autre part, le nombre de CAL pour lesquels la deuxième more de C2 est /R/ est un peu plus élevé, avec un peu plus de 100 formes. Cependant, la compétition entre les différents patterns est plus difficile dans ce contexte puisque le principe

général est une option valide. Ce qui laisse encore une fois peu de formes à partir desquelles extraire un pattern.

Si l'on considère le lexique dans son ensemble, le pattern de Suppression apparaît comme une sous-règle à appliquer lorsque le principe général est inadéquat, étant présent dans 130 formes. Cependant, cela est peut-être encore trop peu pour qu'il soit généralisé par les locuteurs et donc productif dans la création de nouvelles formes. Une fréquence de types de 130 formes peut paraître un nombre assez important pour abstraire un pattern, mais ce nombre ne reflète probablement pas le nombre de formes effectivement connues par chaque locuteur. Les CAL étant des abréviations récentes et par nature optionnelles, il se peut que la variation dans l'exposition de chaque locuteur à ces formes soit considérable, notamment en comparaison avec des processus morphologiques établis de longue date, qui s'appliquent de manière systématique, et qui sont utilisés fréquemment au quotidien. Par exemple, un petit groupe de locuteurs passionnés de culture populaire pourrait être à l'origine de la création et de l'emploi de la plupart des CAL, tandis que des locuteurs plus distants de cette culture ne seraient familiers qu'avec les CAL les plus communs. Si l'on suppose que le nombre de formes existantes suivant le pattern de Suppression est trop faible pour que les locuteurs le généralisent, alors leur comportement concernant la sélection d'un pattern irrégulier reposerait principalement sur leurs biais a priori. En d'autres termes, ils ne pourraient pas suivre un comportement de probability-matching.

Une description bayésienne du processus d'apprentissage des différents patterns morphologiques utilisés dans la formation des CAL permet d'expliquer la variabilité d'utilisation des différents patterns par les participants à l'expérience de création de nouvelles formes. Néanmoins, cette explication ne résout pas entièrement le problème des différences observées dans l'emploi du pattern de Suppression dans les deux études. Tout au plus, en répondant à la question « pourquoi le pattern de Suppression n'est pas productif dans la création de nouvelles formes ? », elle soulève une nouvelle question : « comment le pattern de Suppression peut apparaître comme une sous-règle pour les formes irrégulières existantes si les locuteurs ne l'utilisent pas comme une sous-règle quand ils créent de nouvelles formes irrégulières ? ». Cette nouvelle question est plus difficile à aborder, mais on peut considérer plusieurs éléments de réponse.

Premièrement, tous les locuteurs ne sont pas des innovateurs, donc tous les locuteurs ne créent pas de CAL. Même si ce n'est pas une préférence pour l'ensemble de la population, le pattern de Suppression pourrait être préféré par les locuteurs qui emploient régulièrement des mots d'emprunt et sont enclins à créer des CAL. En d'autres termes, le pattern de Suppression pourrait être le pattern irrégulier préféré dans les biais a priori des innovateurs. Nous avons vu en section 1.3.3 que la nature itérative de l'apprentissage linguistique avait pour conséquence de rapidement amplifier de légères préférences présentes dans l'input transmis. En théorie, il est donc envisageable que le pattern de Suppression devienne productif au niveau communautaire même si au niveau des individus il n'est en réalité que légèrement préféré par les locuteurs qui construisent l'input de la communauté.

Deuxièmement, il est important de rappeler que le lexique de CAL n'est pas représentatif de l'ensemble des CAL créés, mais seulement des formes ayant survécu assez longtemps dans la communauté des locuteurs japonais pour être inventoriés. Même si le pattern de Suppression n'est pas la préférence naturelle de la plupart des locuteurs quand il s'agit de créer de nouvelles formes, peut-être crée-t-il des formes qui sont moins appréciés par la plupart des locuteurs que celles créées par les autres patterns, augmentant ainsi les chances que ces formes soient mémorisées et réemployées dans la communauté. En d'autres termes, dans les biais a priori des locuteurs qui ne favorisent pas le pattern de Suppression, l'acceptabilité de ce pattern pourrait n'être que légèrement en-dessous de celle de leur pattern préféré, tandis que les autres patterns seraient totalement inacceptables. Pour rappel, Smith (2009) montrait que dans des populations d'agents bayésiens apprenant une langue artificielle de manière itérative avec une transmission plus complexe que de « un à un », comme c'est le cas dans une véritable communauté de locuteur, les préférences les plus répandues dans l'input initial ne sont pas forcément celles qui émergent comme structure en fin de chaîne de transmission. La nature complexe des interactions entre locuteurs rend donc d'autant plus opaques les raisons permettant au pattern de Suppression d'émerger comme la sous-règle à employer pour les formes irrégulières. À défaut de pouvoir retracer tout le processus de transmission de ce pattern dans le réseau de locuteurs, nous pouvons faire l'hypothèse qu'il est considéré comme au moins acceptable par certains locuteurs et qu'il est promu par des dynamiques de niveau communautaire.

Enfin, une situation linguistique dans laquelle les locuteurs n'ont été exposés qu'à peu d'input et sont motivés uniquement par leurs préférences individuelles ne serait pas vraiment

pratique, notamment dans des situations qui présentent beaucoup de variabilité. Il serait sans doute plus avantageux pour les locuteurs d'être enclins à converger rapidement vers une convention, quelle qu'elle soit, tant que celle-ci ne va pas strictement à l'encontre de leur jugement d'acceptabilité. En termes d'apprentissage bayésien, cela signifie que les locuteurs peuvent réajuster leurs croyances sur l'état du monde de manière plus ou moins drastique avec chaque nouvel input, donnant ainsi plus ou moins d'importance aux nouveaux inputs par rapport à leurs préférences individuelles. Par conséquent, même les locuteurs n'ayant pas pour préférence initiale le pattern de Suppression peuvent rapidement converger vers celui-ci, s'ils y sont régulièrement exposés.

Le même raisonnement pourrait s'appliquer pour expliquer comment le pattern de Maintien est devenu le principe général de formation. Identifier les raisons pour lesquelles un nouveau processus morphologique apparaît dans une langue dépasse le cadre de cette thèse¹², mais lorsqu'un tel processus apparaît, il doit au moins comprendre un pattern général selon lequel de nouveaux mots peuvent être créés. À moins de considérer la phonologie et la morphologie comme des procédés purement analogiques (ce qui n'est pas notre positionnement, voir section 1.3.1), les processus morphologiques sont, par nature, un pattern de transformation d'une racine. Les locuteurs qui créent de nouveaux mots doivent suivre un certain pattern, même si celui-ci est totalement idiolectal. La variabilité présente même parmi les CAL existants ne provoquant pas de conflit avec les contraintes phonologiques pourrait être le reflet des diverses préférences de certains locuteurs concernant le pattern qui devrait être le principe général de formation des CAL. Mais les patterns concurrents au Maintien produisent probablement des formes peu aptes à être mémorisées et réemployées par la communauté, pour les mêmes raisons évoquées ci-dessus : la nature des innovateurs ; la manière dont l'ensemble de la population interagit d'une part, et hiérarchise les différents patterns d'autre part ; et l'inclinaison générale pour la convergence induite par la situation. Puis, plus les formes suivant le pattern de Maintien sont réemployées par la communauté, plus l'effet de sa fréquence de

¹² Dans le cas de CAL japonais, on pourrait argumenter qu'il existe un besoin d'abrégier les mots d'emprunt à cause de leur tendance à être particulièrement longs par rapport au reste du lexique japonais. En effet, préserver la structure syllabique du japonais requiert souvent l'épenthèse de voyelles dans l'adaptation de mots d'emprunt, pour empêcher l'apparition de clusters consonantiques ou de consonnes finales. Cela entraîne l'addition d'autant de mores à la longueur des mots. De plus, l'ampleur du problème est doublée dans le cas de mots d'emprunt composés. Les abrégier à une longueur de quatre mores les rend plus proches de la longueur moyenne des mots japonais.

types s'amplifie, jusqu'à ce que les préférences de la plupart des locuteurs basculent, faisant de ce pattern le plus productif de manière générale.

4.3.4 Conclusion

En conclusion de ces deux premières études, voyons les éléments de réponses qu'apporte la comparaison des données du lexique et de l'expérience à nos deux questions de recherche relatives à la construction des représentations qui guident la productivité des patterns morphologiques :

- 1) Les locuteurs reproduisent-ils simplement les probabilités de l'input dans leur output (c.-à-d. ont-ils un comportement de type « probability-matching ») ?
- 2) Sinon, quels autres facteurs entrent en jeu ?

L'absence de productivité du pattern de Suppression dans les nouvelles formes créées, malgré sa fréquence dans les formes existantes répond à la question 1) par la négative. D'autres études avaient déjà identifiés des situations linguistiques dans lesquelles les locuteurs n'adoptaient pas un comportement de probability-matching (Becker et al., 2011 ; Dawdy-Hesterberg, 2014 ; Kapatsinski, 2010). Nos résultats viennent appuyer ces observations en présentant un autre cas de figure dans lequel les locuteurs ne reproduisent pas les probabilités de l'input dans leur output : lorsque le nombre de formes existantes est trop faible pour abstraire certains patterns.

En réponse à la question 2), les résultats de l'expérience de création de nouvelles formes indiquent que les locuteurs possèdent des biais a priori pour certains patterns. Ces biais sont particulièrement apparents lorsque les locuteurs ne disposent que d'un input limité pour abstraire certains patterns. Cela est cohérent avec une vision bayésienne du processus d'apprentissage. De plus, la différence entre les données du lexique et de l'expérience suggère que des processus de niveau communautaire, notamment la convergence morphologique, aurait joué un rôle crucial dans l'émergence non seulement du principe général de formation, mais également d'une sous-règle contextuelle dans les formes existantes. La section suivante aura pour but d'explorer plus avant l'effet de la convergence morphologique dans la formation des CAL.

Chapitre 5 – Expérience ESP - CAL japonais : convergence morphophonologique dans un processus morphologique récent

Après avoir étudié les formes existantes de CAL ainsi que les nouvelles formes créées par les locuteurs dans un contexte expérimental, nous avons constaté des écarts inattendus entre la fréquence de types de certains patterns et leur productivité. Cela indique que malgré l'importance des probabilités dans la généralisation de patterns, les locuteurs ne parviennent pas toujours à reproduire les probabilités de l'input dans leur output. Dans le cas du pattern de Suppression dans les CAL, on peut conclure, en accord avec les principes d'un raisonnement bayésien, que les locuteurs se basent essentiellement sur leurs biais a priori par manque de formes existantes pouvant les convaincre de la productivité de ce pattern. Cependant, nous savons qu'avec une longue exposition, même dans une approche bayésienne, il est attendu que les locuteurs adoptent un comportement proche du probability-matching.

Il a été montré, par ailleurs, que les locuteurs pouvaient faire preuve de convergence, entre autres, au niveau morphologique. L'idée de convergence morphologique peut sembler de prime abord incompatible avec celle selon laquelle après un long entraînement les locuteurs ont un comportement proche du probability-matching. En effet, si chaque nouvel input a son importance dans les représentations du locuteur, après une longue exposition, non seulement ses biais initiaux ne devraient plus énormément peser dans la balance, mais les nouveaux inputs devraient également être, en quelque sorte, noyés dans la masse. Cependant, la convergence à différents niveaux linguistiques se produit assez rapidement, et ne peut être attribuée à une longue exposition à une variante atypique pour le locuteur. Une manière de concilier ces deux idées serait de reconsidérer l'importance accordée aux nouveaux inputs dans ce modèle d'apprentissage bayésien. Si les nouveaux inputs ont une forte importance, alors il est possible que les locuteurs convergent rapidement vers une variante atypique lorsqu'une situation particulière l'exige, mais qu'ils retrouvent leur variante habituelle dès qu'ils sont de nouveau exposés à un input habituel. Dans l'expérience présentée ci-après, nous allons explorer la manière dont les locuteurs traitent de nouveaux inputs reçus d'un partenaire d'interaction dans la création de nouvelles formes.

5.1 Méthode

5.1.1 Matériel

Comme dans l'expérience de création de nouvelles formes précédente, chaque participant voyait un total de 120 items. Cependant, cette expérience se divise en trois phases distinctes : un pré-test ; une phase d'interaction avec un partenaire artificiel ; et un post-test. Les items étaient donc cette fois répartis en trois sous-listes de 40, correspondant aux trois phases du paradigme ESP (le déroulement de l'expérience est détaillé en section 5.1.2). Les mots d'emprunts utilisés pour former les items de l'expérience ESP sont issus des listes de C1 et C2 utilisées dans l'expérience de création de nouvelles formes (désormais « expérience baseline »). Seule la moitié des C2 utilisés dans l'expérience baseline ont été utilisés dans l'expérience ESP. Ainsi, il n'y avait pour cette dernière qu'une seule liste de C1 et de C2 (respectivement annexe 5 et 8). Concernant les caractéristiques phonologiques des C2, la liste était structurée de la même manière que pour l'expérience baseline, soit en dix types de C2 chacun représentés par douze mots. Cependant, en ne conservant que la moitié des C2, il n'était pas possible de conserver les sets d'items avec des différences minimales sur le plan phonologique. La liste de C2 comprend donc les types d'items suivants :

- 12 C2 *sans conflit* d'une longueur de **plus de trois mores, sans consonne bilabiale** à l'initiale (qui dans l'expérience baseline faisaient partie d'un set **d'un seul mot**).
- 12 C2 *sans conflit* d'une longueur de **plus de trois mores, sans consonne bilabiale** à l'initiale (qui dans l'expérience baseline faisaient partie d'un set **de trois mots** avec les deux types suivants).
- 12 C2 *sans conflit* d'une longueur **de trois mores, sans consonne bilabiale** à l'initiale.
- 12 C2 *sans conflit* d'une longueur de **plus de trois mores, avec consonne bilabiale** à l'initiale.
- 12 C2 /R/ avec allongement d'une **voyelle non haute** et une consonne **non bilabiale** initiale.
- 12 C2 /R/ avec allongement d'une **voyelle haute** et une **consonne non bilabiale** initiale.
- 12 C2 /R/ avec allongement d'une **voyelle non haute** et une **consonne bilabiale** initiale.
- 12 C2 /R/ avec allongement d'une **voyelle haute** et une **consonne bilabiale** initiale.
- 12 C2 /Q/ d'une longueur **de trois mores**.

- 12 C2 /Q/ d'une longueur de **plus de trois mores**.

La liste de 120 items utilisée dans cette étude est donc très similaire à celles présentées dans l'expérience baseline. Néanmoins, dans le paradigme ESP, seulement un tiers de cette liste apparaît dans chaque phase de l'expérience. Pour ne pas fausser les résultats, il est donc crucial que les sous-listes de 40 items présentées dans chaque phase soient équilibrées du point de vue du traitement que reçoivent, en moyenne, les items qui les composent. C'est en cela que les données de l'étude précédente ont été utilisées comme baseline. L'idée principale pour la formation de ces sous-listes équilibrées était de former des trios de C2 ayant reçus des traitements similaires dans l'expérience baseline, puis de répartir aléatoirement un membre de ces trios dans chacune des trois sous-listes. Le détail de la procédure pour former ces trios est donnée ci-après.

Tout d'abord, afin de s'assurer que nous disposions de C2 pouvant potentiellement mener à l'utilisation de tous les patterns, nous avons exclu les quelques C2 pour lesquels un ou plusieurs patterns n'avaient jamais été utilisés dans l'expérience baseline. Puis, pour chaque type de C2 décrit ci-dessus, nous avons constitué les trios de C2 qui avaient reçu en moyenne les traitements les plus similaires. Pour calculer cela, nous avons attribué à tous les trios possibles un score basé sur les différences entre les proportions d'utilisation de chacun des patterns pour chaque C2. Par exemple, voici les résultats obtenus pour les C2 n°1, 2 et 3 de la liste 1 lors de l'expérience baseline :

Tableau 10 : proportion d'utilisation (en %) de chaque pattern pour les C2 n° 1, 2, et 3 de la liste 1 dans l'expérience baseline.

Liste	Numéro C2	Anglais	% Maintien	% Suivi	% Suppression	% Remplacement
1	1	cancel	86.89	1.64	6.56	4.92
1	2	crowd	86.89	4.92	4.92	3.28
1	3	royal	70.49	1.64	21.31	6.56

Le score du trio de C2 « liste 1 n° 1 – liste 1 n° 2 – liste 1 n° 3 » a été calculé en additionnant la différence (en valeur absolue) entre la proportion de Maintien du C2 n°1 et n°2, puis du C2 n°1 et n°3, puis du C2 n°2 et n°3 et ainsi de suite pour les autres patterns. Cela donne :

$$\begin{aligned}
& |86.89 - 86.89| + |86.89 - 70.49| + |86.89 - 70.49| + |1.64 - 4.92| + |1.64 - 1.64| + |4.92 - 1.64| \\
& + |6.56 - 4.92| + |6.56 - 21.31| + |4.92 - 21.31| + |4.92 - 3.28| + |4.92 - 6.56| + |3.28 - 6.56| \\
& = \mathbf{78.7}
\end{aligned}$$

Après avoir calculé le score de toutes les combinaisons de trios possibles, le trio avec le score le plus faible était retenu pour être utilisé dans l'expérience ESP. Puis, toutes les combinaisons incluant au moins un des C2 faisant parti du trio déjà sélectionné étaient exclues. Par souci de pouvoir réutiliser les mêmes combinaisons de C1 et C2 dans l'expérience ESP que dans l'expérience baseline, étaient également exclues toutes les combinaisons qui comportaient un C2 dont le numéro était identique à l'un des C2 du trio retenu, même s'il appartenait à une liste différente dans l'expérience baseline. Après ces exclusions, le trio avec le score le plus faible était de nouveau sélectionné. Nous avons répété la procédure jusqu'à obtenir 4 trios dont le traitement était proche dans l'expérience baseline pour chaque type de C2. Étant donné qu'il y a dix types de C2 différents, nous avons ainsi obtenu une unique liste de 120 C2, contrairement aux deux listes de l'expérience baseline. Chacun des 40 trios a ensuite été divisé, de manière aléatoire, pour former les trois sous-listes de 40 items qui seraient présentées aux participants. Le fait que chaque sous-liste comporte un membre de chacun des trios permettant d'assurer un certain équilibre entre celles-ci.

5.1.2 Procédure

Dans cette expérience, nous avons appliqué le paradigme ESP (Von Ahn & Dabbish, 2004) à une tâche de création de nouveaux CAL. Ce paradigme consistait, à l'origine, en un jeu en ligne dans lequel deux joueurs devaient attribuer des labels à des images. Le but du jeu étant que les deux joueurs donnent le même label sans pouvoir se concerter. Dans ce jeu, les joueurs doivent non seulement faire appel à leurs propres représentations mais également se projeter dans celle de leur partenaire de jeu afin de réaliser une action conjointe (Clark, 1996). Ainsi, les joueurs peuvent avoir l'illusion de deviner, comme par télépathie, les pensées de l'autre joueur, ce qui explique le nom attribué à ce paradigme : ESP (Extra Sensory Perception).

Ce paradigme a été appliqué à l'étude de la convergence morphologique dans la formation du passé de pseudo-verbos en anglais dans une précédente étude de Rácz et al. (2020). Contrairement au format original du jeu de Von Ahn & Dabbish (2004) qui faisait intervenir

deux véritables joueurs humains connectés en ligne, l'étude de Rácz et al. (2020) met un joueur humain face à un robot programmé pour converger ou diverger par rapport au comportement du joueur. L'expérience se déroulant entièrement sur ordinateur, les participants ne sont pas nécessairement conscients d'interagir avec un robot. Plusieurs mesures détaillées ci-après ont été mises en place pour rendre crédible l'éventualité que le partenaire soit humain. Le contraire aurait pu limiter toute forme de convergence (Beckner et al., 2016). Dans la lignée de Rácz et al. (2020), nous avons appliqué le paradigme ESP à l'étude de la convergence morphologique, avec un partenaire artificiel, dans les CAL.

L'expérience prend la forme d'un jeu en ligne en trois phases. Dans la première phase, il était demandé aux participants de créer de nouveaux CAL à partir de paires de mots anglais dans une tâche assez similaire à celle présentée dans l'expérience de création de nouvelles formes. De la même manière que précédemment, pour chaque item, il était affiché à l'écran : une paire de mots anglais écrits en *katakana* ; les mêmes mots écrits en alphabet latin ; quatre boutons de réponses correspondant aux quatre patterns communs de formation de CAL ; et une barre de progression. Ceci est illustré sur la figure 13, ci-dessous. Les instructions pour cette phase étaient de choisir l'abréviation la plus naturelle pour cette paire de mots, sans nécessairement se soucier du sens de l'abréviation ainsi créée (voir annexe 2 pour les instructions complètes).

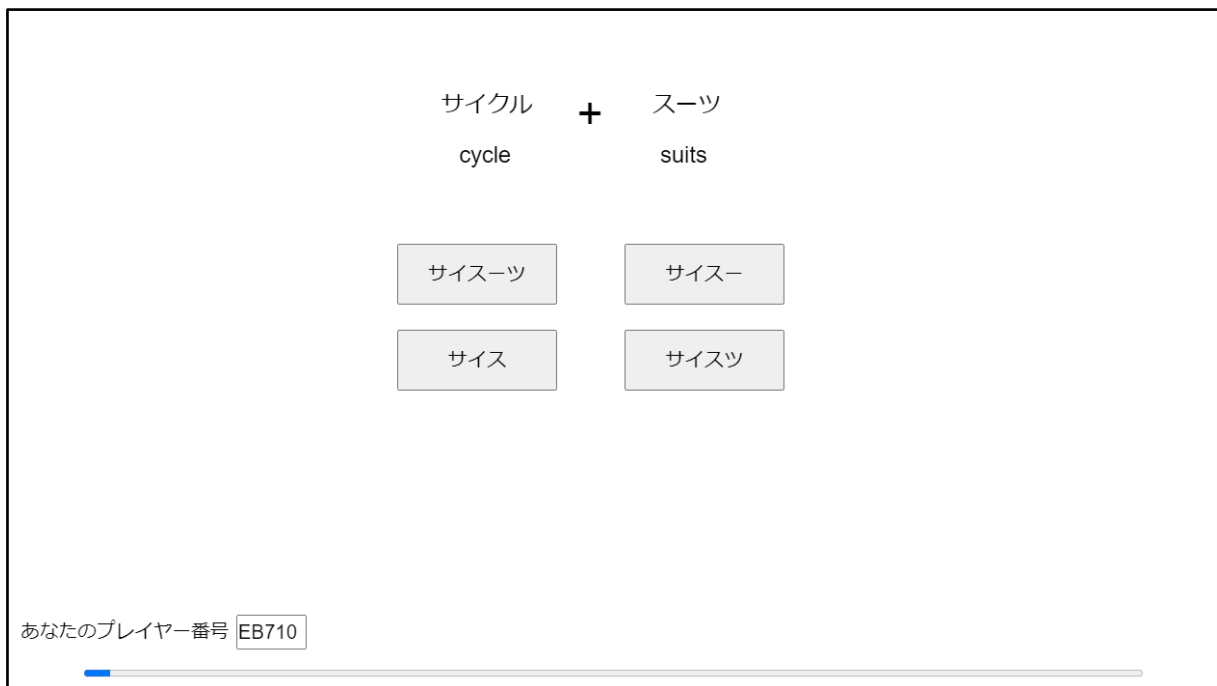


Fig.13 : exemple d'un item tel que présenté en phase 1 de l'expérience ESP.

Du point de vue des informations affichées à l'écran, la principale différence avec l'expérience de création de nouvelles formes se situe dans la présence en bas à gauche d'un numéro de joueur (une série de chiffres et de lettres générée de manière aléatoire) lorsque le participant démarrait l'expérience. Cet élément a été ajouté par souci de réalisme, notamment pour la deuxième phase de jeu qui implique la connexion à un faux partenaire de jeu.

Avant la deuxième phase de jeu, il était indiqué aux participants qu'ils allaient être connectés en ligne à un autre joueur et que le but de cette phase était, pour les deux joueurs, de donner les mêmes réponses. Les instructions stipulaient qu'en cas de réponses concordantes, les deux joueurs recevaient un point, et encourageaient à viser un score élevé, mais sans que ce score n'ait une quelconque valeur ou incidence par ailleurs. Le premier joueur devait choisir une réponse en premier, avant que le deuxième joueur ne choisisse à son tour.

Après avoir lu les instructions, le participant pouvait lancer la connexion à l'autre joueur. Toujours par souci de réalisme, une mise en scène de connexion se déroulait alors, avec un message indiquant : « La connexion est en cours (cela peut prendre jusqu'à une minute) ... ». Après quelques secondes, la fausse connexion était établie et le partenaire de jeu était présenté par son numéro de joueur (également une suite aléatoire de chiffres et de lettres), qui était ensuite visible en bas à droite de l'écran lors de la phase de jeu (voir figure 14 ci-dessous). L'utilisation de numéros de joueurs générés aléatoirement permettait de renforcer l'aspect réaliste de l'expérience tout en limitant d'éventuels effets d'ordre sociolinguistique qu'auraient pu générer l'utilisation de noms ou de pseudonymes.

En réalité, le partenaire de jeu était un robot dont le comportement s'adaptait à celui du participant dans la première phase (voir section 5.1.3 pour plus de détails). De plus, même si les instructions indiquaient que l'attribution des rôles entre premier joueur et deuxième joueur était décidée aléatoirement, le rôle du second joueur était systématiquement attribué au participant. Ainsi, pour chaque item, le participant devait attendre que le partenaire artificiel donne sa réponse avant de pouvoir répondre. Ce dernier était programmé pour « réfléchir » entre trois et neuf secondes avant de répondre, dans le but de renforcer l'illusion d'un véritable partenaire humain dont le temps de réponse est variable.

La figure 14 ci-dessous illustre cette phase de réflexion et les informations affichées à l'écran lors de la deuxième phase de jeu. En plus des informations déjà visibles dans la première phase (voir figure 13), on peut remarquer : le numéro de joueur du partenaire artificiel, en bas

à droite ; le score, affiché en bas au milieu ; et une phrase indiquant « veuillez attendre la réponse de "YG267"... » sous les boutons de réponses, verrouillés pendant ce moment de réflexion. Après ce moment de réflexion, les boutons de réponses étaient déverrouillés et le participant était invité à répondre par une indication textuelle. En cas de réponses concordantes avec le robot, une indication textuelle « Correct ! "YG267" a également choisi X » s'affichait, et le score augmentait d'un point. En cas de réponses discordantes, une indication textuelle « Dommage... vous avez choisi X mais "YG267" a choisi Y ». Dans les deux cas, les indications textuelles restaient dix secondes à l'écran, puis le jeu passait automatiquement à l'item suivant.

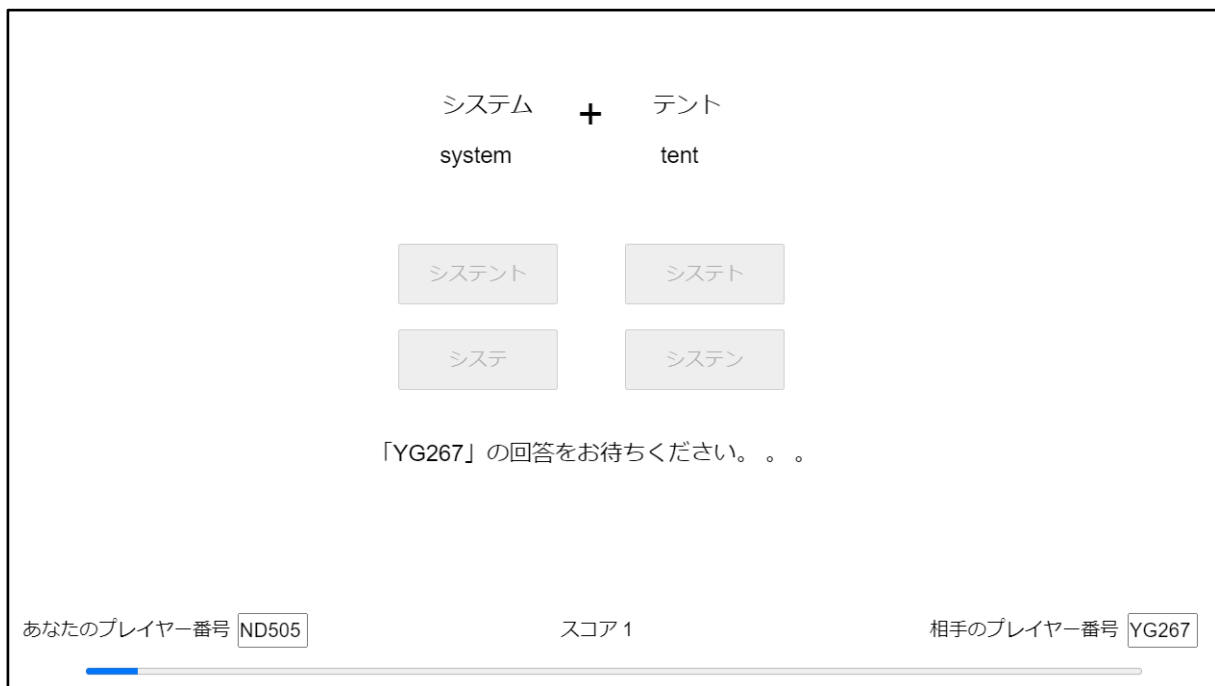


Fig.14 : exemple d'un item tel que présenté en phase 2 de l'expérience ESP, pendant le moment de réflexion du robot.

Dans la troisième phase, le participant se voyait déconnecté de l'autre joueur. Il lui était alors demandé de créer une nouvelle série d'abréviations, selon une méthode identique à la première phase. Les informations affichées à l'écran étaient donc les mêmes que pour la première phase (figure 13). Les instructions indiquaient simplement que la procédure était la même que dans la première phase et n'incitaient pas à adopter une quelconque attitude vis-à-vis des phases précédentes. Les participants étaient donc entièrement libres dans le choix de leurs réponses dans cette phase.

Pour former les items présentés à chaque participant, au début de l'expérience, chaque C2 de la liste utilisée pour l'expérience ESP (voir section précédente, 5.1.1) était associé un C1. Bien que les sets de C2 avec des différences minimales sur le plan phonologique de l'expérience baseline n'ont pas pu être conservés, chaque C1 et C2 ont préservé le numéro de set qui leur était attribué pendant celle-ci. Ainsi, dans l'expérience ESP également, les combinaisons possibles entre C1 et C2 étaient limitées aux mots partageant le même numéro de set. Cela a permis de s'assurer que les C2 apparaissaient avec les mêmes C1 dans les deux expériences et donc de limiter une éventuelle influence des C1 sur les résultats. Une fois les items formés, pour chaque participant, l'ordre d'apparition des sous-listes dans chacune des phases, ainsi que l'ordre des items à l'intérieur de ces sous-listes, étaient déterminés aléatoirement.

Pour chaque item, le pattern qu'affichait chacun des boutons de réponses était déterminé aléatoirement. Par ailleurs, afin de repérer les participants qui n'effectueraient pas la tâche sérieusement, notamment en cliquant rapidement sur n'importe quel bouton, nous avons enregistré pour chaque item le temps de réponse ainsi que la position du bouton sélectionné. Cette expérience a également été programmée sur Construct 2, un logiciel de création de jeux (multi-joueurs) pour navigateur web en 2D (Scirra, 2020 ; release 279).

5.1.3 Partenaire artificiel

Le partenaire artificiel qui entre en jeu lors de la deuxième phase est un robot dont les réponses se basent sur celles fournies par le participant dans la première phase. Il peut adopter trois comportements différents qui correspondent à nos trois conditions expérimentales. Chaque participant sera donc exposé à un seul de ces trois comportements, qui peuvent être décrits comme suit :

- La condition « sans changement » : le robot enregistre le nombre d'utilisations de chacun des patterns par le participant lors de la première phase et reproduit exactement le même nombre d'utilisations dans ses propres réponses dans la deuxième phase.
- La condition « sur-régularisateur » : le robot produira davantage de formes régulières dans la deuxième phase que le nombre produit par le participant lors de la première phase.
- La condition « sous-régularisateur » : le robot produira moins de formes régulières dans la deuxième phase que le nombre produit par le participant lors de la première phase.

Il est délicat d'établir une stratégie idéale concernant la manière dont le robot doit augmenter ou réduire le nombre de formes régulières qu'il produit en fonction de ce que le participant a produit en phase 1. Dans l'étude de Rácz et al. (2020) le partenaire artificiel était également programmé pour adopter ce type de comportement vis-à-vis du taux de régularisation des formes créées. La stratégie qui était adoptée par le robot consistait à ajouter ou enlever 40% au nombre de formes régulières produites. Par exemple, si le participant produisait 10 formes régulières en phase 1, alors le robot « sans changement » en produisait 10, le « sur-régularisateur » en produisait 14, et le « sous-régularisateur » en produisait 6. Cette stratégie, bien que satisfaisante pour mettre en lumière des effets de convergence morphologique nous semble avoir au moins deux inconvénients.

Le premier est que les participants qui régularisent peu rencontre systématiquement un robot dont le comportement diverge peu de leur propre comportement, tandis que les participants qui régularisent beaucoup rencontrent systématiquement un robot dont le comportement diverge fortement du leur. Cela peut être problématique si l'on cherche à savoir si les participants sont aussi enclins à converger de l'irrégularité vers la régularité, que l'inverse.

Un deuxième inconvénient de cette stratégie est qu'elle présente des effets de plancher et de plafond pour les participants qui sont extrêmement réguliers ou irréguliers. Plus précisément, avec 52 items par phase dans leur étude, Rácz et al. (2020) observent de tels effets pour des participants dont le taux de régularisation en phase 1 est inférieur à 6% ou supérieur à 70%. Au-delà de 70%, le robot « sur-régularisateur » ne peut fournir assez de formes régulières en respectant la proportion de 40% d'augmentation. Par exemple, si le participant crée 38 formes régulières (soit environ 73%) en phase 1, alors pour respecter une augmentation de 40% le robot devrait fournir 53 formes régulières en phase 2. Or il n'y a que 52 items. En dessous de 6% de formes régulières, une diminution de 40% ne correspond qu'à des fractions d'items, ce qui peut engendrer des différences de proportions considérables en raison des arrondis. Par exemple, face à un participant n'ayant produit qu'une seule forme régulière, le robot « sous-régularisateur » produira aussi une seule forme régulière (l'arrondi de $1 - (1 * 0.4) = 0.6$; soit 0% de diminution), tandis que face à un participant ayant produit deux formes régulières, le robot « sous-régularisateur » produira également une seule forme régulière (l'arrondi de $2 - (2 * 0.4) = 1.2$; soit 50% de diminution). Pour l'homogénéité de leurs résultats, Rácz et al. (2020) ont dû exclure de tels participants, ce qui nous semble regrettable, notamment dans notre cas d'étude où nous cherchons à évaluer l'influence de nouveaux inputs sur des locuteurs ayant de

forts biais a priori. Toutefois, il nous semble crucial de retenir que quelle que soit la stratégie adoptée, le taux de régularisation du robot doit évoluer de manière proportionnelle en fonction du nombre de formes régulières produites par le participant en phase 1.

Ainsi, la stratégie que nous avons retenue consiste à mettre les participants face à un robot qui diverge peu de leur propre comportement lorsqu'il va dans le sens de la préférence du participant (c.-à-d. lorsqu'un participant qui régularise beaucoup est appairé avec le robot « sur-régularisateur », et vice-versa) et qui diverge fortement lorsqu'il va dans le sens contraire de la préférence du participant (c.-à-d. lorsqu'un participant qui régularise peu est appairé avec le robot « sur-régularisateur », et vice-versa). Nous avons opté pour cette solution en supposant que lorsqu'un participant est déjà largement convaincu par la nécessité de créer des formes régulières ou irrégulières, un partenaire possédant une préférence similaire et même légèrement plus appuyée est en soi une forte validation des biais du participant. Au contraire, pour avoir un impact aussi fort dans le sens contraire des biais du participant, il est sans doute nécessaire de l'exposer à un comportement qui diverge plus fortement du sien.

La formule selon laquelle les deux robots calculent le nombre de formes régulières qu'ils vont produire est donc légèrement différente. En ce qui concerne le robot « sur-régularisateur », il appliquera la formule suivante :

$$\text{round}((n + (0.3 * (1 - n)) * 40)$$

Dans cette formule, n représente la proportion de formes régulières produites par le participant en phase 1. En toute logique, $(1 - n)$ correspond donc à la proportion de formes irrégulières qu'il a produites. Pour déterminer la proportion de formes régulières qu'il produira, le robot « sur-régularisateur » prendra donc 30% de cette proportion de formes irrégulières et les ajoutera à la proportion de formes régulières produite par le locuteur en phase 1. Le tout est multiplié par 40 (puisque les phases de jeu comprennent 40 items) puis arrondi pour obtenir un nombre rond d'items. Par exemple : si le participant a produit 6 formes régulières, le robot « sur-régularisateur » produira 16 formes régulières ; et si le participant a produit 30 formes régulières, le robot en produira 33. Au maximum, le robot « sur-régularisateur » augmentera de 12 le nombre de formes régulières produites par rapport au participant en phase 1, et ce dans le cas où le participant n'a produit aucune forme régulière. À mesure que le nombre de formes régulières produites par le participant augmente, l'importance de l'augmentation du nombre de formes régulières produites par le robot diminue.

Concernant le robot « sous-régularisateur », il appliquera la formule suivante :

$$\text{round}((n - (0.3 * n)) * 40)$$

Le principe reste essentiellement le même à la différence que le robot « sous-régularisateur » va soustraire 30% de la proportion de formes régulières créées par le participant en phase 1. Pour reprendre les mêmes exemples que précédemment : si le participant a produit 6 formes régulières, le robot « sous-régularisateur » produira 4 formes régulières ; et si le participant a produit 30 formes régulières, le robot produira 21 formes régulières. Au maximum, le robot « sous-régularisateur » diminuera de 12 le nombre de formes régulières produites par rapport au participant en phase 1, et ce dans le cas où le participant n'a produit que des formes régulières. A l'inverse du robot « sur-régularisateur », plus le participant produit de formes régulières en phase 1, et plus importante sera la diminution du nombre de formes régulières produites par le robot « sous-régularisateur ».

Une différence majeure entre notre cas d'étude et celui de Rácz et al. (2020) réside dans le nombre de patterns irréguliers proposés comme réponses possibles. Tandis que dans le cas de la formation du passé en anglais, seuls deux patterns sont possibles : un régulier et un irrégulier ; dans le cas des CAL, quatre patterns sont disponibles : un régulier et trois irréguliers. Comme notre expérience ESP vise à explorer l'influence de nouveaux inputs même dans une situation où les locuteurs ont eu une longue exposition, nous nous intéressons en particulier à la proportion de formes régulières que les participants vont produire dans les différentes conditions expérimentales. Par conséquent, nous avons cherché à réduire l'influence que pouvait avoir l'existence de différentes options de patterns irréguliers dans notre cas d'étude. En d'autres termes, nous ne souhaitons pas que les réponses du robot rendent un pattern irrégulier plus ou moins remarquable aux yeux du participant. Pour cela, nous avons programmé le robot pour que son utilisation des patterns irréguliers coïncide avec leurs proportions d'utilisation parmi les formes irrégulières produites par le participant en phase 1. Prenons pour exemple un participant qui, en phase 1, a produit : 20 formes suivant le pattern de Maintien ; 7 formes suivant le pattern de Suivi ; 3 suivant le pattern de Suppression ; et 10 suivant le pattern de Remplacement. Ce participant ayant produit 20 formes irrégulières au total, les proportions d'utilisation de chaque pattern irrégulier parmi les formes irrégulières produites sont donc : 0.35 pour le pattern de Suivi ; 0.15 pour le pattern de Suppression ; et 0.5 pour le pattern de Remplacement. S'il est appairé avec le robot « sous-régularisateur », celui-ci devra produire 14

formes régulières, soit 6 de moins que le participant en phase 1. Pour l'utilisation des patterns irréguliers, le robot se basera sur le nombre de formes produites dans chaque pattern par le participant et y ajoutera ces 6 formes réparties comme suit :

- 2 formes pour le pattern de Suivi, car $round(6*0.35)=2$
- 1 forme pour le pattern de Suppression, car $round(6*0.15)=1$
- 3 formes pour le pattern de Remplacement, car $round(6*0.5)=3$

Les réponses du robot « sous-régularisateur » contiendront donc : 14 patterns de Maintien ; 9 patterns de Suivi ; 4 patterns de Suppression ; et 13 patterns de Remplacement.

De la même manière le robot « sur-régularisateur » prélèvera un nombre proportionnel de formes dans les différents patterns irréguliers pour augmenter le nombre de formes régulières qu'il va produire. En prenant le même participant hypothétique en exemple, le robot « sur-régularisateur » devra produire 26 formes régulières, soit 6 de plus que le participant en phase 1. Sur la base des mêmes calculs que ci-dessus, le robot « sur-régularisateur » produira 2 patterns de Suivi ; 1 pattern de Suppression ; et 3 patterns de Remplacement de moins par rapport aux formes produites par le participant en phase 1. Ainsi, les réponses du robot « sur-régularisateur » contiendront : 26 patterns de Maintien ; 5 patterns de Suivi ; 2 patterns de Suppression ; et 7 patterns de Remplacement.

Cet exemple est résumé dans le tableau 11 ci-dessous. Pour rappel, le robot « sans changement » copie simplement le nombre de formes produites par le participant en phase 1 :

Tableau 11 : exemple de distribution des formes produites pour chacun des comportements du robot avec un participant hypothétique.

	Régulier	Irréguliers		
	Maintien	Suivi	Suppression	Remplacement
Participant phase 1	20	7	3	10
Sous-régularisateur	14	9	4	13
Sur-régularisateur	26	5	2	7
Sans changement	20	7	3	10

En raison des arrondis, il est possible que le robot se retrouve avec un total de 39 ou 41 formes à produire. Afin de corriger cela, le robot sélectionnera un pattern irrégulier de manière

aléatoire et y ajoutera ou en soustraira une forme, pour obtenir le nombre exact de 40 formes à produire.

Après avoir établi le nombre de formes qu'il doit produire pour chaque pattern, le robot doit décider quel item recevra quel traitement. Si le robot avait procédé de manière aléatoire dans cette décision, cela pourrait résulter en des comportements assez irréalistes pour un être humain. Par exemple, le robot pourrait utiliser le pattern de Maintien souvent avec les C2 /Q/ et rarement avec les C2 *sans conflit*, ce qui serait hautement inattendu de la part d'un vrai locuteur. Dans l'étude de Rácz et al. (2020), le partenaire artificiel pouvait adopter trois stratégies différentes, à savoir :

- Le comportement « typique » : régulariser prioritairement les pseudo-verbos anglais qui montraient un fort taux de régularisation dans une expérience baseline menée au préalable.
- Le comportement « atypique » : régulariser prioritairement les pseudo-verbos qui montraient un faible taux de régularisation dans l'expérience baseline.
- Le comportement « aléatoire » : choisir aléatoirement quels pseudo-verbos seraient réguliers/irréguliers.

Cet aspect du comportement du robot permettait de contrôler si la convergence morphologique était influencée simplement par le taux de régularisation de l'interlocuteur ou également par la nature des formes que ce dernier décide de régulariser. Pour des raisons techniques et théoriques, nous n'avons pas souhaité explorer davantage cet aspect. Il est montré dans les résultats de Rácz et al. (2020) que le principal prédictor du taux de régularisation d'un pseudo-verbe en phase 3 (c.-à-d. après l'interaction avec le partenaire artificiel) est son taux de régularisation dans l'expérience baseline. La typicalité du comportement du robot n'a qu'une influence mineure, bien que significative, sur les items que le participant choisira de régulariser. Bien que cette légère influence ait des implications théoriques intéressantes sur la finesse des informations qui sont prises en compte dans le phénomène de convergence morphologique, ce facteur n'en demeure pas moins secondaire.

Par ailleurs, introduire différents niveaux de typicalité dans le comportement d'un robot dans notre cas d'étude s'avérait problématique pour plusieurs raisons. Tout d'abord, contrairement aux pseudo-verbos anglais dont le plus faible taux de régularisation dans l'expérience baseline est d'environ 35%, il existe de nombreux C2 dans les CAL dont le taux

de régularisation est proche de 0% : les C2 /Q/. Un robot atypique aurait donc systématiquement régularisé des C2 /Q/ prioritairement, au détriment d'autres types de C2, ce qui serait atypique au-delà de toute vraisemblance pour un locuteur humain. Ensuite, dans un cas d'étude proposant quatre patterns de réponse au lieu de deux, il est plus difficile de programmer des comportements cohérents en termes de typicalité. Par exemple, la solution qui nous aurait semblé la plus souhaitable était de programmer le robot « typique » pour qu'il sélectionne le nombre nécessaire d'items sur lesquels utiliser le pattern de Maintien parmi ceux dont le C2 a obtenu un haut taux d'utilisation de ce pattern dans l'expérience baseline. Puis, pour les patterns irréguliers, commencer par celui pour lequel le robot doit créer le plus de formes, choisir les items dont le C2 a obtenu un haut taux d'utilisation de ce pattern et qui n'a pas déjà été sélectionné pour être formé selon le pattern de Maintien. Et ainsi de suite jusqu'au pattern pour lequel le robot doit créer le moins de formes. On pourrait alors s'interroger sur la typicalité des formes créées pour ce dernier pattern qui n'obtient automatiquement que les restes qui n'ont pas été sélectionnés au préalable. D'autres comportements sont envisageables, mais aucune solution ne nous paraissait pleinement satisfaisante pour inclure ce facteur de typicalité qui, par ailleurs, ne semble jouer qu'un rôle secondaire dans la convergence morphologique. Nous avons donc dû renoncer à l'inclusion de ce facteur dans notre étude.

La solution pour laquelle nous avons opté pour que le partenaire artificiel puisse choisir quels items recevraient quel traitement est donc la suivante :

- Le robot commence par le pattern pour lequel il doit créer le plus de forme, peu importe s'il s'agit du pattern régulier ou non.
- Le robot sélectionne la quantité nécessaire d'items en choisissant ceux qui disposent des C2 qui ont obtenu les meilleurs taux d'utilisation de ce pattern dans l'expérience baseline.
- Le robot passe au second pattern pour lequel il doit créer le plus de forme.
- Le robot sélectionne la quantité nécessaire d'items en choisissant ceux qui disposent des C2 qui ont obtenu les meilleurs taux d'utilisation de ce pattern dans l'expérience baseline et qui n'ont pas déjà été sélectionnés pour être utilisés avec un autre pattern.
- Ainsi de suite jusqu'au pattern pour lequel le robot doit créer le moins d'items.

Si l'on prend pour exemple le robot « sur-régularisateur » du tableau 11, il sélectionnera en premier lieu les 26 items possédant les C2 ayant obtenu les plus hauts taux d'utilisation du

pattern Maintien dans l'expérience baseline, et il leur appliquera ce pattern. Puis il sélectionnera les 7 items possédant les C2 ayant obtenu les plus hauts taux d'utilisation du pattern de Remplacement dans l'expérience baseline mais sur lesquels il n'a pas déjà appliqué le pattern de Maintien, et il leur appliquera le pattern de Remplacement. Ensuite, il sélectionnera les 5 items possédant les C2 ayant obtenu les plus hauts taux d'utilisation du pattern de Suivi dans l'expérience baseline mais sur lesquels il n'a pas encore appliqué le pattern de Maintien ou de Remplacement, et il leur appliquera le pattern de Suivi. Enfin, il appliquera le pattern de Suppression aux deux items restants.

Toutes les réponses du robot sont donc ainsi préparées et stockées dans sa mémoire dès le début de la phase 2. Pour chaque item, le moment de réflexion du robot est donc entièrement factice, et lorsque que le participant répond, sa réponse est simplement comparée à la réponse que le robot a en mémoire pour cet item.

5.1.4 Participants

Les participants, 287 au total, ont été recrutés sur une plateforme en ligne de crowdsourcing : Crowdworks (www.crowdworks.jp). Afin d'avoir une population homogène et des conditions de passation aussi proches que possibles dans les diverses expériences présentées dans cette thèse, nous aurions souhaité utiliser la même plateforme de crowdsourcing pour l'ensemble des expériences. Nous avons malheureusement dû transiter de Lancers vers Crowdworks pour des raisons techniques. En effet, dans la période entre les passations de l'expérience baseline pour les CAL et celles des autres expériences, la plateforme Lancers a modifié ses conditions d'utilisation, rendant son utilisation plus compliquée depuis l'extérieur du Japon, et impossible avec les moyens à notre disposition. Le fonctionnement de ces plateformes reste cependant très similaire et nous n'avons pas de raison de croire que ce changement ait pu avoir une influence considérable sur nos résultats, dans la mesure où les passations d'aucune de nos expériences ne se sont faites à cheval sur les deux plateformes.

Comme précédemment, les participants sont restés anonymes mais sont identifiables par leur pseudonyme sur le site de Crowdworks. De plus, une série de questions d'ordre démographique était posée après la tâche expérimentale. Les informations ainsi collectées pour chaque participant sont : l'âge ; le sexe ; la profession ; le lieu de résidence actuel ; la région ou le pays d'origine ; les précédentes régions ou pays de résidence ; l'âge auquel le japonais est

devenu une langue employée au quotidien ; les problèmes de vue ou de lecture ; le niveau de maîtrise des langues étrangères ; et le niveau d'éducation. L'âge moyen des participants est de 38.09 ans (écart type = 9.54). Quant au sexe : 2 participants n'ont pas souhaité l'indiquer ; 154 étaient des femmes ; et 131 étaient des hommes.

5.2 Résultats

De la même manière que dans l'expérience précédente, les temps de réponses de certains participants, particulièrement courts ou longs, sont probablement dû à un manque d'engagement ou de concentration lors de la réalisation de la tâche. Par conséquent, la première étape de notre analyse a été d'exclure ces participants non coopératifs. Cette fois encore nous avons choisi comme critère d'exclusion la transformation logarithmique du temps de réaction médian des participants. Bien qu'il suive une distribution plus éloignée de la distribution normale que dans l'expérience baseline, nous n'avons pas trouvé de critère plus approprié en ce sens. Nous avons donc préféré rester cohérent avec nos choix précédents. Toutefois, nous n'avons pas inclus dans le calcul du temps médian de réponse des participants les items de la phase 2 car, lors de celle-ci, il leur est demandé d'attendre la réponse du robot (entre trois et neuf secondes, aléatoirement) avant de répondre à leur tour, ce qui pouvait considérablement modifier leurs propres temps de réponse. Ainsi, ont été exclus de nos analyses les participants dont le log du temps de réponse médian (dans les phases 1 et 3) était éloigné de plus de deux écarts types par rapport à la moyenne du log du temps de réponse médian (dans les phases 1 et 3) de l'ensemble des participants. Huit participants ont été exclus en raison de temps de réponse trop courts, et cinq participants en raison de temps de réponse trop longs. Les résultats présentés ci-après sont donc ceux des 274 participants restants.

Lors des passations, les participants recevaient de manière aléatoire une des trois conditions expérimentales (correspondant aux trois comportements possibles pour le robot), jusqu'à ce qu'environ 90 participants aient vu chaque condition. Nous limitons simplement la possibilité d'apparition d'une condition si celle-ci atteignait cet objectif avant les autres. La répartition des participants dans les différentes conditions est donc la suivante : 91 participants pour le robot « sans changement » ; 93 participants pour le robot « sur-régularisateur » ; et 90 pour le robot « sous-régularisateur ».

Dans l'analyse de nos résultats, nous comparerons principalement les réponses des participants lors de la phase 1 et lors de la phase 3. En effet, le paradigme ESP se présente en trois phases qui sont comparables à un pré-test (phase 1), un entraînement (phase 2), et un post-test (phase 3). L'intérêt de ce type de design expérimental est de pouvoir mesurer l'influence qu'aura eu la phase d'entraînement sur le comportement des participants, et ce en comparant les résultats de la phase de pré-test à ceux de la phase de post-test. Par ailleurs, ce sont les phases de pré-test et de post-test qui présentent une tâche expérimentale aussi identique que possible et dont les résultats sont naturellement comparables. Par conséquent, seules les phases 1 et 3 seront prises en compte dans nos analyses portant sur l'influence du robot sur le comportement des participants.

Les principaux résultats de cette expérience se situent dans l'évolution de la proportion de formes régulières produites (c.-à-d. d'utilisation de pattern de Maintien) entre la phase 1 et la phase 3. Les figures 15 à 17 montrent cette évolution au cours des différentes phases de jeu en fonction du comportement du robot. La figure 15 présente les résultats pour les items *sans conflit*, la figure 16 ceux des items /R/, et la figure 17 ceux des items /Q/.

Pour les items *sans conflit*, on peut constater que les participants ont créé plus de formes régulières en phase 3 qu'en phase 1, et ce dans toutes les conditions. Cela indique que les participants créent plus de formes régulières à mesure que le jeu progresse. Le comportement du robot a également une légère influence dans le sens attendu, en cela que dans la condition « sur-régularisateur », l'augmentation du nombre de formes régulières est plus importante que pour la condition « sous-régularisateur ». Un modèle linéaire mixte généralisé a été appliqué aux données des items *sans conflit* avec : le type de pattern utilisé (régulier vs irrégulier) comme variable dépendante binaire ; le comportement du robot (« sans changement » ; « sur-régularisateur » ; « sous-régularisateur ») et la phase (1 vs 3) en facteur fixe ; et le participant et le numéro du C2 en effets aléatoires. Des tests du rapport de vraisemblance entre différents modèles ont été effectués pour évaluer la significativité des effets des différents facteurs. Les comparaisons de modèles ont révélé un effet significatif de la phase ($\chi^2=132.06$; $p<0.001$), mais un effet non significatif du robot ($\chi^2=2.7707$; $p=0.25$). L'effet de l'interaction de ces deux facteurs s'avère également significatif ($\chi^2=10.2$; $p<0.01$), en cela que l'influence du robot est plus importante en phase 3 qu'en phase 1 pour la condition « sur-régularisateur ». Afin d'observer les différences de groupes d'une phase à l'autre ou au sein d'une même phase, toutes

les comparaisons post hoc par paires ont été réalisées avec une correction de Tukey¹³. Le tableau 12 ci-dessous présente les ratios z et les valeurs p de ces comparaisons. Les trois premières lignes, en vert, nous indiquent qu'il n'y a pas d'effet significatif du robot dans la phase 1, ce qui signifie que les groupes de participants exposés aux différentes conditions sont homogènes et que les différences observées en phase 3 peuvent donc effectivement être attribuées à des facteurs propres à l'expérience. Les trois lignes suivantes, en rouge, nous indiquent que la phase a un effet significatif pour toutes les conditions. Enfin, les trois lignes suivantes, en bleu, nous indiquent que le comportement des participants en phase 3 n'est pas significativement différent en fonction du comportement du robot.

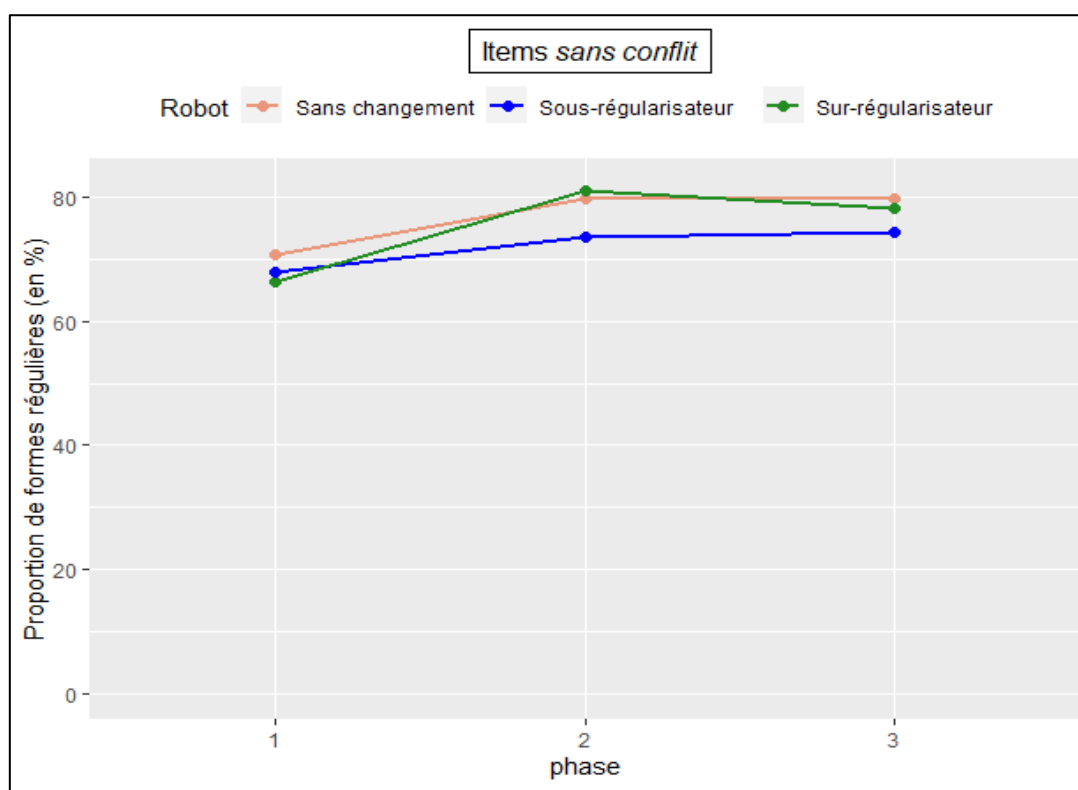


Fig.15 : évolution de la proportion (en %) de formes régulières au cours des phases de jeu, en fonction du comportement du robot, pour les items *sans conflit*.

¹³ Toutes les comparaisons post hoc de ce type présentées dans cette thèse ont été réalisées sur R à l'aide du package *lsmeans* (Lenth, 2016).

Tableau 12 : ratios z et valeurs p des comparaisons post hoc par paires avec correction de Tukey pour les items *sans conflit*.

Comparaison					
Phase	Robot	Phase	Robot	Ratio z	Valeur p
1	sans-changement	1	sur-régularisateur	-1.411	0.7205
1	sans-changement	1	sous-régularisateur	-1.07	0.8933
1	sur-régularisateur	1	sous-régularisateur	0.331	0.9995
1	sans-changement	3	sans-changement	6.438	<0.001
1	sur-régularisateur	3	sur-régularisateur	8.962	<0.001
1	sous-régularisateur	3	sous-régularisateur	4.57	<0.001
3	sans-changement	3	sur-régularisateur	-0.102	1
3	sans-changement	3	sous-régularisateur	-2.1	0.2869
3	sur-régularisateur	3	sous-régularisateur	-2.005	0.3396
3	sans-changement	1	sous-régularisateur	-4.445	<0.001
1	sans-changement	3	sous-régularisateur	1.288	0.792
3	sans-changement	1	sur-régularisateur	-4.805	<0.001
1	sans-changement	3	sur-régularisateur	3.302	<0.05
1	sur-régularisateur	3	sous-régularisateur	2.694	0.0763
3	sur-régularisateur	1	sous-régularisateur	-4.349	<0.001

Concernant les items /R/, on observe également une augmentation du nombre de formes régulières dans la condition « sans changement », une augmentation plus importante dans la condition « sur-régularisateur », et une diminution dans la condition « sous-régularisateur ». Le même type de modèle linéaire mixte généralisé que précédemment a été appliqué aux données des items /R/. Des tests du rapport de vraisemblance entre différents modèles ont également été effectués pour évaluer la significativité des effets des différents facteurs. Les comparaisons de modèles ont révélé un effet significatif de la phase ($\chi^2=21.654$; $p<0.001$), du robot ($\chi^2=9.6$; $p<0.01$) et de l'interaction de ces deux facteurs ($\chi^2=80.929$; $p<0.001$) en cela que le robot a, en toute logique, un effet plus important en phase 3 qu'en phase 1. L'évolution de la proportion de formes régulières créées au cours des différentes phases de jeu pour les items /R/ est présentée dans la figure 16 ci-dessous. Comme pour les items *sans conflit* toutes les comparaisons post hoc par paires ont été réalisées avec une correction de Tukey. Le tableau 13 ci-dessous présente les ratios z et valeurs p de ces comparaisons. Les trois premières lignes, en vert, nous indiquent que pour les items /R/, il n'y a pas non plus d'effet significatif du robot dans la phase 1, ce qui confirme l'homogénéité des groupes de participants. Les trois lignes suivantes, en rouge, nous indiquent que la phase a un effet significatif pour toutes les conditions pour les items /R/

également. Enfin, les trois lignes suivantes, en bleu, nous indiquent que l'effet du robot en phase 3 est principalement dû à la différence entre la diminution du nombre de formes régulières causée par le robot sous-régularisateur et l'augmentation provoquée par le robot sur-régularisateur.

Enfin, pour les items /Q/, on voit que la proportion de formes régulières reste proche de zéro, indépendamment de la phase ou du comportement du robot. Le même modèle statistique a été appliqué à ces données et les mêmes tests du rapport de vraisemblance ont été réalisés. Les comparaisons de modèles n'ont révélé aucun effet significatif de la phase ($\chi^2=2.9299$; $p=0.09$), du robot ($\chi^2=1.0408$; $p=0.59$) ou de l'interaction de ces facteurs ($\chi^2=2.1628$; $p=0.34$). L'évolution de la proportion de formes régulières créées au cours des différentes phases de jeu pour les items /Q/ est présentée dans la figure 17 ci-dessous.

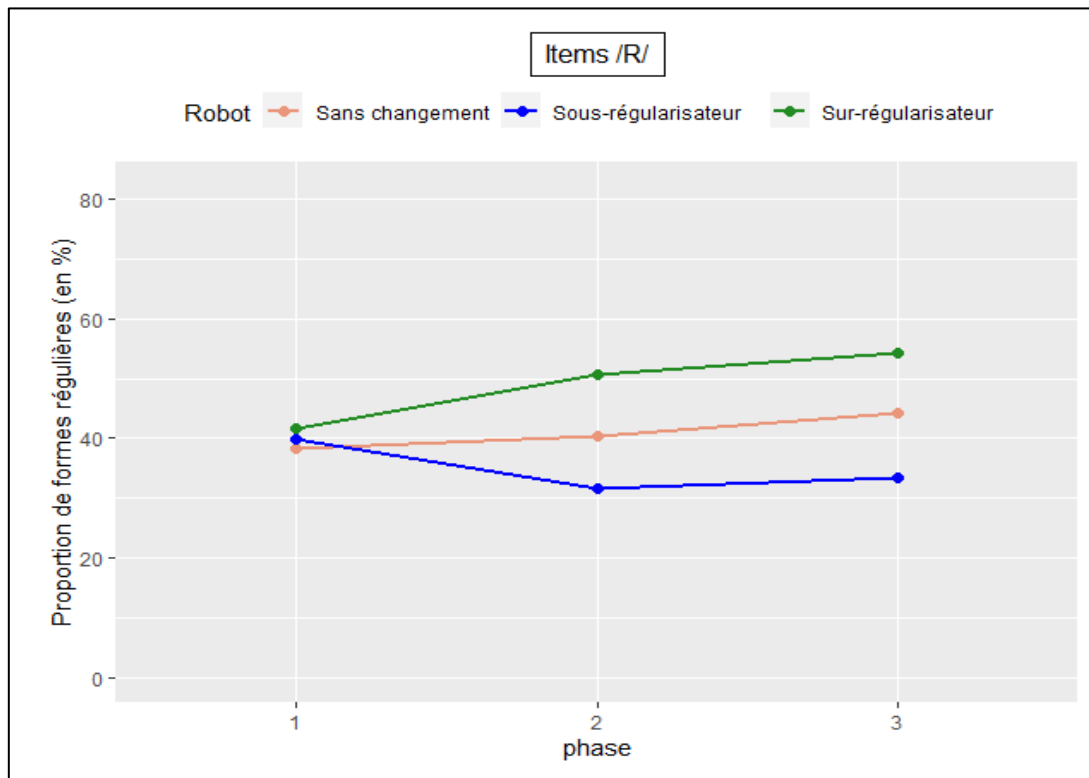


Fig.16 : évolution de la proportion (en %) de formes régulières au cours des phases de jeu, en fonction du comportement du robot, pour les items /R/.

Tableau 13 : ratios z et valeurs p des comparaisons post hoc par paires avec correction de Tukey pour les items /R/.

Comparaison					
Phase	Robot	Phase	Robot	Ratio z	Valeur p
1	sans-changement	1	sur-régularisateur	0.808	0.9662
1	sans-changement	1	sous-régularisateur	0.433	0.9981
1	sur-régularisateur	1	sous-régularisateur	-0.37	0.9991
1	sans-changement	3	sans-changement	3.434	<0.01
1	sur-régularisateur	3	sur-régularisateur	8.494	<0.001
1	sous-régularisateur	3	sous-régularisateur	-4.255	<0.001
3	sans-changement	3	sur-régularisateur	2.811	0.0557
3	sans-changement	3	sous-régularisateur	-2.693	0.0766
3	sur-régularisateur	3	sous-régularisateur	-5.498	<0.001
3	sur-régularisateur	1	sous-régularisateur	-3.748	<0.01
1	sur-régularisateur	3	sous-régularisateur	-2.128	0.2725
1	sans-changement	3	sous-régularisateur	-1.317	0.7759
3	sans-changement	1	sous-régularisateur	-0.946	0.9346
1	sans-changement	3	sur-régularisateur	4.198	<0.001
3	sans-changement	1	sur-régularisateur	-0.583	0.9922

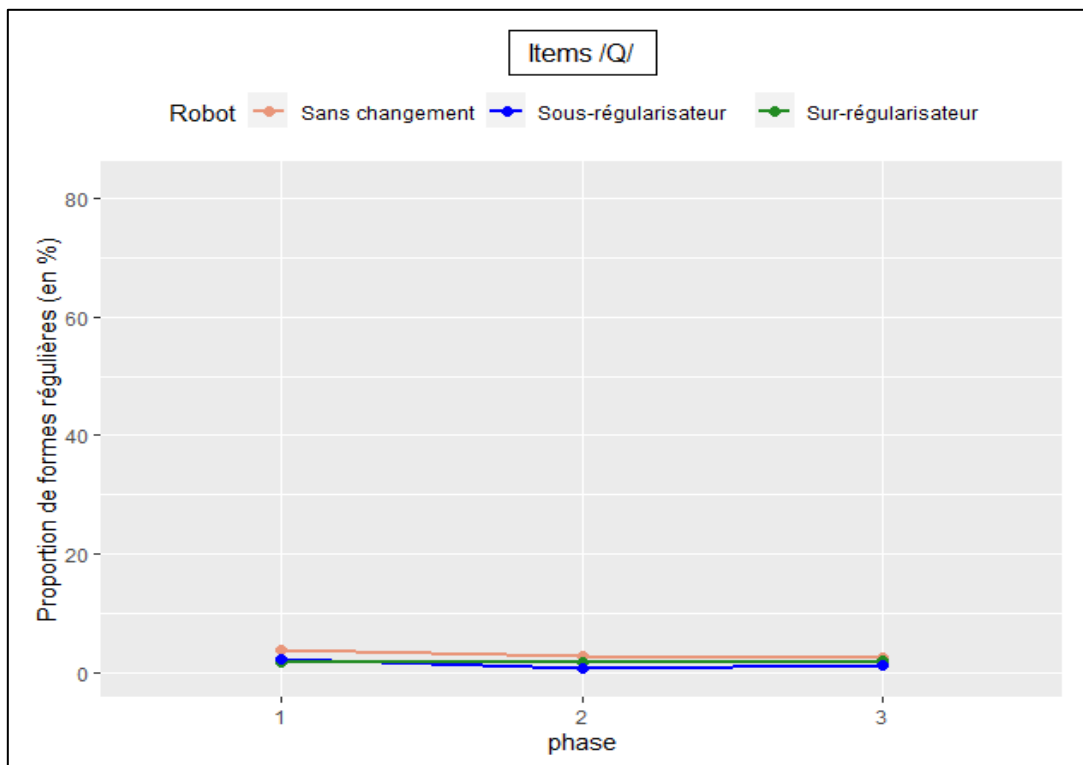


Fig.17 : évolution de la proportion (en %) de formes régulières au cours des phases de jeu, en fonction du comportement du robot, pour les items /Q/.

Pour finir, rappelons que cette expérience n'était pas élaborée pour tester l'effet d'un quelconque facteur influant sur l'utilisation des différents patterns irréguliers. En effet, le robot avait été programmé pour éventuellement influencer sur la proportion de formes régulières produites par les locuteurs, tout en conservant leurs proportions d'utilisation des différents patterns irréguliers. Pourtant, en observant la distribution des formes irrégulières créées entre les différents patterns irréguliers, on peut constater que la proportion d'utilisation de certains patterns varie dans certaines conditions entre la phase 1 et 3. La figure 18 ci-dessous montre les proportions d'utilisation de chaque pattern dans les phases 1 et 3 en fonction du comportement du robot, et ce dans l'ensemble des items. On peut y voir que les proportions d'utilisation des patterns irréguliers sont toutes légèrement inférieures en phase 3 qu'en phase 1, notamment dans les conditions « sans changement » et « sur-régularisateur ». Cependant, dans la condition « sans changement » le pattern de Suppression subit une diminution plus importante que les autres patterns. Il est invraisemblable que cette diminution soit motivée par la fréquence de ce pattern dans le lexique, où il est largement préféré aux deux autres patterns irréguliers, et elle est donc probablement en lien avec des facteurs propres à l'expérience. Il est possible que des modifications dans la proportion de formes régulières produites aient des répercussions inattendues sur l'emploi des patterns irréguliers. Notre expérience n'étant pas prévue pour adresser ces questions, nous les laissons en suspens pour le moment et espérons pouvoir les explorer dans de futurs travaux.

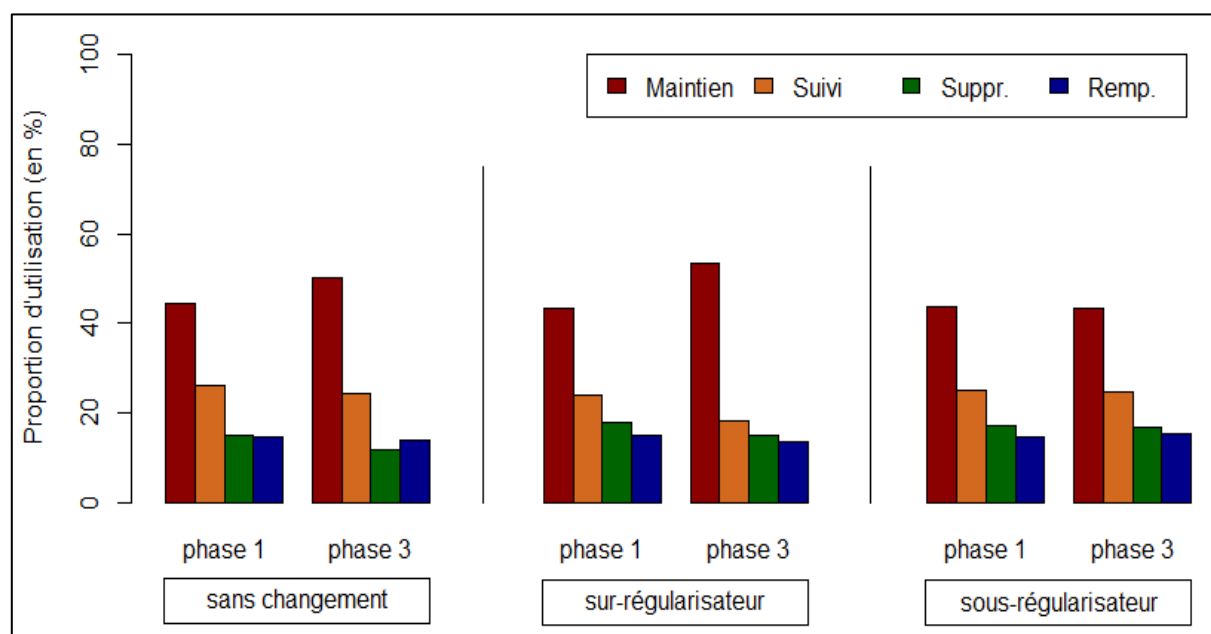


Fig.18 : proportions d'utilisation (en %) des différents patterns dans les phases 1 et 3, en fonction du comportement du robot.

5.3 Discussion

Bien qu'initialement élaborée pour explorer le phénomène de convergence morphologique, manifestée par l'influence du robot sur la régularité des formes créées, l'expérience ESP montre qu'un autre facteur joue un rôle important dans le comportement des joueurs : l'habitation à la tâche. L'effet de la phase sur le nombre de formes régulières produites indique que les locuteurs tendent naturellement vers plus de régularité à mesure qu'ils s'habituent à la tâche. On peut noter que cette observation est, en soit, un élément de plus en défaveur de l'idée du probability-matching. En effet, si les locuteurs ne font que reproduire les probabilités d'un input donné dans leur output, alors, leur comportement ne devrait pas évoluer spontanément au cours du temps. Au contraire, on devrait s'attendre à ce que l'habitation à la tâche consolide les préférences des locuteurs. Pourtant, dans les résultats de l'expérience ESP, l'effet de la phase tend à augmenter le nombre de formes régulières produites, dans toutes les conditions, et pour tous les types d'items sauf les items /Q/.

S'il semble logique que l'effet de la phase ne provoque pas d'augmentation de formes régulières pour les items /Q/, puisqu'il n'existe aucune forme régulière ou presque dans les CAL existants, on peut en revanche s'interroger sur l'effet de la phase sur les items /R/. En effet, puisqu'il y a davantage de formes irrégulières que régulières dans les CAL existant dont /R/ est la deuxième more de C2 (voir figure 3 page 36), un effet d'habitation ne devrait pas nécessairement favoriser la création de davantage de formes régulières pour ces items, surtout si l'on considère que la création de nouvelles formes suit un comportement de probability-matching. Une explication envisageable serait que les participants ne font pas de distinction catégorique entre les items *sans conflit* et les items /R/, ce qui ne serait pas sans rappeler les observations de Kapatsinski (2010) sur la manière dont un pattern général peut éclipser un pattern limité à un contexte phonologique. Contrairement à la more /Q/, la more /R/ peut assez aisément être maintenue en position finale de mot. On peut alors envisager que les locuteurs ne font pas de différence entre la more /R/ et les mores *sans conflit* et, lorsqu'ils s'habituent à la tâche, créent plus de formes régulières sauf dans les cas où le pattern régulier serait vraiment incongru (c.-à-d. pour les items /Q/). Néanmoins, nous avons vu jusqu'à présent plusieurs arguments qui vont à l'encontre de cette explication. La première raison pour laquelle nous considérons que la more /R/ occupe une place particulière dans les représentations des locuteurs est que les probabilités observées quant à son traitement dans les formes existantes diffèrent

sensiblement des mores *sans conflit*. De plus l'existence de cette distinction est soutenue par les résultats de l'expérience baseline et ceux de la première phase de l'expérience ESP, montrant que les items /R/ sont effectivement traités différemment des items *sans conflit* en cela que les proportions d'utilisation des différents patterns pour ces deux types d'items sont significativement différentes.

Afin de trouver une explication à cet effet de la phase, nous avons regardé si l'on pouvait observer un effet similaire dans les données de l'expérience baseline. Pour que cette nouvelle observation soit cohérente avec les données des différentes phases de l'expérience ESP, nous avons vérifié si la proportion de formes régulières parmi les 40 premiers items rencontrés par chaque participant dans l'expérience baseline était moins importante que celle des 40 items suivants, puis des 40 derniers items. Un modèle linéaire mixte généralisé a été appliqué à l'ensemble des données de l'expérience baseline avec : le type de pattern utilisé (régulier vs irrégulier) comme variable dépendante binaire ; la pseudo-phase (40 premiers items vs items 41 à 80 vs 40 derniers items) comme facteur fixe ; et le participant et le numéro du C2 en effets aléatoires. L'analyse du modèle ne montre pas d'effet significatif du prédicteur ($z=-0.806$; $p=0.42$).

Les données de l'expérience baseline ne montrant pas d'effet d'habituation similaire à l'expérience ESP, cela nous amène à conclure que les spécificités du paradigme ESP ont favorisé ce comportement. Le fait de placer les participants dans une situation d'interaction les a peut-être encouragés à adopter un comportement plus régulier en général. Même s'ils sont conscients que la more /R/ doit être traitée différemment des mores *sans conflit*, le fait d'adopter plus fréquemment le pattern de Maintien pour l'ensemble des items où il est permis pourrait sembler être une bonne stratégie pour obtenir le maximum de réponses communes avec leur partenaire de jeu. Malgré tout, les résultats de l'expérience ESP, notamment avec le robot « sous-régularisateur » montrent qu'un nouvel input extérieur peut contrecarrer cet effet d'habituation. Autrement dit, l'effet de la convergence morphologique est plus important que la tendance naturelle des participants à régulariser davantage à mesure que le jeu progresse.

L'effet de la convergence morphologique est particulièrement apparent dans les données des items /R/ (figure 16) puisque ce sont les seules à montrer un effet significatif du comportement du robot. L'effet de la phase est bien présent pour ces items, puisqu'on peut constater que le nombre de formes régulières augmente entre la phase 1 et la phase 3 même

dans la condition « sans changement ». Mais on observe également dans ces données que le robot « sur-régularisateur » entraîne une augmentation plus importante du nombre de formes régulières que le seul effet de la phase. Aussi, le robot « sous-régularisateur » entraîne une diminution du nombre de forme régulières produites, et ce même si l'effet de la phase agit dans la direction contraire. Pour cette dernière condition, on peut noter que l'effet de la phase est bien présent dans le fait que la diminution soit de moindre importance que l'augmentation provoquée par le robot « sur-régularisateur ».

Au vu de ces résultats, il serait tentant de croire que le robot a plus d'effet sur les items /R/ que sur les autres items, notamment *sans conflit*, précisément car ceux-ci sont ceux qui montrent le plus de variabilité. Cependant certaines spécificités de ce cas d'étude et du protocole expérimental utilisé ne permettent pas d'arriver à de telles conclusions. Comme indiqué en section 5.1.3, pour choisir quel item recevra quel traitement le robot est programmé pour appliquer les différents patterns de manière « typique », c.-à-d. en se basant sur les résultats de l'expériences baseline. Or, dans le cas des CAL, à quelques exceptions près, les meilleurs scores pour l'utilisation du pattern de Maintien reviennent aux items *sans conflit* ; les pires scores pour ce pattern reviennent aux items /Q/ ; et les items /R/ se situent au milieu. Par conséquent, demander au robot d'augmenter ou diminuer le nombre de formes régulières produites en respectant ce comportement typique revient en grande partie à augmenter ou diminuer le nombre de formes régulières produites pour les items /R/. En d'autres termes, la différence de comportement entre le robot et le participant en phase 1 se situe essentiellement dans leur traitement des items /R/, ce qui explique que la convergence soit plus importante pour ces items. Par exemple, en moyenne, les participants ont produit 18 formes régulières, dont 11 items *sans conflit*, en phase 1. Dans cette situation, le robot « sur-régularisateur » devrait donc créer 25 formes régulières, pour lesquelles il choisirait en priorité les 16 items *sans conflit* de la sous-liste utilisée dans cette phase ; et le robot « sous-régularisateur » devrait créer 13 formes régulières, qui seraient donc toutes des items *sans conflit*. Pour simplifier, pour les items *sans conflit*, le participant moyen produira 11 formes régulières et sera confronté à un robot « sur-régularisateur » qui en produira 16 ou « sous-régularisateur » qui en produira 13 (soit une différence de 3 formes entre les deux robots) ; tandis que pour les items /R/, le participant moyen produira 7 formes régulières et sera confronté à un robot « sur-régularisateur » qui en produira 9 ou « sous-régularisateur » qui en produira 0 (soit une différence de 9 formes entre les deux robots). Dans ces conditions, il est difficile d'établir si la variabilité présente dans les formes

existantes correspondant à un certain type d'items a une influence sur l'intensité de la convergence morphologique, ou s'il s'agit d'un biais causé par le protocole même de l'expérience. Nous laissons cette question en suspens pour le moment. Nous y reviendrons dans le chapitre 8 consacré à l'étude de la convergence morphologique dans des contextes où la variation est minimale.

Quoi qu'il en soit, le fait que le robot ait une influence sur le traitement des items /R/ et pas les autres types d'items nous révèle plusieurs choses. Tout d'abord, comme nous l'avons vu, la convergence morphologique peut aller à l'encontre d'un effet d'habituation et potentiellement faire pencher la balance du côté des patterns irréguliers pour le traitement de certains items plutôt que la généralisation du pattern régulier.

Ensuite, cela confirme que les items /R/ sont bien traités différemment des autres items. La convergence morphologique s'appliquant aux items /R/ ne s'étend pas aux items /Q/, ce qui indique que certaines contraintes phonologiques bloquent totalement l'effet de la convergence. Plus généralement, les résultats des items /Q/ indiquent que la productivité de certains patterns peut être entièrement verrouillée dans des contextes phonologiques particuliers. Ni le fait d'être un pattern général très productif ; ni l'effet d'habituation, ni la convergence morphologique ne peuvent accroître la productivité de ces patterns.

L'effet de convergence observé pour les items /R/ n'encourage pas non plus les participants à modifier significativement leur traitement des items *sans conflit*. En effet, pour ces items, on observe un effet significatif de la phase et de l'interaction de la phase avec le comportement du robot, mais pas du comportement du robot en lui-même. Nous avons constaté dans l'expérience baseline que les participants étaient sensibles à certaines caractéristiques phonologiques dans la création de nouvelles formes, en cela qu'ils traitent les différents types d'items de manière différente. Pour autant, il était envisageable que les participants ne soient pas sensibles à ces mêmes critères dans la manière dont ils convergent avec le partenaire artificiel. Le fait d'être encouragés à produire plus ou moins de formes régulières pour les items /R/ aurait pu inciter les participants à en produire également plus ou moins pour les items *sans conflit*, même en étant conscients que ces deux types d'items sont catégorisés différemment en raison de leurs caractéristiques phonologiques. Néanmoins, les résultats de l'expérience ESP montrent que ce n'est pas le cas. Le processus de convergence morphologique paraît sensible

aux mêmes distinctions phonologiques que celles qui guident les locuteurs dans la manière de créer de nouvelles formes de manière générale.

Bien que la convergence morphologique n'ait d'effet significatif que dans le traitement que les participants appliquent aux items /R/, le fait qu'elle influe sur la proportion de formes régulières produites nous renseigne sur une caractéristique intéressante de ce phénomène : la convergence peut influencer sur le comportement des locuteurs même lorsque ceux-ci ont été exposé à un input suffisamment important pour que les probabilités dans les formes qu'ils créent témoignent d'un comportement proche du probability-matching. Autrement dit, les locuteurs peuvent converger avec leur partenaire même lorsque celui-ci va à l'encontre de fortes probabilités dans les formes connues du locuteur. Pour rappel, nous avons vu dans l'expérience baseline que bien que les locuteurs n'aient pas assimilé le pattern de Suppression comme sous-règle à employer lorsque le pattern de Maintien n'est pas appliqué, les proportions de nouvelles formes régulières créées selon le type d'items suivent celles du lexique. Nous avons conclu que les locuteurs avaient un entraînement suffisant en ce qui concerne l'emploi du pattern régulier pour leur permettre de reproduire les probabilités de l'input dans les nouvelles formes créées. Notons qu'un tel comportement est attendu que l'on considère que l'apprentissage se fait selon un comportement de probability-matching ou que l'on considère une approche bayésienne du processus d'apprentissage. Or, l'expérience ESP ne nous renseigne pas sur l'influence de la convergence morphologique sur l'utilisation des différents patterns irréguliers, un aspect sur lequel les locuteurs n'ont qu'une expérience limitée, mais bien sur son influence sur la proportion de formes régulières créées, un aspect sur lequel les locuteurs démontrent une bonne connaissance des probabilités présentes dans les formes existantes.

Cette observation nous incite à penser que la convergence morphologique n'est pas simplement un mécanisme de gestion de l'indécision ou de la variation entre les locuteurs, mais pourrait jouer un rôle primordial dans l'émergence et l'évolution de patterns morphologiques. On parle généralement de convergence lorsque deux locuteurs qui utilisent des variantes différentes entre en contact, car c'est dans ce contexte que son effet est le plus flagrant. Mais la convergence morphologique peut être définie comme la faculté des locuteurs à s'adapter rapidement à leur interlocuteur, à de nouveaux inputs. Dans une communauté homogène de locuteurs où chacun possède un système mature et stable issu d'un long entraînement, cette faculté a pour effet de continuellement renforcer les biais des locuteurs avec de nouveaux inputs cohérents avec ceux-ci. Mais sitôt qu'un nouvel input vient contredire ces biais, les locuteurs

sont aptes à s'y adapter. L'expérience ESP montre que les nouveaux inputs produits par le partenaire artificiel ont une influence sur les participants même après que leur interaction s'est achevée (c.-à-d. en phase 3). Toutefois, cette influence restera probablement de court terme car, à l'issue de l'expérience, les participants ne seront plus exposés à ces inputs atypiques et ils se réadapteront aux inputs fournis par la communauté de locuteurs.

La convergence morphologique, ainsi définie, pourrait expliquer des situations telles que celles des CAL japonais dans lesquelles les régularités observées dans les formes existantes ne se retrouvent pas dans les nouvelles formes créées. Dans ce processus morphologique récent, tous les locuteurs ne sont pas exposés aux mêmes inputs et ne contribuent pas de la même manière à la production de nouveaux inputs. Une certaine préférence pour le pattern de Suppression pour le traitement des mores problématiques parmi les utilisateurs réguliers de CAL pourrait suffire à entraîner, par imitation, l'ensemble de la communauté vers cette tendance, même si la plupart des locuteurs n'ont à ce stade pas de claire préférence sur cet aspect de la formation des CAL.

Pour conclure, voyons les éléments de réponse qu'apporte l'expérience ESP à nos questions de recherche relatives à la construction des représentations qui guident la productivité des patterns morphologiques :

- 1) Les locuteurs reproduisent-ils simplement les probabilités de l'input dans leur output (c.-à-d. ont-ils un comportement de type « probability-matching ») ?
- 2) Sinon, quels autres facteurs entrent en jeu ?

Les résultats de l'expérience ESP apportent deux nouveaux arguments pour une réponse négative à la première question. Tout d'abord, même si les raisons provoquant l'effet de la phase sont difficiles à cerner, le fait que le comportement des participants évolue spontanément au cours de la tâche n'est pas compatible avec l'idée que ces derniers reproduisent simplement les probabilités de l'input dans leur output. Ensuite, le fait que leur comportement soit significativement influencé par les quelques nouveaux inputs qu'ils reçoivent du partenaire artificiel indique, a minima, qu'ils ne produisent pas des nouvelles formes en se basant sur les strictes probabilités de l'ensemble de l'input accumulé au cours de leur expérience de locuteur.

Cette observation indique, en réponse à la question 2), que la convergence morphologique est également à prendre en compte parmi les facteurs qui influent sur l'émergence des représentations qui guident la productivité morphologique. L'effet de la

convergence montre que les locuteurs accordent une importance particulière aux nouveaux inputs qu'ils reçoivent, au moins sur le court terme. Ainsi, ils peuvent adopter des comportements qui vont à l'encontre des probabilités présentes dans l'ensemble de l'input auquel ils avaient été exposés précédemment. On peut alors s'interroger sur l'influence qu'a le nombre de formes existantes sur l'effet de la convergence. L'expérience ESP montre que les participants peuvent converger en allant à l'encontre de patterns pour lesquels ils sont capables d'adopter un comportement proche du probability-matching, témoignant d'une exposition à ces patterns assez importante.

Pour autant, le processus de formation des CAL reste un processus morphologique devenu productif récemment dans la langue et soumis à beaucoup de variation. Cette situation linguistique particulière pourrait être propice au phénomène de convergence puisque les nouveaux inputs y sont plus importants, en termes de strictes proportions, dans l'input global des locuteurs. Dans les sections qui suivent, nous explorerons plus avant l'importance des nouveaux inputs dans l'évolution des représentations des locuteurs en étudiant l'effet de la convergence morphologique dans un processus morphologique fréquemment employé par les locuteurs et qui n'est que marginalement soumis à la variation.

Chapitre 6 – Les flexions verbales japonaises

Les expériences sur les CAL japonais nous ont permis d'explorer différents aspects relatifs à l'émergence de régularités dans des processus morphologiques récents et sujet à la variation. En comparant les formes existantes avec les nouvelles formes créées dans un cadre expérimental avec ou sans partenaire, nous avons pu observer le rôle de différents facteurs dans l'abstraction de patterns. Nous avons vu, entre autres, que les probabilités observées dans les formes existantes ne se retrouvaient pas toujours dans les formes créées, ce qui est cohérent avec l'idée que les biais a priori des locuteurs jouent un rôle plus important dans des situations où l'input auquel les locuteurs sont exposés est limité.

Cependant, on peut s'interroger sur le fait que tout ou partie des phénomènes observés dans les expériences précédentes peuvent s'expliquer par la nature récente du processus morphologique étudié. En particulier, nous avons vu avec l'expérience ESP que les participants convergeaient vers leur partenaire de jeu même pour certains aspects pour lesquels l'input auxquels ils ont été exposés est assez important pour leur permettre d'adopter un comportement proche du probability-matching. En effet, bien que les proportions de formes régulières produites par les participants sans partenaire (c.-à-d. dans l'expérience baseline et la phase 1 de l'expérience ESP) respectent les proportions observées dans le lexique, la présence d'un partenaire peut entraîner la production de plus ou moins de formes régulières. Mais cet effet de convergence observé dans les CAL est-il véritablement encouragé par le caractère récent et variable du processus morphologique étudié, ou serait-il aussi présent dans d'autres situations ?

Afin d'apporter des éléments de réponse à cette question, il convient de comparer les résultats portant sur les CAL japonais à un processus morphologique ne possédant pas ces caractéristiques. Dans les expériences présentées dans les chapitres 7 et 8, nous avons appliqué, dans la mesure du possible, les mêmes protocoles expérimentaux que précédemment à un processus morphologique bien établi dans la langue japonaise et qui ne présente que peu de variation dans les formes existantes : les flexions verbales.

Ce chapitre est consacré à la description des différents patterns que suivent les verbes japonais dans leurs flexions. Nous étudierons dans la section 6.1 les deux principaux patterns de flexions verbales existants en japonais. Puis, nous verrons en section 6.2 les fréquences de

types de ces deux patterns dans l'ensemble du lexique. En section 6.3, nous aborderons la productivité de ces patterns dans les verbes qui ont effectivement intégré le lexique japonais récemment. Enfin, en section 6.4 nous présenterons une série d'études basées sur la flexion de pseudo-verbes nous renseignant sur la productivité des différents patterns dans la création de nouvelles formes.

6.1 Les verbes japonais et leurs flexions

D'après l'analyse traditionnelle des verbes japonais (Bloch, 1946), à l'exception de deux verbes irréguliers (*suru* 'faire' et *kuru* 'venir'), ceux-ci sont répartis en deux groupes en fonction de leurs flexions et de la nature du phonème final du radical : les verbes *yodan* et les verbes *ichidan*. Les verbes *yodan* possèdent un radical s'achevant typiquement par une consonne et leur forme inaccomplie (c.-à-d. leur forme du dictionnaire) est suivie du suffixe *-u* (ex : *nom-u* 'boire'). Les verbes *ichidan* possèdent un radical s'achevant par une voyelle et leur forme inaccomplie est suivie du suffixe *-ru* (ex : *tabe-ru* 'manger'). Tous les radicaux *ichidan* ont *i* ou *e* comme voyelle finale.

Bien que les flexions verbales soient les mêmes pour tous les verbes *ichidan*, les flexions des verbes *yodan* dépendent du phonème final du radical, notamment pour leur forme accomplie. Il est important de noter que certains verbes *yodan* possèdent un radical se terminant par *ir* ou *er*, auquel cas, leur forme inaccomplie est phonologiquement indistinguishable de celle d'un verbe *ichidan*. Il existe même des paires de verbes dont la forme inaccomplie est identique, bien que l'un soit *yodan* (*kir-u* 'couper'), et l'autre *ichidan* (*ki-ru* 'porter un vêtement'). Cette ambiguïté n'est pas présente dans leurs formes fléchies, comme l'indique le tableau 14 ci-dessous qui illustre les différentes flexions possibles en japonais (modifié et traduit depuis Batchelder [1999]). Notons également que certains verbes *yodan*, comme le verbe *ka-u* dans le tableau ci-dessous, ont un radical se terminant par une voyelle. Il n'y a malgré cela pas de doute concernant le fait que ces verbes suivent le pattern *yodan*, au vu de leurs flexions. Notamment, leur forme inaccomplie est suivie du suffixe *-u* des verbes *yodan*, et non du suffixe *-ru* caractéristique des verbes *ichidan*. La voyelle finale du radical de ces verbes étant variable, nous l'indiquons par le caractère « V » dans le tableau ci-dessous.

Tableau 14 : les différentes flexions verbales en japonais.¹⁴

Type	Glose	Inacc.	Conj.	Nég.	Acc.	Volitif	Cond.
Ichidan							
		<i>x-ru</i>	<i>x</i>	<i>x-nai</i>	<i>x-ta</i>	<i>x-yoo</i>	<i>x-reba</i>
Voyelle finale							
e	‘manger’	<i>tabe-ru</i>	<i>tabe</i>	<i>tabe-nai</i>	<i>tabe-ta</i>	<i>tabe-yoo</i>	<i>tabe-reba</i>
i	‘couper’	<i>ki-ru</i>	<i>ki</i>	<i>ki-nai</i>	<i>ki-ta</i>	<i>ki-yoo</i>	<i>ki-reba</i>
Yodan							
		<i>x-u</i>	<i>x-i</i>	<i>x-anai</i>	variable	<i>x-oo</i>	<i>x-eba</i>
Phonème finale							
r	‘porter’	<i>kir-u</i>	<i>kir-i</i>	<i>kir-anai</i>	<i>kitta</i>	<i>kir-oo</i>	<i>kir-eba</i>
t	‘attendre’	<i>mat-u</i>	<i>mat-i</i>	<i>mat-anai</i>	<i>matta</i>	<i>mat-oo</i>	<i>mat-eba</i>
V	‘acheter’	<i>ka-u</i>	<i>ka-i</i>	<i>kawanai</i>	<i>katta</i>	<i>ka-oo</i>	<i>ka-eba</i>
m	‘lire’	<i>yom-u</i>	<i>yom-i</i>	<i>yom-anai</i>	<i>yonda</i>	<i>yom-oo</i>	<i>yom-eba</i>
b	‘appeler’	<i>yob-u</i>	<i>yob-i</i>	<i>yob-anai</i>	<i>yonda</i>	<i>yob-oo</i>	<i>yob-eba</i>
n	‘mourir’	<i>shin-u</i>	<i>shin-i</i>	<i>shin-anai</i>	<i>shinda</i>	<i>shin-oo</i>	<i>shin-eba</i>
k	‘écrire’	<i>kak-u</i>	<i>kak-i</i>	<i>kak-anai</i>	<i>kaita</i>	<i>kak-oo</i>	<i>kak-eba</i>
g	‘nager’	<i>oyog-u</i>	<i>oyog-i</i>	<i>oyog-anai</i>	<i>oyoida</i>	<i>oyog-oo</i>	<i>oyog-eba</i>
s	‘prêter	<i>kas-u</i>	<i>kas-i</i>	<i>kas-anai</i>	<i>kas-ita</i>	<i>kas-oo</i>	<i>kas-eba</i>
Irréguliers							
<i>suru</i>	‘faire’	<i>suru</i>	<i>shi</i>	<i>shinai</i>	<i>shita</i>	<i>shiyoo</i>	<i>sureba</i>
<i>kuru</i>	‘venir’	<i>kuru</i>	<i>ki</i>	<i>konai</i>	<i>kita</i>	<i>koyoo</i>	<i>kureba</i>

Le fait qu’un verbe japonais se terminant par *ru* puisse, en théorie, aussi bien suivre le pattern de flexion *yodan* que *ichidan* est l’aspect de ce processus morphologique qui sera au

¹⁴ Nos transcriptions des verbes japonais dans les sections qui suivent ne reflètent pas les cas d’allophonie comme il serait attendu de la romanisation Hepburn utilisée jusqu’à présent. Ces transcriptions, plus phonologiques, rendent plus transparents les paradigmes des verbes qui subissent ces allophonies. Ainsi, nous transcrivons par exemple les flexions du verbe ‘attendre’ : *matu* / *mati* / *matanai* / etc. plutôt que *matsu* / *machi* / *matanai* / etc.

cœur de nos études. En effet, ce chevauchement de paradigme pose un défi aux locuteurs dans la création de nouvelles formes. Lorsqu'ils sont confrontés à la forme inaccomplie d'un nouveau verbe (ou d'un pseudo-verbe dans un cadre expérimental) se terminant en *ru*, deux patterns de flexion se proposent à eux. Nous allons donc nous intéresser aux comportements des locuteurs quant au traitement de nouvelles formes sur cet aspect précis. Dans le reste de ce chapitre, nous verrons ce que la littérature nous apprend sur les fréquences de types et la productivité des différents patterns de flexion verbale en nous focalisant sur la potentielle compétition entre le pattern *yodan* se terminant en *r*, et le pattern *ichidan*. Puis nous présenterons dans les chapitres 7 et 8 les tâches de création de nouvelles formes, également orientées sur la compétition entre ces deux patterns, que nous avons menées.

6.2 Fréquences de types dans les formes existantes

Nous avons souligné plusieurs fois l'importance des probabilités dans l'input sur la productivité des patterns morphologiques. Aussi, une première étape dans notre étude de la productivité des patterns *yodan* se terminant en *r* et *ichidan* a été de collecter des informations sur leurs fréquences de types. Pour cela, nous nous sommes servi du dictionnaire machine-readable JMdict, un dictionnaire japonais-multilingue développé par Breen (2004). Ce dictionnaire contient environ 190 000 entrées et est largement utilisé dans des logiciels de traitement automatique du japonais (traduction instantanée, saisie intuitive, dictionnaires en ligne, etc.). La nature machine-readable de ce dictionnaire implique que sa consultation passe par un logiciel indépendant. Nous avons choisi de consulter JMdict sur le site jisho.org, un dictionnaire en ligne exploitant cette ressource qui figure parmi les recommandations de la documentation fournie par Breen. Le tableau 15 ci-dessous présente le nombre de verbes répertoriés dans JMdict en fonction du groupe auxquels ils appartiennent ainsi que leurs terminaisons. Nous donnons plus de détails pour certaines terminaisons précises qui nous donnent une idée sur les fréquences de types qui pourraient entrer en jeu dans l'issue de la compétition entre les deux patterns de flexions possibles pour les verbes se terminant en *ru*.

Tableau 15 : fréquences de types des différents paradigmes verbaux dans les formes existantes répertoriées dans le dictionnaire JMdict.

Ichidan	32.9% (3669 verbes)
<hr/>	
Par terminaison	
e-ru	29.2% (3256 verbes)
i-ru	3.7% (413 verbes)
Yodan	67.1% (7479 verbes)
<hr/>	
Par terminaison	
r-u	22.9% (2562 verbes)
er-u	1.3% (143 verbes)
ir-u	3.4% (377 verbes)
ar-u	10.3% (1154 verbes)
ur-u	3.3% (374 verbes)
or-u	4.6% (514 verbes)
t-u	2.6% (285 verbes)
V-u	7.3% (810 verbes)
m-u	5.8% (648 verbes)
b-u	0.9% (108 verbes)
n-u	0.1% (10 verbes)
k-u	8.6% (955 verbes)
g-u	1.3% (148 verbes)
s-u	17.5% (1953 verbes)
<hr/>	

On peut tout d'abord constater que les verbes *yodan* sont deux fois plus nombreux que les verbes *ichidan*, ce qui pourrait présager d'une certaine productivité du pattern *yodan*. Puis, au sein des verbes *yodan*, on observe de grande disparité dans les phonèmes finaux de ces derniers. Les verbes se terminant par *s* et par *r* représentent plus de 60% de tous les verbes *yodan*. Les verbes en *r*, en particulier, sont les plus nombreux et représentent 34% des verbes *yodan*. On peut donc s'attendre à ce que le pattern *yodan* finissant en *r* soit particulièrement

productif dans la création de nouveaux verbes. Notons toutefois que la pattern *yodan* en *r* seul est moins fréquent que le pattern *ichidan*.

Qu'en est-il alors de la productivité des deux patterns possibles pour les verbes en *ru* ? Tout d'abord, ces données confirment que la compétition ne devrait exister que pour les verbes se terminant en *eru* ou *iru*, puisque le radical d'aucun verbe *ichidan* ne s'achève sur une autre voyelle. Ensuite, on observe que les probabilités sont très différentes pour les verbes terminant en *eru* et *iru*. D'une part, quasiment tous les verbes en *eru* (plus de 95%) sont des verbes *ichidan*. D'autre part, pour les verbes en *iru*, les probabilités sont bien moins écrasantes puisque le pattern *ichidan* ne représente qu'environ 52% de ces cas. Dans une perspective de probability-matching, il serait attendu que la voyelle qui précède la terminaison d'un nouveau verbe en *ru* soit déterminante dans l'application de l'un ou l'autre des patterns : les verbes en *eru* suivraient quasiment toujours le pattern *ichidan*, les verbes en *aru*, *oru* ou *uru* suivraient systématiquement le pattern *yodan*, et les verbes en *iru* auraient une probabilité proche de 50/50 de suivre l'un ou l'autre.

6.3 Productivité des différents patterns dans les formes existantes

Le fait que les flexions verbales soit un processus morphologique établi de longue date dans la langue nous permet de nous faire une idée de la productivité réelle des différents patterns. En effet, en observant les nouveaux verbes qui sont effectivement créés et ont intégré le japonais moderne, nous pouvons d'ores et déjà observer si ces patterns sont employés pour créer de nouvelles formes, et s'ils le sont, si cela se fait dans des proportions semblables à ce que l'on observe dans l'input.

Batchelder (1999) explique que l'évaluation de la productivité des pattern *ichidan* et *yodan* s'avère difficile, car peu de nouveaux verbes entrent dans la langue japonaise. Plutôt que de créer de nouveaux verbes, la stratégie la plus souvent employée pour la création d'expressions verbales est de suffixer le verbe *suru* 'faire' aux mots que l'on souhaite employer comme verbe. Cela est particulièrement productif pour la création de verbes à partir de mots d'emprunt. Par exemple : *kurikku-suru*, littéralement 'to do a click', correspond au verbe 'to

click' en anglais¹⁵. Cependant, il est possible de créer de nouveaux verbes en japonais, notamment dans l'argot ou pour créer des formes éphémères le temps d'une conversation, des occasionalismes. Ces nouveaux verbes, suivent tous le pattern *yodan* avec consonne finale *r*. Par exemple : ググる *guguru*, est un verbe créé récemment et qui pourrait se traduire en anglais par 'to google' (c.-à-d. chercher sur Google), sur la base du mot d'emprunt グーグル *guuguru* 'Google'.

Comme dans cet exemple, beaucoup de ces verbes sont basés sur un mot qui semble tronqué juste avant une more débutant par *r* avant de se voir attribué la terminaison verbale *ru*, ce qui peut laisser penser que la présence d'un *r* dans le mot peut favoriser la création de tels verbes. Par exemple : ダブる *daburu* 'to double' est basé sur ダブル *daburu* 'double' ; et ネグる *neguru* 'to neglect' est basé sur ネグレクト *negurekuto* 'neglect'. Cependant, ce n'est pas une condition nécessaire pour la création de ces verbes, comme le montrent les exemples : コピる *kopiru* 'to copy', basé sur コピ *kopi* 'copy' ; ou パニクる *panikuru* 'to panic', basé sur パニック *panikku* 'panic'. Dans tous les cas, le *ru* final des formes verbales semble être considéré à part du reste du mot, comme une terminaison verbale. Cela apparaît notamment dans la forme orthographique de ces verbes, où *ru* est systématiquement transcrit en *hiragana* <る>, le syllabaire utilisé entre-autres pour les mots japonais autochtones, tandis que le reste du mot est transcrit en *katakana*, le syllabaire utilisé pour les mots d'emprunts. Cela est d'autant plus apparent si l'on compare l'orthographe d'un nom finissant par *ru*, comme *guuguru* et *daburu*, où *ru* est transcrit en *katatakana* <ル>, et celle du verbe correspondant, respectivement *guguru* et *daburu*, où *ru* est transcrit en *hiragana* <る>.

Nous avons vu précédemment qu'un verbe se terminant par *ru* pouvait suivre dans ses flexions aussi bien le pattern *ichidan* que le pattern *yodan* se terminant par *r*. Néanmoins, si l'on observe les nouveaux verbes qui intègrent la langue japonaise, le paradigme *ichidan* semble totalement improductif et tous les nouveaux verbes créés en japonais suivent donc le pattern *yodan* se terminant par *r*. Ainsi, le verbe *guguru* devient : *gugutta* et non **gututa* à l'accompli ; *guguranai* et non **guginai* à la négation ; etc. Puisqu'il est utilisé pour créer des occasionalismes, le pattern *yodan* finissant en *r* s'avère être, a minima, créatif. Or, la distinction

¹⁵ Les traductions des exemples sont ici données en anglais pour rendre plus transparente la correspondance entre le mot emprunté et le verbe ou l'expression verbale créée à partir de celui-ci.

entre la *créativité* et la *productivité* d'un pattern n'est pas nécessairement bien établie. Cette distinction entre, d'un côté, les processus créatifs créant des formes intentionnelles et imprédictibles, et de l'autre, les processus productifs créant des formes non intentionnelles mais prédictibles (Bauer, 2006) est bien souvent subjective (Haspelmath, 2002), voire contredite par des données expérimentales Dal & Namer (2018). Nous considérons donc qu'il n'y a pas de différence fondamentale entre *créativité* et *productivité*. Ainsi, nous dirons que la construction verbale par suffixation de *-suru* (ex : *kurikku-suru* 'to [do a] click') est très productive en japonais. Et en ce qui concerne la création de verbes, le pattern *yodan* se finissant par *r* montre une certaine productivité (ex : *guguru* 'to google'), mais le pattern *ichidan* n'est pas du tout productif. Notons que cette première observation sur les formes existantes en japonais est en contradiction avec ce qui serait attendu d'un comportement de probability-matching basé sur les fréquences de types présentées dans la section précédente. Si l'on observe les verbes récemment intégrés dans le lexique japonais, la productivité des différents patterns ne laisse place à aucune variation : les nouveaux verbes japonais suivent tous le pattern *yodan* se terminant par *r*.

6.4 Productivité des patterns dans la création de nouvelles formes

Il existe dans la littérature plusieurs études impliquant des tâches exigeant des participants qu'ils appliquent les différents patterns de flexion verbale japonais à de nouvelles formes. Bien que ces études ne soient pas nécessairement focalisées sur la productivité des différents patterns applicable aux verbes en *ru*, chacune peut nous renseigner sur ce sujet connexe. Dans ces études, il est demandé à des participants japonais de donner les formes fléchies inconnues de pseudo-verbes en se basant sur des phrases exemples incorporant d'autres flexions de ces pseudo-verbes. En somme, il leur est demandé d'appliquer un pattern de flexion verbale en leur donnant assez d'informations pour que le pattern à employer soit entièrement prédictible à partir des formes qui leur sont présentées, à la manière d'un wug test (Berko, 1958).

Les premières études de ce type sont celles de De Chene, (1982) visant à vérifier la validité du découpage morphologique des verbes *yodan* en un radical avec, en règle générale, une consonne en position finale suivi d'un suffixe *-u*, plutôt qu'un radical avec une voyelle finale suivi d'un suffixe avec une consonne initiale *-Cu*, par exemple. Bien que ce débat dépasse le cadre de cette thèse, on peut toutefois noter dans les résultats de De Chene que les participants

proposent des réponses inattendues dans leur choix des patterns de flexions verbales, malgré le fait que ceux-ci soient entièrement prédictibles au vu des formes qui leur sont présentées. Les pseudo-verbos utilisés dans l'expérience de De Chene sont des verbes *yodan* au radical se terminant par *z* ou par *b*. Notons qu'il n'existe aucun radical verbal se terminant par *z* dans la langue japonaise. Bien que de tels pseudo-verbos devraient nécessairement être considérés comme des verbes à finale consonantique (c.-à-d. des *yodan*), cela peut expliquer que les participants aient des difficultés à leur appliquer un pattern flexionnel.

Vance (1987) a montré que dans une expérience de type wug test, les locuteurs japonais pouvaient appliquer le pattern *yodan* à un pseudo-verbe *kap-u*, dont la consonne finale *p* ne se retrouve dans aucun radical verbal japonais, de la même manière qu'ils l'appliquent à des pseudo-verbos tels que *hom-u*, et *hok-u* possédant une consonne finale que l'on retrouve dans d'autres radicaux verbaux japonais. Cependant, Vance (1991) échoue à reproduire ces résultats dans une expérience similaire, dans laquelle cette fois-ci les participants ont appliqué le pattern *yodan* moins souvent sur le pseudo-verbe *kap-u* que sur *hom-u* et *hok-u*. Quoi qu'il en soit, il est intéressant de noter dans les résultats de De Chene que les participants n'appliquaient pas de manière systématique le pattern *yodan* comme attendu, ni pour les pseudo-verbos se terminant en *z*, ni pour ceux terminant en *b*. Les formes attendues étant ici définies comme celles suivant le schéma en haut de la section *yodan* du tableau 14 : *x-u* à l'inaccompli ; *x-i* à la forme conjonctive ; *x-anai* à la négation ; etc. Sur les 750 réponses collectées, seulement 46.3% (N=347) correspondaient à une simple application du pattern *yodan* sans modification du radical.

Parmi les autres types de réponses fournies, certaines consistaient à modifier de manière inattendue le radical, par élision et/ou épenthèse, avant d'y appliquer le pattern *yodan*. Par exemple, pour le pseudo-verbe *kiz-u*, produire la forme négative *kizuranai* (plutôt que *kiz-anai*), est assimilable à une modification du radical en *kizur*, par épenthèse de *ur*, avant l'application du suffixe de négation *-anai* correspondant au pattern *yodan*. De même, produire la forme *kiranai*, est assimilable à une modification du radical en *kir*, par élision du *z* et épenthèse du *r*, avant l'application du suffixe de négation *yodan*. Au moins 21,9% (N=164) des réponses observées peuvent être ainsi catégorisées comme suivant le pattern *yodan* de manière irrégulière.

De manière plus surprenante, une partie des réponses fournies consistaient à modifier le radical, par élision et/ou épenthèse, avant d'y appliquer le pattern *ichidan*. Par exemple, pour

le pseudo-verbe *kiz-u*, produire la forme négative *kizu-nai*, est assimilable à une modification du radical en *kizu* (un radical se terminant par une voyelle, donc *ichidan*), par épenthèse de *u*, avant l'application du suffixe de négation *-nai* correspondant au pattern *ichidan*. De même, produire la forme *kinai*, est assimilable à une modification du radical en *ki*, par élision du *z*, avant l'application du suffixe de négation *ichidan*. La production de ces formes est inattendue pour deux raisons. Premièrement, nous avons vu que le pattern *ichidan* n'était pas productif en japonais : les nouveaux verbes qui sont créés en japonais ne suivent jamais le pattern de flexion *ichidan*. Deuxièmement, le radical des verbes *ichidan* existants se terminent tous par la voyelle *i* ou *e*. Or, ces réponses montrent une certaine productivité du pattern *ichidan*, et ce même pour des radicaux se terminant par d'autres voyelles que *i* et *e*, comme dans l'exemple *kizu-nai* ci-dessus. Il y a donc dans l'input auxquels les locuteurs japonais sont exposés des raisons à la fois phonologiques et statistiques pour que de telles formes ne soient pas créées. Pourtant 13.6% (N=102) des réponses collectées par De Chene peuvent ainsi être catégorisées comme des formes *ichidan*. On remarque d'ores et déjà une différence entre les formes qui rejoignent le lexique commun, suivant toujours le pattern *yodan* en *r*, et le fait que les participants à cette expérience créent de manière individuelle des nouvelles formes suivant le pattern *ichidan*.

La deuxième série d'études sur l'application des patterns de flexions verbales japonais à des pseudo-verbes est celle menée par Vance (1987, 1991), brièvement évoquée précédemment. Dans une série de trois expériences, Vance vise à déterminer si certains critères phonologiques influent sur la difficulté à appliquer les patterns *ichidan* et *yodan* à des pseudo-verbes. Ces expériences prennent la forme d'un questionnaire à choix multiples dans lequel les participants lisent tout d'abord une ou plusieurs phrases intégrant un pseudo-verbe, avant de choisir la forme appropriée de ce même pseudo-verbe pour compléter une autre série de phrase à trou. Les formes sous lesquelles le pseudo-verbe apparaît dans les phrases à lire fournissent assez d'informations pour qu'il n'y ait aucun doute, en théorie, sur le pattern à employer pour créer ses flexions. Le participant doit ensuite sélectionner parmi un choix multiple des formes inédites de ce même pseudo-verbe pour compléter les phrases à trou.

La première expérience (Vance, 1987) porte sur le pattern *yodan* uniquement et les pseudo-verbes utilisés sont : *hok-u*, *hom-u*, *yon-u* et *kap-u*, parmi lesquels *kap-u* présente une consonne finale qui n'apparaît dans aucun verbe attesté en japonais. Une seule phrase à lire, et par conséquent une seule forme du pseudo-verbe, était présentée aux participants avant qu'il ne leur soit demandé de compléter les phrases à trou. Premièrement, les résultats de cette

expérience montrent que les participants n'emploient le pattern *yodan* de manière systématique pour aucun pseudo-verbe. Le pseudo-verbe obtenant le plus haut taux de réponses attendues en fonction des informations présentées est *hom-u*, avec 64% (N=128) réponses attendues sur 200. Deuxièmement, les résultats montrent que le traitement de *kap-u* ne s'avère pas significativement plus variable que celui des autres pseudo-verbes.

La deuxième expérience (Vance, 1991) vise à reproduire la première expérience en corrigeant et améliorant son protocole expérimental. Deux différences notables avec la première expérience sont à noter. Tout d'abord, afin de faciliter la tâche, deux phrases à lire, et par conséquent deux formes du pseudo-verbe, sont présentées aux participants avant de leur demander de compléter les phrases à trou. Ensuite, le verbe *mur-u* remplace le verbe *yon-u*, afin de tester si le fait que le pattern *yodan* se terminant par *r* soit productif en japonais encourage les locuteurs à appliquer ce pattern à des pseudo-verbes se terminant par *r*. Les résultats montrent une fois de plus que les participants font preuve de variabilité dans leur traitement des différents verbes, malgré une amélioration du taux de réponses attendues en fonction des informations présentées (73% [N=146] de réponses attendues sur 200 pour le verbe *hom-u*). Comme évoqué précédemment, cette fois-ci le traitement du verbe *kap-u* par les participants s'est avéré significativement plus variable. En revanche, avec 71.5% (N=143) de réponses attendues sur 200, le traitement du pseudo-verbe *mur-u* ne se montre pas significativement moins variable que *hom-u* et *hok-u*.

Notons que parmi les choix de réponses proposés aux participants dans ces deux expériences, au moins une des formes est assimilable à une modification du radical suivi d'une application du pattern *ichidan*, de la même manière que dans les expériences de De Chene (1982). Par exemple, pour la flexion du volitif du pseudo-verbe *hom-u*, les trois choix de réponse étaient : *homoo*, la forme attendue selon le pattern *yodan* ; *homuroo* une modification du radical en *homur* suivie d'une application du pattern *yodan* ; et *homuyoo*, une modification du radical en *homu* suivie d'une application du pattern *ichidan*. Parmi les 54 réponses inattendues fournies pour ce pseudo-verbe, 53.7% (N=29) suivaient le pattern *ichidan*, montrant une fois encore que ce pattern peut avoir une certaine productivité, même pour des radicaux ne se terminant pas par *i* ou *e*, et même dans des situations où les locuteurs possèdent des informations encourageant l'emploi du pattern *yodan*.

La troisième expérience (également présentée par Vance [1991]) porte sur le paradigme *ichidan*. Vance remarque en introduction de cette expérience que ses précédentes études n'abordent pas le paradigme *ichidan* parce qu'elles partent du principe qu'il serait trivial pour les locuteurs japonais d'appliquer le pattern *ichidan* à des pseudo-verbos, bien que cette supposition ne soit soutenue par aucune donnée objective. Cette réflexion peut paraître étonnante quand on sait que les verbes étant entrés récemment dans la langue japonaise suivent tous le pattern *yodan* se terminant par *r*. Mais dans l'usage et l'apprentissage de la langue, il est compréhensible que le pattern *ichidan* puisse être considéré comme trivial à appliquer. En effet, si l'on se réfère au tableau 14, on peut constater que le pattern *ichidan* s'applique à des verbes ne pouvant se terminer qu'en *iru* et *eru*, et qu'il s'applique de la même manière, sans irrégularité à ces deux types de verbes. En revanche, le pattern *yodan* s'applique à des verbes possédant des phonèmes finaux variés qui déterminent certaines de leurs flexions et peut ainsi paraître plus complexe à généraliser à de nouvelles formes. Vance a donc voulu tester si cette intuition sur la trivialité du pattern *ichidan* se vérifiait de manière expérimentale, en reprenant dans les grandes lignes le même protocole que précédemment mais en l'appliquant cette fois à deux pseudo-verbos, *riru* et *meru*, suivant le pattern de flexion *ichidan*. Une différence notable avec les expériences précédentes est que ces verbes pouvaient apparaître dans deux conditions différentes : dans la première condition, parmi les réponses aux choix, il n'y avait aucun choix qui correspondrait à une application du pattern de flexion *yodan* ; tandis que dans la seconde condition, un tel choix était présent parmi les réponses possibles. Par exemple, pour le pseudo-verbe *riru*, si la forme de négation était demandée, la forme *riranai* n'était pas parmi les réponses possibles dans la première condition, mais elle l'était dans la deuxième condition. Dans tous les cas, deux phrases à lire étaient présentées aux participants avant qu'il ne leur soit demandé de compléter les phrases à trou. L'une de ces phrases contenait le pseudo-verbe sous sa forme inaccomplie (c.-à-d. *riru* ou *meru*) qui laisse une ambiguïté sur le pattern à employer pour ses flexions. Dans l'autre phrase, le pseudo-verbe apparaissait sous la forme conjonctive, ne laissant en théorie aucun doute sur le fait que le pattern *ichidan* devait être employé¹⁶.

¹⁶ Les pseudo-verbos apparaissent plus précisément à la forme désidérative. Cette forme, pouvant se traduire par 'vouloir faire X', est formée en suffixant *-tai* à la forme conjonctive du verbe. Il n'en demeure pas moins que la partie correspondant à la forme conjonctive du pseudo-verbe ne laisse aucun doute sur le fait que le pattern *ichidan* doive être employé.

Les résultats de cette troisième expérience montrent que dans la condition où aucune forme suivant le pattern *yodan* n'était présente dans les possibilités de réponses, l'application du pattern *ichidan* à des pseudo-verbos était très fréquente, sans être systématique pour autant. Le taux de réponses attendues en fonction des informations présentées étant de 88.7% (71 sur 80) pour *riru*, et de 90% (72 sur 80) pour *meru*. En revanche le traitement des différents verbes s'avère plus variable dans la condition dans laquelle une réponse suivant le pattern *yodan* était disponible. La proportion de réponses attendues chute alors à 37.5% (30 sur 80) pour *riru*, et à 56.25% (45 sur 80) pour *meru*. Cela indique que malgré le fait que le pattern *ichidan* soit en théorie applicable à de nouvelles formes, la compétition avec le pattern productif des verbes *yodan* se terminant par *r* limite considérablement sa productivité, les participants préférant le pattern *yodan* dans la majorité des cas, quitte à ignorer un indice explicite encourageant à employer le pattern *ichidan*. Notons que le pseudo-verbe *meru* semble encourager l'emploi du pattern *ichidan* dans cette seconde condition davantage que le pseudo-verbe *riru*. Bien que la différence ne soit pas significative pour les données de Vance, nous verrons par la suite que les pseudo-verbos se terminant en *eru* entraîne parfois effectivement un emploi plus fréquent du pattern *ichidan* que les verbes se terminant en *iru*. Cela est probablement dû à l'abondance de verbes *ichidan* en *eru* (notamment si l'on compare au nombre de verbes *yodan* en *eru*) dans les formes existantes.

Les expériences de Vance confirment les observations formulées sur la base des résultats de De Chene (1982) concernant la productivité des différents pattern de flexions verbales dans la création de nouvelles formes. Bien que les verbes ayant intégré le lexique commun japonais suivent tous le pattern *yodan* en *r*, suggérant que seul ce pattern serait productif, non seulement ce pattern n'est pas employé de manière systématique à un niveau individuel par des locuteurs en milieu expérimental, mais le pattern *ichidan* montrent également une certaine productivité dans les nouvelles formes qu'ils créent. En revanche, Vance montre que malgré l'apparente trivialité de l'application du pattern *ichidan* à de nouvelles formes, il n'est pas appliqué de manière systématique même lorsque les informations présentées aux participants l'encouragent, notamment si une forme concurrente suivant le pattern *yodan* est également proposée parmi les réponses possibles.

La troisième série d'études portant sur la création nouvelles formes de pseudo-verbos japonais est celle menée par Batchelder (1999) et Batchelder & Ohta (2000). Dans une première étude, Batchelder (1999) a exploré la productivité des deux patterns communément considérés

comme productifs dans la création de nouvelles expressions verbales en japonais : le pattern *yodan* se terminant en *r* ; et la suffixation du verbe *suru*. Pour rappel, la suffixation du verbe *suru* ‘faire’ est particulièrement productive en japonais (ex : *kurikku-suru* ‘cliquer’). Dans un wug test similaire à ceux présentés dans les expériences précédentes, Batchelder a testé et comparé l’aptitude des locuteurs japonais à généraliser ces deux patterns à des pseudo-verbes. Les pseudo-verbes utilisés étaient formés à partir des 16 mêmes radicaux, 8 finissant en *o*, et 8 finissant en *i* (ex : *gano*, *nimi*, etc.). Selon les conditions, ces radicaux pouvaient être présentés comme suivant le pattern de flexion *yodan* se terminant en *r* (ex : *ganor-u*, *nimir-u*), ou présentés dans une expression verbale intégrant le verbe *suru* (ex : *gano-suru*, *nimi-suru*). Les pseudo-verbes étaient présentés aux participants dans un cours dialogue, d’abord dans leur forme inaccomplie (ex : *ganoru/ganosuru*, *nimiru/nimisuru*), puis dans leur forme accomplie (ex : *ganotta/gano-shita*, *nimitta/nimi-shita*) ou négative (ex : *gano-ranai/gano-shinai*, *nimi-ranai/nimi-shinai*). Le participant devait ensuite fournir la forme accomplie ou négative qui n’apparaissait pas dans le dialogue. Le principal résultat de cette étude est que ni la voyelle (*o* vs *i*) ni le type de verbes (*yodan* vs *-suru*) n’a d’effet significatif sur la proportion de réponses inattendues en fonction des informations présentées (12.3% en moyenne) produites par les participants. Ceci indique que les locuteurs japonais ne sont pas davantage capables d’appliquer le pattern le plus productif (*-suru*) qu’un pattern moins productif (*yodan* en *r*) à des pseudo-verbes, remettant en cause la productivité de ces patterns. Par ailleurs, Batchelder rapporte des réponses inattendues consistant à appliquer le pattern *ichidan* (ex : *ganota/ganonai*, *nimita/niminai*), réputé comme non productif, mais aucun chiffre précis n’est fourni sur ce type de réponses.

Dans une expérience suivante, Batchelder & Ohta (2000) comparent la productivité des patterns *ichidan* et *yodan* finissant en *r* (désormais simplement « *yodan* »). L’expérience est similaire à la précédente, jusqu’aux radicaux verbaux utilisés. Cependant, cette fois, 8 des radicaux finissent en *e*, et 8 des radicaux finissent en *i* (ex : *gane*, *nimi*, etc.). Aussi, selon les conditions, ces radicaux pouvaient être présentés suivant le pattern de flexions *yodan* (ex : *ganer-u*, *nimir-u*) ou le pattern *ichidan* (ex : *gane-ru*, *nimi-ru*). De la même manière que précédemment, ces pseudo-verbes étaient présentés dans un dialogue sous leur forme inaccomplie (ex : *ganeru*, *nimiru*), puis accomplie (ex : *ganetta/gane-ta*, *nimitta/nimi-ta*) ou négative (ex : *gane-ranai/gane-nai*, *nimi-ranai/nimi-nai*). Le participant devait ensuite fournir la flexion inaccomplie ou négative qui n’apparaissait pas dans le dialogue. Le principal résultat

de cette étude est que pour les radicaux présentés comme suivant le pattern *ichidan*, le pattern *ichidan* était employé dans 52.7% des cas, tandis que le pattern *yodan* était appliqué dans 44.3% des cas. En revanche, pour les radicaux présentés comme suivant le pattern *yodan*, le pattern *yodan* était employé dans 75.5% des cas, tandis que le pattern *ichidan* était appliqué dans 19.6% des cas. Notons que les participants n'étaient pas limités par un choix multiple de réponses dans cette expérience. Les réponses restantes ne peuvent être considérées comme suivant l'un ou l'autre de ces deux patterns. Par ailleurs, les résultats montrent que la voyelle finale du radical (*i* vs *e*) ne provoque pas de différence significative dans le traitement des pseudo-verbos.

Dans la lignée des autres études présentées dans cette section, les expériences de Batchelder (1999) et Batchelder & Ohta (2000) viennent confirmer que la productivité des différents patterns dans le cadre de tâches de création de nouvelles formes ne correspond pas à ce que l'on peut observer dans les verbes ayant récemment intégré le lexique japonais. De nouveau, le pattern *yodan* dont on pouvait prédire qu'il serait le seul productif n'est pas employé de manière systématique, et le pattern *ichidan*, supposément improductif, est pourtant employé par les participants dans la création de nouvelles formes. Ces résultats montrent également que le pattern consistant à créer une expression verbale avec *-suru*, le plus productif au vu des formes intégrées récemment au lexique commun, ne s'avère pas plus productif que le pattern *yodan* au niveau des créations individuelles des participants.

Globalement, les résultats de ces différentes études font ressortir non seulement les similitudes mais surtout les différences entre la productivité des différents patterns au niveau des formes existantes dans le lexique commun et les nouvelles formes créées par les locuteurs. Tout d'abord, on peut constater que la productivité du pattern *yodan* est supérieure à celle du pattern *ichidan* dans les deux cas, mais les résultats des expériences ne proposent pas une image aussi tranchée que celle obtenue par la simple observation des nouvelles formes qui sont entrées dans la langue récemment. Bien qu'aucun nouveau verbe du lexique commun ne suive le pattern *ichidan*, il arrive qu'il soit utilisé dans la création de nouvelles formes, même dans des situations où les informations disponibles encouragent l'emploi du pattern *yodan*. Réciproquement, le pattern *yodan*, que suivent tous les verbes récemment intégrés à la langue, n'est pas employé de manière systématique dans la création de nouvelles formes. Ensuite, nous avons vu dans la section précédente que le pattern de suffixation du verbe *-suru* était très productif dans le lexique commun. Il est par ailleurs structurellement facile à reproduire puisqu'il consiste à simplement suffixer les mêmes formes au nom que l'on souhaite transformer en expression

verbale. Pourtant, ce pattern ne montre pas de productivité supérieure au pattern *yodan* dans l'expérience de Batchelder & Ohta (2000). On peut donc constater à travers cette série d'expériences et l'observation du lexique que les participants n'adoptent pas un comportement de probability-matching lorsqu'ils créent de nouveaux verbes. Cela est d'autant plus apparent lorsque l'on constate que dans la création de nouvelles formes, les verbes en *eru* ne suivent pas le pattern *ichidan* plus souvent que les verbes en *iru*, malgré un écart considérable dans les probabilités de l'input concernant ces deux types de verbes. Ainsi, le cas des flexions verbales est intéressant en cela qu'il présente deux patterns en compétition dont la productivité diffère selon que l'on observe les formes existantes et les nouvelles formes créées dans un cadre expérimental, et ce dans un processus morphologiques bien établi et fréquemment employé par les locuteurs.

Cependant toutes les expériences que nous avons présentées dans cette section ont pour point commun d'encourager l'utilisation de certains patterns par les informations qui sont présentées aux participants. Or, il est difficile d'évaluer objectivement la productivité des patterns dans la création de nouvelles formes lorsque les conditions expérimentales orientent le comportement des participants vers l'emploi de certains patterns. En ce sens, il serait plus pertinent de tester quel pattern serait sélectionné lorsqu'aucune information ne favorisant l'emploi d'un pattern ou l'autre ne serait présentée aux participants. L'expérience décrite ci-après s'attèlera donc à cette tâche.

Chapitre 7 – Expérience baseline - flexions verbales : flexion de nouvelles formes verbales japonaises

Les expériences précédentes sur l'application des différents paradigmes verbaux à des pseudo-verbos nous présentent quelques indices sur la productivité des différents patterns. Elles soulignent un écart entre les probabilités observées dans les nouveaux verbes effectivement employés par la communauté des locuteurs et la compétence des locuteurs à appliquer les différents patterns à de nouvelles formes. Cependant, aucune de ces expériences ne met les patterns *ichidan* et *yodan* en compétition sans encourager l'emploi de l'un ou l'autre, ce qui permettrait d'évaluer la productivité relative de ces deux patterns. L'expérience présentée dans cette section a deux objectifs : le premier est de proposer une tâche de création de nouvelles formes n'orientant pas le choix des locuteurs, afin de déterminer si la productivité des différents patterns est basée sur les probabilités observées dans les verbes qui sont entrés dans la langue récemment et/ou celles observées dans l'ensemble des formes existantes ; le second est d'obtenir des données sur la manière dont une série de pseudo-verbos sont traités par les locuteurs japonais sans influence extérieure, afin de servir de baseline à l'expérience sur la convergence présentée dans le chapitre 8.

7.1 Méthode

7.1.1 Matériel

Dans cette expérience, nous avons présenté aux participants les formes inaccomplies de pseudo-verbos se terminant en *ru*, ce qui laisse l'ambiguïté sur le pattern à employer pour former ses flexions. Chaque pseudo-verbe était présenté dans un court dialogue écrit entre deux amis (voir annexe 10 pour la liste complète des dialogues). Le premier personnage utilisait le pseudo-verbe à la forme inaccomplie, puis le second personnage répondait en utilisant le même pseudo-verbe à la forme négative ou accomplie. Dans la phrase du second personnage, la place du pseudo-verbe était remplacée par un trou. Afin de pouvoir comparer les résultats des expériences sur les flexions verbales avec ceux des expériences sur les CAL, nous avons créé

une liste de 120 pseudo-verbos qui ont servi à former nos items (voir annexe 9 pour la liste complètes des pseudo-verbos utilisés). Nos items sont ici définis comme une combinaison d'un pseudo-verbe et d'un dialogue dans lequel il apparaît. Tous les pseudo-verbos utilisés se terminent en *ru*, et la liste comprends autant de pseudo-verbos en *iru* ou *eru*, que de verbos en *aru*, *uru*, ou *oru*. L'hypothèse derrière ce choix est que, dans ce contexte d'ambiguïté, la dernière voyelle du radical pourrait avoir un effet important, en cela que les verbos finissant en *iru* et *eru* auraient plus de chances de suivre le pattern *ichidan*, puisque seules ces terminaisons sont possibles dans les *ichidan* existants. Les 60 pseudo-verbos de ces deux catégories (*iru/eru* vs *aru/oru/uru*) sont répartis équitablement en ce qui concerne la dernière voyelle du radical. Par ailleurs, afin d'éviter d'avoir trop de formes similaires, deux longueurs sont possibles pour les pseudo-verbos : trois mores ; ou quatre mores. Les formes possédant la même terminaison sont réparties équitablement en termes de longueur, entre trois ou quatre mores. Ainsi, la liste de 120 pseudo-verbos contient :

- 15 pseudo-verbos en *iru* d'une longueur de **trois mores**.
- 15 pseudo-verbos en *iru* d'une longueur de **quatre mores**.
- 15 pseudo-verbos en *eru* d'une longueur de **trois mores**.
- 15 pseudo-verbos en *eru* d'une longueur de **quatre mores**.
- 10 pseudo-verbos en *aru* d'une longueur de **trois mores**.
- 10 pseudo-verbos en *aru* d'une longueur de **quatre mores**.
- 10 pseudo-verbos en *uru* d'une longueur de **trois mores**.
- 10 pseudo-verbos en *uru* d'une longueur de **quatre mores**.
- 10 pseudo-verbos en *oru* d'une longueur de **trois mores**.
- 10 pseudo-verbos en *oru* d'une longueur de **quatre mores**.

Afin de nous assurer que les pseudo-verbos utilisés ont une structure phonologique à la fois plausible et diversifiée, nous avons générés la partie des pseudo-verbos qui précède les terminaisons décrites ci-dessus à l'aide d'un générateur de notre création se basant sur la fréquence d'apparition des phonèmes japonais. Pour le choix des voyelles qui composent le radical des pseudo-verbos, en dehors de la voyelle de la terminaison qui respecte les critères décrits ci-dessus, le générateur utilisait les fréquences présentées dans la figure 19 ci-dessous, reprise et modifiée de Labruno (2006a) et basée sur les données de Campbell (1999). Ces

données sont issues d'un comptage portant sur des énoncés en japonais moderne issus d'un corpus de 503 phrases prélevées dans des journaux ou magazines et comportant tout type de mots (noms, verbes, etc.).

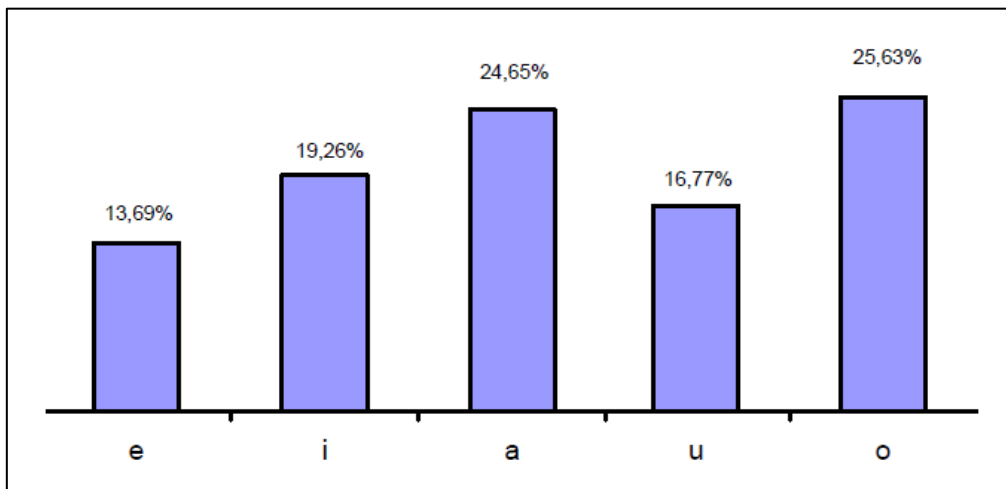


Fig.19 : fréquence d'apparition des voyelles fournies au générateur de pseudo-verbes.

Étant donné que les consonnes japonaises n'ont pas la même fréquence d'apparition en position initiale et médiale de mot, les données sur lesquelles se base le générateur pour le choix des consonnes sont celles présentées dans la figure 20, tiré de Labrune (2006a), adapté de Kokuritsu Kokugo Kenkyūjo (1984). Ces données représentent les fréquences d'apparition des différentes consonnes dans les noms bimoraïques de la strate *Yamato* en japonais moderne, en fonction de la position (initiale vs médiale) dans le mot.

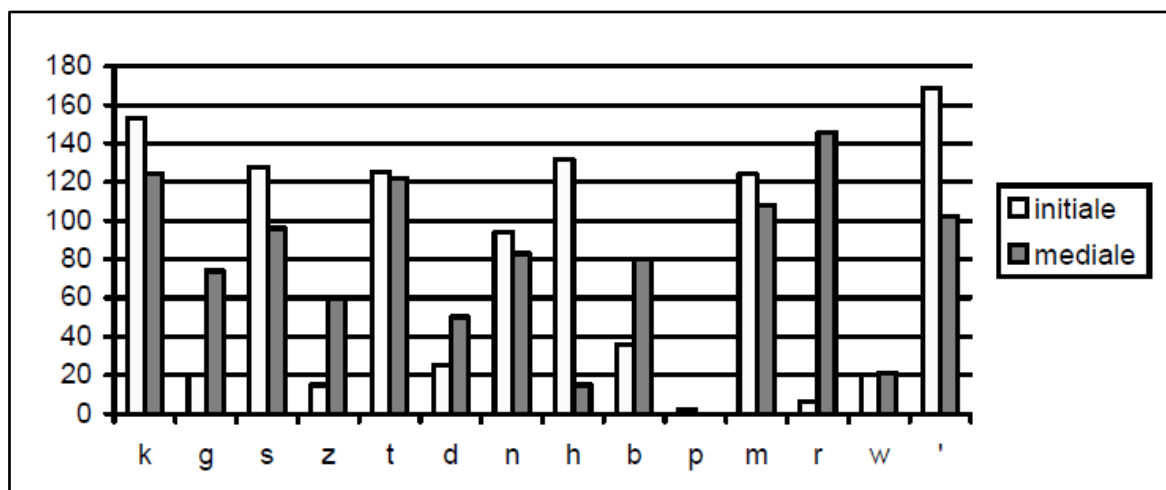


Fig.20 : fréquences d'apparition des consonnes fournies au générateur de pseudo-verbes. Le symbole « ' » représente une attaque vide.

Parmi les verbes japonais de quatre mores, il est possible que la deuxième more soit la nasale homorganique /N/, comme dans le verbe *kaNjiru* ‘ressentir’. Pour permettre au générateur de créer de tels verbes, nous avons ajouté la possibilité que /N/ soit la consonne de la deuxième more, avec une probabilité d’environ 1,5% (à peu près équivalente à celles des consonnes les moins fréquentes dans cette position). Le cas échéant, le générateur supprimera la voyelle de la deuxième more, puisque /N/ est un segment moraïque à part entière et ne peut être accompagné d’une voyelle. De même, la source utilisée ne proposait pas de données pour la semi-consonne *y*. La même probabilité d’apparition d’environ 1,5% lui a donc été attribuée.

Enfin, le générateur respecte une série de contraintes phonologiques et phonotactiques du japonais. Tout d’abord, la consonne *w* ne peut être suivie que de la voyelle *a*, et la consonne *y* ne peut être suivie que de *a*, *u*, ou *o*. Si le hasard fait suivre ces consonnes d’une voyelle incompatible, le générateur conservera la consonne et sélectionnera une nouvelle voyelle aléatoirement, jusqu’à en obtenir une compatible. Ensuite, le générateur respecte une contrainte phonologique qui agit dans l’ensemble du lexique à l’exception des mots d’emprunt et qui impose que les morphèmes japonais ne contiennent pas plus d’une obstruante voisée. Si, en se basant sur les fréquences d’apparition fournies, le générateur crée un verbe contenant deux obstruantes voisées ou plus, il le détruit et en génère un nouveau, jusqu’à obtenir une forme licite. Le bon fonctionnement du générateur entraîne inévitablement qu’il génère régulièrement des verbes déjà existants en japonais. Ces formes ont été exclues manuellement et de nouvelles ont été générées jusqu’à obtenir les 120 pseudo-verbes qui nous ont servi d’items.

Une fois les 120 formes inaccomplies des pseudo-verbes ainsi créées, les quatre flexions utilisées dans cette expérience ont également été créées par simple application des patterns de flexions verbales : négation *ichidan* (ex : *tadakeru* > *tadakenai*) ; négation *yodan* (ex : *tadakeru* > *tadakeranai*) ; accomplie *ichidan* (ex : *tadakeru* > *tadaketa*) ; et accomplie *yodan* (ex : *tadakeru* > *tadaketta*). Nous avons ensuite cherché chacune de ces formes dans le corpus *chūnagon* (voir section 4.1.1), afin de nous assurer qu’il ne retournerait aucune occurrence. A noter que les chevauchements de paradigme ont imposé de rechercher chacune de ces formes individuellement. En effet, bien qu’il n’existe pas de verbe *tadakeru* en japonais, et que cette forme ne retourne aucun résultat dans le corpus *chūnagon*, la forme *tadaker-anai*, sa flexion de négation *yodan*, pourrait également correspondre à un hypothétique verbe *ichidan* à la négation : *tadakeru* > *tadakeru-nai*. De même, la forme *tadaketta*, sa flexion accomplie *yodan*, pourrait correspondre à la forme accomplie de deux verbes *yodan* hypothétiques :

tadake-u et *tadaket-u*. Vérifier qu’aucune de ces formes n’apparaît dans le corpus *chūnagon* permet non seulement de s’assurer que le pseudo-verbe est bien un pseudo-verbe, mais également qu’aucun chevauchement de paradigme avec un verbe existant ne pourrait influencer les participants à opter pour une forme plutôt qu’une autre. En suivant le même raisonnement, les autres flexions de bases (conjonctif, volitif et conditionnel) ont également été vérifiées, pour couvrir les cas où un verbe existant mais rare n’apparaîtrait, par exemple, qu’au conditionnel dans le corpus *chūnagon*, et pourrait éventuellement influencer sur nos résultats par un chevauchement de paradigme.

Enfin, une vigilance particulière a été portée aux verbes en *eru*, car certaines flexions verbales consistent à modifier le radical d’un verbe (*ichidan* ou *yodan*) de sorte à créer un verbe *ichidan* finissant en *eru*. Ce nouveau radical pourra ensuite être suivi par les flexions de base présentées dans la section *ichidan* du tableau 14. Parmi ces flexions, on trouve : le passif ; le potentiel (c.-à-d. ‘pouvoir faire X’) ; le factitif (c.-à-d. ‘faire faire X’) ; et le factitif-passif (c.-à-d. ‘être contraint de faire X’). Le tableau 16 ci-dessous présente ces flexions. Notons que le factitif-passif possède une forme abrégée dans le langage courant dans laquelle le *e* et le *r* (transcrits entre parenthèses) à la jonction des suffixes de factitif et de passif sont élidés.

Il existe donc un risque qu’un pseudo-verbe en *eru* soit perçu comme la forme potentielle, passive, factitive, ou factitive-passive d’un autre verbe. Par exemple, *tadakeru* pourrait être la forme potentielle d’un hypothétique verbe *tadaku*. Nous avons donc vérifié qu’aucun verbe existant ne pouvait être retro-formé à partir des pseudo-verbes en *eru* en suivant ces flexions. Après ces vérifications sur le corpus *chūnagon*, nous avons également demandé une seconde vérification d’un locuteur natif, qui nous a confirmé l’inexistence de ces formes, ou de formes pouvant être confondues avec celles-ci par chevauchement de paradigme.

Tableau 16 : les flexions verbales du japonais entraînant la création d'un radical *ichidan* se finissant par *e*.

<i>Ichidan</i>			
	Modèle	Exemple	Glose
Inaccompli	<i>x-ru</i>	<i>tabe-ru</i>	manger
Passif	<i>x-rare-ru</i>	<i>tabe-rare-ru</i>	être mangé
Potentiel	<i>x-rare-ru</i>	<i>tabe-rare-ru</i>	pouvoir manger
Factitif	<i>x-sase-ru</i>	<i>tabe-sase-ru</i>	faire manger
Factitif-passif	<i>x-sas(e-r)are-ru</i>	<i>tabe-sas(e-r)are-ru</i>	être contraint de manger

<i>Yodan</i>			
	Modèle	Exemple	Glose
Inaccompli	<i>x-u</i>	<i>yom-u</i>	lire
Passif	<i>x-are-ru</i>	<i>yom-are-ru</i>	être lu
Potentiel	<i>x-e-ru</i>	<i>yom-e-ru</i>	pouvoir lire
Factitif	<i>x-ase-ru</i>	<i>yom-ase-ru</i>	faire lire
Factitif-passif	<i>x-as(e-r)are-ru</i>	<i>yom-as(e-r)are-ru</i>	être contraint de lire

Comme mentionné précédemment, les verbes étaient présentés dans un court dialogue écrit entre deux amis. Dans ces dialogues, le premier personnage utilisait le pseudo-verbe à la forme inaccomplie, puis le second personnage répondait en utilisant le même pseudo-verbe à la forme négative ou accomplie. Les deux personnages s'expriment dans un japonais standard familier, comme le ferait deux jeunes amis parlant d'activités ou d'actions que l'on effectue dans le vie de tous les jours. Voici un exemple de dialogue dans lequel une forme accomplie est attendue pour remplacer le trou dans la phrase du deuxième personnage :

En japonais :

萌 : 明日[pseudo-verbe]のはどうか？

翔太 : 三日前_____から、ちょっと早すぎると思う。

Traduction :

Moe : Ça te dirait de [pseudo-verbe] demain ?

Shōta : Comme on a déjà _____ il y a trois jours, je pense que c'est un peu trop tôt.

Les dialogues dans lesquels peuvent apparaître les pseudo-verbes dans l'expérience sont également au nombre de 120. Ils sont divisés en 60 dialogues pour lesquels une forme accomplie doit venir compléter la phrase à trou, et 60 dialogues pour lesquels c'est une forme négative qui doit être utilisée. Par conséquent, chaque participant devait créer 60 formes accomplies et 60 formes négatives.

Afin que chaque participant crée autant de formes de chaque flexion pour chaque type de pseudo-verbes (en termes de longueur et de terminaison), nous avons réparti manuellement et de manière équilibrée les pseudo-verbes en deux catégories : ceux dont la forme n°1 serait l'accompli et la forme n°2 la négation et ceux pour lesquels ce serait l'inverse. Les participants ne créaient que l'une des deux formes pour chaque pseudo-verbe. Ainsi, les cinq premiers pseudo-verbes en *aru* de trois mores ont l'accompli pour forme n°1 et la négation pour forme n°2, tandis que c'est l'inverse pour les cinq autres pseudo-verbes en *aru* de trois mores. Notons qu'un équilibre parfait ne peut pas être obtenu pour les types de pseudo-verbes possédant 15 représentants (ex : pseudo-verbes en *iru* de trois mores). Ainsi, seuls les sept premiers verbes en *iru* ont l'accompli pour forme n°1 et la négation pour forme n°2, tandis que c'est l'inverse pour huit autres pseudo-verbes en *iru* de trois mores. Le détail de la répartition des pseudo-verbes dans ces deux catégories est disponible dans l'annexe 9.

7.1.2 Procédure

Dans cette expérience diffusée en ligne, il était demandé aux participants de choisir parmi deux propositions de réponses (pattern *ichidan* vs pattern *yodan*) la flexion de pseudo-verbes en se basant sur leur forme inaccomplie. La forme inaccomplie du pseudo-verbe leur était présentée dans la première phrase d'un dialogue écrit, dans lequel elle était affichée en bleu et en gras afin de la rendre facilement identifiable. Dans sa forme inaccomplie, le pseudo-verbe était écrit entièrement en *hiragana*. La figure 21 ci-dessous présente un item tel qu'affiché à l'écran lors de la passation de l'expérience. Les instructions invitaient les participants à sélectionner la forme verbale qui leur semblait la plus naturelle, sans prêter trop attention au sens des verbes (voir annexe 3 pour les instructions complètes). Deux boutons de réponses correspondant à l'application du pattern *ichidan* et *yodan* sur le pseudo-verbe, également écrites en *hiragana*, étaient également visibles. Dans toutes les phases de jeu et pour chaque item, la répartition du pattern *ichidan* et *yodan* entre les deux boutons de réponses était déterminée

aléatoirement. De plus, afin d'identifier les participants non coopératifs, la position du bouton sélectionné et le temps de réponse pour chaque item étaient enregistrés. Une barre de progression était présente en bas de l'écran.

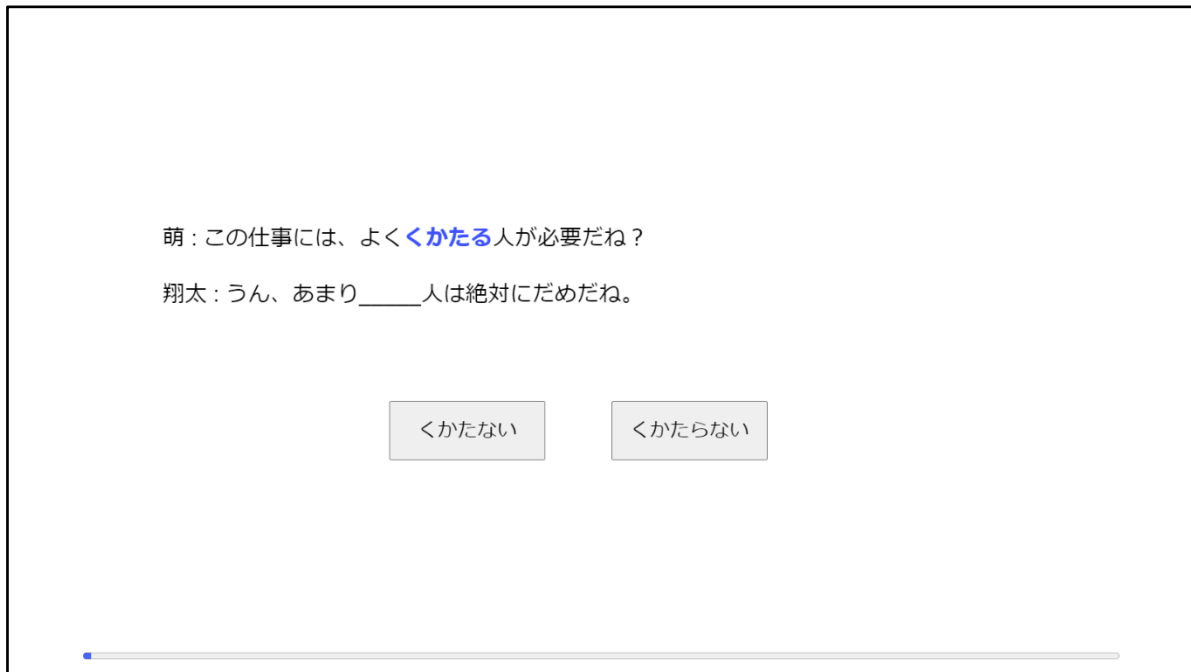


Fig.21 : exemple d'un item tel que présenté dans l'expérience baseline-flexions verbales.

Au démarrage de l'expérience, pour chaque participant, il était déterminé aléatoirement s'il verrait les formes n°1 ou les formes n°2 des pseudo-verbos. Puis, l'ordre d'apparition des pseudo-verbos était également déterminé de manière aléatoire. Lorsque vient le tour de chaque pseudo-verbe d'apparaître, un des soixante dialogues correspondant à la forme dans laquelle il doit être présenté est sélectionné de manière aléatoire pour qu'il y soit intégré et ainsi former un item complet (c.-à-d. un dialogue + un verbe). Les dialogues et les pseudo-verbos n'apparaissent qu'une fois pour chaque participant. Cette expérience a également été programmée sur Construct 2, un logiciel de création de jeux (multi-joueurs) pour navigateur web en 2D (Scirra, 2020 ; release 279).

7.1.3 Participants

Le recrutement des participants s'est déroulé via une plateforme de crowdsourcing : Crowdworks (www.crowdworks.jp). Nous avons ainsi collecté les données de 66 participants. Tout comme dans les expériences précédentes, nos participants sont restés anonymes mais

restent identifiables à l'aide de leur pseudonyme sur le site de Crowdworks. Les mêmes informations étaient demandées à la fin de l'expérience, à savoir : l'âge ; le sexe ; la profession ; le lieu de résidence actuel ; la région ou le pays d'origine ; les précédentes régions ou pays de résidence ; l'âge auquel le japonais est devenu une langue employée au quotidien ; les problèmes de vue ou de lecture ; le niveau de maîtrise des langues étrangères ; et le niveau d'éducation. L'âge moyen des participants est de 37.59 ans (écart type = 11.22). Quant au sexe : 31 étaient des femmes ; et 35 étaient des hommes.

7.2 Résultats

Une fois encore, les temps de réponse particulièrement longs ou courts de certains participants sont probablement dû à un manque d'attention ou d'implication dans la tâche. Nous avons donc commencé notre analyse par le repérage et l'exclusion des données de ces participants visiblement non collaboratifs. Pour cela, nous avons choisi le même critère que précédemment : la transformation logarithmique du temps de réponse médian des participants. Au-delà de la cohérence avec nos choix précédents, ce critère est d'autant plus pertinent pour cette expérience qu'il suit une distribution très proche de la distribution normale. Nous avons donc exclu de nos analyses tous les participants dont le log du temps de réponse médian était éloigné de plus de deux écarts types par rapport à la moyenne du log du temps de réponse médian de l'ensemble des participants. Cela a entraîné l'exclusion de quatre participants dont les temps de réponse étaient trop faibles, mais aucun participant n'affichait de temps de réponses trop longs. Les résultats présentés ci-après correspondent donc aux données des 62 participants restants. Parmi ceux-ci, 30 ont vu les formes n°1 des pseudo-verbos et 32 ont vu leurs formes n°2.

Le résultat principal de cette étude réside dans l'influence de la voyelle de la terminaison dans la proportion de formes *yodan* créée dans l'expérience. Cependant, un second aspect auquel nous nous sommes intéressés est l'influence potentielle de deux facteurs inclus dans l'expérience principalement pour diversifier la tâche et la rendre moins répétitives pour les participants : la forme fléchie demandée (accomplie vs négation) et la longueur du verbe dans sa forme inaccomplie (quatre mores vs trois mores). La figure 22 ci-dessous présente l'influence des différentes combinaisons de ces deux facteurs sur les proportions d'utilisation du pattern *yodan*, en fonction de la voyelle de la terminaison du verbe. On peut constater que les verbes

en *eru* sont en général moins souvent fléchis selon le pattern *yodan* que tous les autres verbes. On remarque également que les verbes en *iru* ne se différencient que très peu des verbes en *aru*, *uru* et *oru*.

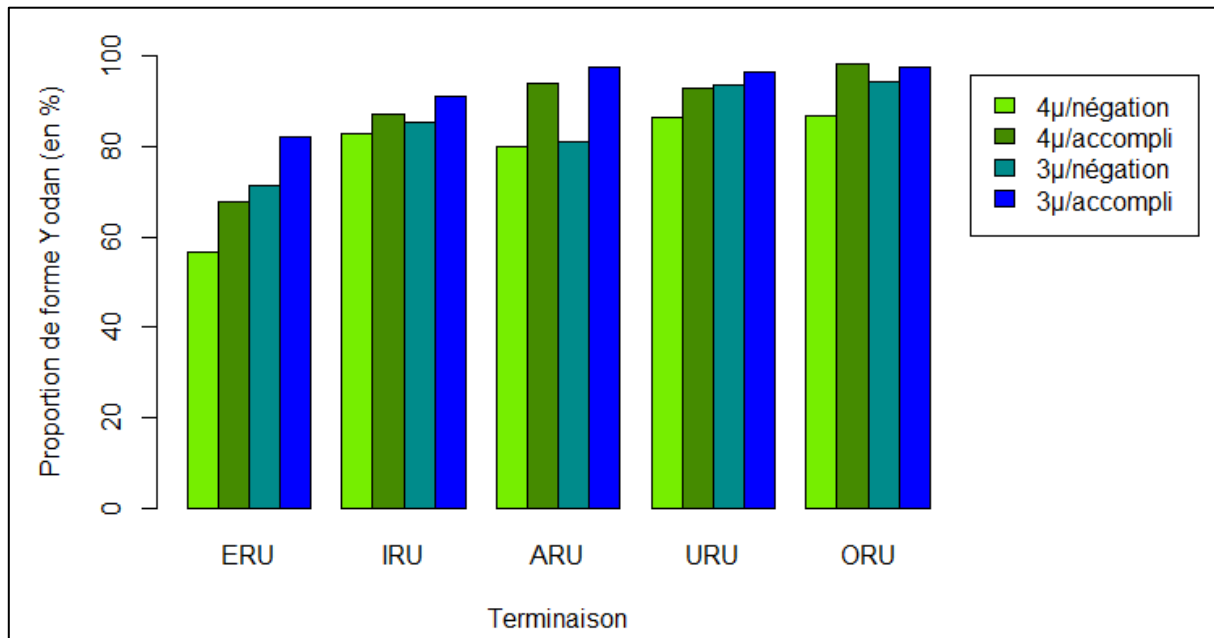


Fig.22 : influence de la forme demandée et de la longueur des verbes sur la proportion de formes *yodan* produites.

Un modèle linéaire généralisé a été appliqué à l'ensemble des données avec : le type de flexion utilisé (*ichidan* vs *yodan*) comme variable dépendante binaire ; et la voyelle de la terminaison, la forme fléchie demandée (accomplie vs négation) et la longueur du verbe dans sa forme inaccomplie (quatre mores vs trois mores) en facteur fixe. Ce modèle étant assez complexe au niveau de ses prédicteurs, il ne nous a pas été possible d'ajouter les participants et les items en tant que facteurs aléatoires comme précédemment. En effet, malgré nos tentatives, un modèle linéaire mixte généralisé de ce type ne pouvait pas converger avec nos données. Des tests du rapport de vraisemblance entre différents modèles ont été effectués pour évaluer la significativité des effets des différents facteurs. Les comparaisons de modèles montrent des effets significatifs : de la voyelle de la terminaison ($\chi^2=-527.07$; $p<0.001$) ; de la longueur des verbes ($\chi^2=-73.281$; $p<0.001$) ; de la forme fléchie demandée ($\chi^2=-114.94$; $p<0.001$) ; de l'interaction entre la voyelle de la terminaison et la forme fléchie demandée ($\chi^2=-155.01$; $p<0.001$), en cela que demander la flexion de négation entraîne une diminution plus importante de la proportion de formes *yodan* produites pour certaines terminaisons, notamment *eru*, *aru*,

et *oru* ; et de l'interaction entre la voyelle de la terminaison et la longueur du verbe ($\chi^2=-9.5364$; $p<0.05$), en cela qu'une longueur de quatre mores entraîne une diminution plus importante de la proportion de formes *yodan* produites pour certaines terminaisons, notamment *eru*. En revanche, il n'y a pas d'effet significatif de l'interaction entre la longueur du verbe et la forme fléchie demandée ($\chi^2=-1.0887$; $p=0.2968$).

7.3 Discussion

Cette nouvelle expérience nous renseigne sur la productivité des patterns *ichidan* et *yodan* dans la flexion de nouveaux pseudo-verbos. En cela, ses résultats viennent confirmer certaines observations des précédentes études sur le sujet, mais ils nous éclairent également sur la productivité de ces deux patterns lorsqu'aucune information encourageant l'emploi de l'un ou l'autre des patterns n'est présentée aux participants.

Nous avons vu que les nouveaux verbes qui entrent dans la langue japonaise étaient tous du groupe *yodan*, ce qui laisse penser que seul ce pattern serait productif dans la création de nouvelles formes. Pourtant, les études précédentes sur le sujet montraient déjà que dans un cadre expérimental, non seulement les locuteurs japonais n'utilisaient pas ce pattern de manière systématique, mais qu'ils utilisaient également le pattern *ichidan*, bien que moins fréquemment. Les résultats de notre expérience viennent confirmer cette tendance, en cela que le pattern *yodan* est largement majoritaire dans les réponses des participants. Ainsi, on peut constater que lorsque l'ambiguïté subsiste concernant le pattern à employer, le pattern *yodan* sera largement privilégié.

Cependant, dans cette situation, le comportement des participants diffère de ce qui a pu être observé précédemment en plusieurs points. Le plus important de ces points est que la voyelle de la terminaison joue un rôle significatif dans la manière dont le pseudo-verbe sera traité. Dans les réponses des participants, les verbes en *eru* suivent plus souvent le pattern *ichidan* que les autres verbes. On remarque en particulier que la proportion de formes *yodan* pour les verbes en *eru* de quatre mores, lorsque la forme négative est demandée, n'est que légèrement supérieure à 50%. Dans la mesure où le pattern *ichidan* montre une certaine productivité, on peut s'attendre au vu des fréquences de types qu'il soit davantage productif pour les verbes en *eru*. Pour rappel, 95% des verbes en *eru* existants sont des *ichidan*. Cela a donc pu jouer sur la propension des participants à appliquer ce pattern au pseudo-verbos en *eru*.

Quoi qu'il en soit, il serait difficile de décrire ce comportement comme du probability-matching, puisque les résultats de l'expérience sont loin des 95% de formes *ichidan* observées dans le lexique.

Par ailleurs, si les participants font appel à leurs connaissances des probabilités dans les formes existantes pour déterminer le pattern à employer, il semble incohérent que les pseudo-verbos en *iru* ne montrent pas davantage de formes *ichidan* que les pseudo-verbos en *aru*, *oru* ou *uru*. En effet, dans les formes existantes, les verbos en *iru* sont presque aussi souvent des *ichidan* que des *yodan*, tandis que les verbos en *aru*, *oru* et *uru* sont tous des verbos *yodan*. Il serait donc attendu que la proportion de réponses *yodan* pour les pseudo-verbos en *iru* se situe entre les pseudo-verbos en *aru*, *oru* et *uru* d'une part, et les pseudo-verbos en *eru* d'autre part. Rappelons que la sélection des items pour cette expérience a été faite en partie en se basant sur l'hypothèse que les pseudo-verbos en *eru* et *iru* seraient traités différemment de ceux en *aru*, *oru* et *uru*, puisqu'elle contient autant de pseudo-verbos de la première catégorie que de la seconde. Cependant, les résultats ne vont pas dans ce sens et invitent à considérer que les pseudo-verbos en *eru* subissent un traitement particulier par rapport à l'ensemble de tous les autres types de pseudo-verbos.

Une autre observation inattendue dans nos résultats réside dans l'effet de la forme fléchie demandée et de la longueur des pseudo-verbos sur leur traitement. Ces facteurs ont initialement été intégrés pour diversifier l'expérience afin non seulement d'éviter d'avoir trop d'items similaires, mais également de la rendre plus intéressante pour les participants et maximiser leur implication dans la tâche. Pourtant, ils montrent des effets significatifs sur les résultats bien que ces effets soient difficiles à motiver théoriquement. Concernant la longueur des verbos, il faudrait pouvoir vérifier si, dans le lexique, les verbos *yodan* ont plus souvent une longueur de trois mores que les verbos *ichidan*, mais à défaut de données objectives à notre disposition sur ce point précis nous n'avons pas de raison de croire que ce soit le cas. Quant à la forme fléchie demandée, il serait particulièrement inattendu que les verbos *ichidan* apparaissent plus souvent à la négation qu'à l'accompli. D'autant plus que si l'on réfléchit en termes de fréquences de types et non de tokens, cela signifierait que les locuteurs connaissent seulement la forme négative et non la forme accomplie de certains verbos *ichidan*, et vice-versa pour les verbos *yodan*. En l'état, il nous paraît donc assez invraisemblable que ces effets trouvent une explication dans les probabilités du lexique. Par ailleurs, les précédentes études sur le sujet ne mentionnaient aucun de ces effets dans leurs résultats. Une explication relative

au protocole expérimental utilisé est donc à privilégier, mais celle-ci ne nous apparaît pas clairement pour autant. Le processus de flexion verbale étant assez régulier, les locuteurs créant majoritairement des formes *yodan*, la tâche de création de nouvelles formes pourrait paraître étrangement trop facile et redondante pour les participants. La volonté d'éviter le sentiment de simplement donner la même réponse systématiquement pourrait les encourager à chercher des différences minimales dans les items présentés pouvant justifier un changement dans leur comportement. Nous disposons malheureusement de trop peu d'informations pour arriver à une conclusion définitive sur l'influence de ces deux facteurs. Bien que nous laissons ces questions en suspens pour le moment, les potentiels effets du protocole employé, ou plus directement de la longueur des verbes et des flexions demandées sur l'emploi des différents patterns sont autant de pistes à explorer dans de futurs travaux.

Cette première expérience sur les flexions verbales nous renseigne sur la manière dont les locuteurs japonais créent de nouveaux verbes. En cela, elle apporte de nouveaux éléments de réponse à nos questions de recherche relatives à la construction des représentations qui guident la productivité des patterns morphologiques :

- 1) Les locuteurs reproduisent-ils simplement les probabilités de l'input dans leur output (c.-à-d. ont-ils un comportement de type « probability-matching ») ?
- 2) Sinon, quels autres facteurs entrent en jeu ?

En réponse à la question 1), les résultats de cette expérience soulignent une fois de plus l'importance des probabilités de l'input sur celles de l'output tout en fournissant de nouveaux arguments en défaveur de l'idée d'un comportement de probability-matching. La productivité largement supérieure du pattern *yodan* par rapport au pattern *ichidan* est à mettre en lien avec le fait que le pattern *yodan* soit le seul observé parmi les nouveaux verbes qui sont effectivement entrés dans la langue japonaise. Mais sa productivité imparfaite n'est pas entièrement cohérente avec cette observation. Au contraire, le fait que le pattern *ichidan* montre une certaine productivité, notamment avec les verbes en *eru* peut être lié au nombre important de formes existantes suivant ce dernier. Finalement, le comportement des participants semble influencé à la fois par les probabilités dans les nouveaux verbes japonais, ainsi que celles de l'ensemble des formes existantes, mais sans que les résultats ne soient l'exact reflet des probabilités à l'un ou l'autre de ces niveaux.

Concernant la question 2), cette expérience ne met pas vraiment en lumière de nouveaux facteurs entrant en jeu dans la productivité des patterns morphologiques. En revanche, un objectif secondaire de cette première expérience était de servir de baseline afin de créer des listes équilibrées d'items pour appliquer le paradigme ESP au cas des flexions verbales. Dans la section suivante, nous pourrons donc, grâce aux résultats de cette expérience, tester le rôle de la convergence dans un processus morphologique pour lequel les locuteurs sont exposés à un input très fourni.

Chapitre 8 – Expérience ESP - Flexions verbales : convergence morphophonologique dans un processus morphologique non récent

L'un des principaux objectifs de l'étude des flexions verbales dans le cadre de nos travaux est de pouvoir vérifier si la convergence morphologique peut se produire malgré le fait que les locuteurs soient exposés à un input très important. Au sujet de la convergence morphologique, l'expérience ESP sur les CAL suggère qu'elle se produit même dans des situations où les locuteurs adoptent des comportements proches du probability-matching dans la création de nouvelles formes. Toutefois, la formation des CAL reste un processus morphologique récent, montrant beaucoup de variabilité à différents niveaux et proposant un input relativement restreint aux locuteurs. Il est donc possible que cela encourage les locuteurs à être davantage à l'écoute des nouveaux inputs qu'ils reçoivent de leur partenaire d'interaction.

Après avoir observé la manière dont les locuteurs japonais emploient les différents patterns dans la flexion de pseudo-verbos en dehors de toute interaction dans l'expérience baseline, nous pouvons appliquer le paradigme ESP à l'étude de ce processus morphologique. Nous pourrions alors comparer les résultats de cette expérience avec celle portant sur les CAL afin d'évaluer si l'importance du nombre de formes existantes disponibles pour les locuteurs a une influence sur leur propension à converger vers un partenaire d'interaction.

8.1 Méthode

8.1.1 Matériel

Comme dans les autres expériences présentées précédemment, chaque participant voyait 120 items ici définis, comme dans l'expérience baseline, comme un pseudo-verbe inséré dans un court dialogue écrit. Cependant, ils étaient cette fois répartis dans trois sous-listes correspondant aux trois phases de jeu du paradigme ESP, soit 40 items par phase. Les pseudo-verbos et les dialogues utilisés dans l'expérience ESP sont les mêmes que ceux de l'expérience

baseline (annexe 9 et 10 respectivement). Pour rappel, cette liste de 120 pseudo-verbos contient :

- 15 pseudo-verbos en *iru* d'une longueur de **trois mores**.
- 15 pseudo-verbos en *iru* d'une longueur de **quatre mores**.
- 15 pseudo-verbos en *eru* d'une longueur de **trois mores**.
- 15 pseudo-verbos en *eru* d'une longueur de **quatre mores**.

- 10 pseudo-verbos en *aru* d'une longueur de **trois mores**.
- 10 pseudo-verbos en *aru* d'une longueur de **quatre mores**.
- 10 pseudo-verbos en *uru* d'une longueur de **trois mores**.
- 10 pseudo-verbos en *uru* d'une longueur de **quatre mores**.
- 10 pseudo-verbos en *oru* d'une longueur de **trois mores**.
- 10 pseudo-verbos en *oru* d'une longueur de **quatre mores**.

Les données de l'expérience baseline ont été utilisées afin de créer des sous-listes équilibrées pour chaque phase de jeu en termes de nombre de formes *yodan* produites. Afin de constituer ces listes, pour chacun des types de pseudo-verbos ci-dessus, les trois premiers pseudo-verbos ayant obtenu les proportions les plus élevées de formes *yodan* dans l'expérience baseline étaient répartis aléatoirement dans les sous-listes (un pseudo-verbe par sous-liste). Puis le même processus était répété sur les trois pseudo-verbos suivants ayant les proportions les plus élevées de formes *yodan* dans l'expérience baseline. Et ainsi de suite jusqu'à épuisement de tous les pseudo-verbos de chaque type.

Notons que pour les types ne contenant que dix pseudo-verbos, un d'entre eux (celui avec la plus faible proportion de formes *yodan* dans l'expérience baseline) se retrouvait isolé après trois répétitions du processus. Puisque six pseudo-verbos sont concernés, ils ont été répartis comme suit dans les différentes listes : le pseudo-verbe en *uru* de quatre mores et le pseudo-verbe en *aru* de trois mores sont dans la liste 1 ; le pseudo-verbe en *aru* de quatre mores et le pseudo-verbe en *oru* de trois mores sont dans la liste 2 ; le pseudo-verbe en *oru* de quatre mores et le pseudo-verbe en *uru* de trois mores sont dans la liste 3. La répartition entre les pseudo-verbos dont la forme n°1 serait l'accompli et la forme n°2 serait la négation, et vice-versa, a été gardé à l'identique depuis l'expérience baseline.

8.1.2 Procédure

Cette expérience adapte le paradigme ESP, tel que décrit en section 5.1, à l'étude des flexions verbales en japonais. Comme précédemment, l'expérience se déroule sous la forme d'un jeu en ligne en trois phases. Dans la première phase, les participants devaient choisir la flexion de pseudo-verbos dans une tâche similaire à l'expérience baseline. Pour chaque item, l'écran affichait un court dialogue entre deux amis dans lequel le pseudo-verbe apparaissait à la forme inaccomplie dans la phrase du premier personnage, et la réponse du second personnage présentait un trou devant être rempli par ce même pseudo-verbe à la forme accomplie ou négative. Deux boutons de réponses correspondant à l'application du pattern *ichidan* et *yodan* sur le pseudo-verbe, écrit en *hiragana*, étaient également visibles. Dans toutes les phases de jeu et pour chaque item, la répartition du pattern *ichidan* et *yodan* entre les deux boutons de réponses était déterminée aléatoirement. De plus, afin d'identifier les participants non coopératifs, la position du bouton sélectionné et le temps de réponse pour chaque item étaient enregistrés. Une barre de progression était présente en bas de l'écran. Enfin, contrairement à l'expérience baseline, un numéro de joueur correspondant à une suite aléatoire de chiffres et de lettres était affiché en bas à gauche de l'écran. Ce numéro de joueur a été ajouté pour rendre l'expérience multijoueur plus crédible, notamment lors de la phase 2. La figure 23 ci-dessous présente un item lors de la première phase de jeu. Pour cette phase, les instructions invitaient les joueurs à choisir la forme qui leur semblait le plus naturelle pour compléter la phrase à trou, sans se soucier du sens des verbes (voir annexe 4 pour les instructions complètes). Pour chaque participant, l'ordre dans lequel apparaissaient les sous-listes, l'ordre des items dans chaque phase, et le numéro de la forme fléchée demandée étaient déterminés de manière aléatoire. Lorsque venait le tour de chaque pseudo-verbe d'apparaître, il était intégré dans un dialogue pour former un item. Un des soixante dialogues dont la phrase à trou exigeait un verbe à la forme correspondante était sélectionné de manière aléatoire, chaque dialogue et chaque pseudo-verbe n'apparaissant qu'une fois pour chaque participant.

En dehors de ces adaptations au cas des flexions verbales, l'expérience se déroulait ensuite de la même manière que décrit en section 5.1.2 pour le cas des CAL. Le comportement du partenaire artificiel qui intervient dans la phase 2 a également été adapté au processus de flexion verbale. Cela est décrit en détails dans la section suivante (8.1.3). Cette expérience a

également été programmée sur Construct 2, un logiciel de création de jeux (multi-joueurs) pour navigateur web en 2D (Scirra, 2020 ; release 279).

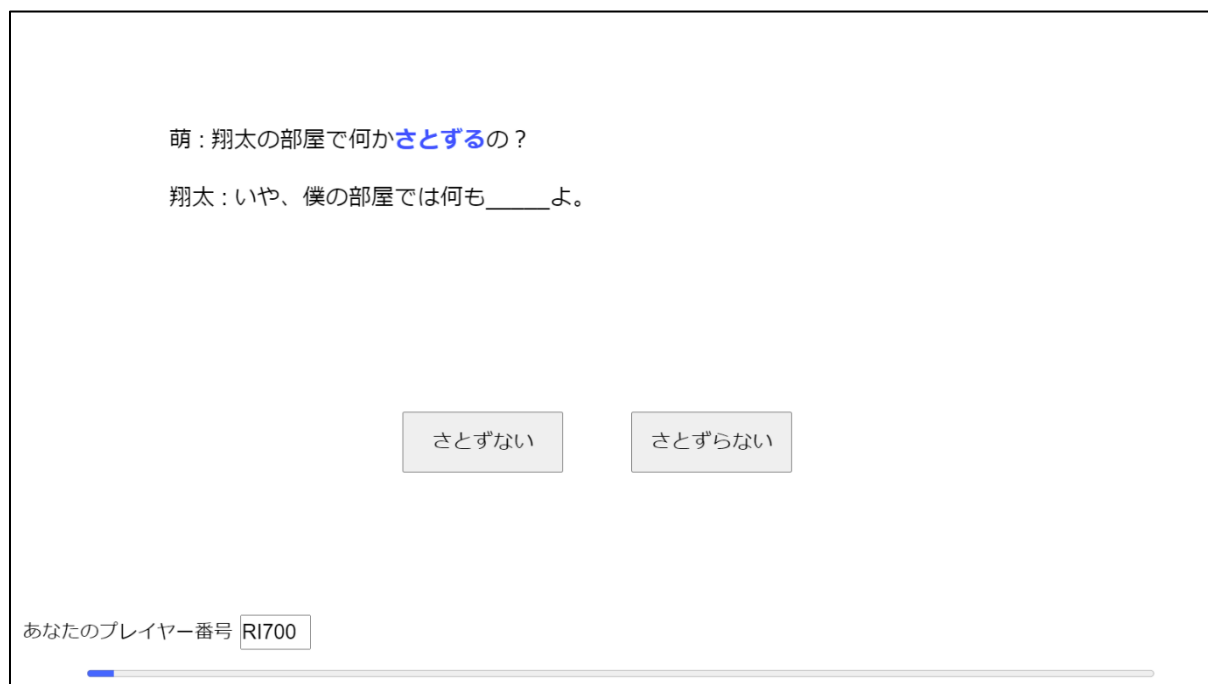


Fig.23 : exemple d'un item tel que présenté dans la phase 1 de l'expérience ESP-flexions verbales.

8.1.3 Partenaire artificiel

Le partenaire artificiel de cette nouvelle expérience a été programmé d'une manière similaire à celui de l'expérience ESP portant sur les CAL. En se basant sur les réponses du participant en phase 1, il pouvait donc adopter trois comportements différents, correspondant à nos trois conditions expérimentales. Bien que l'on puisse débattre du statut régulier du pattern *yodan* pour les flexions verbales japonaises, nous garderons les appellations « sur-régularisateur » et « sous-régularisateur » par souci de cohérence avec l'expérience ESP-CAL. Le pattern *yodan* peut être considéré comme régulier en cela qu'il est le seul pattern productif dans les verbes ayant intégré la langue récemment, et qu'il montre une productivité nettement supérieure au pattern *ichidan* dans les tâches de création de nouvelles formes :

- La condition « sans changement » : le robot enregistre le nombre d'utilisations des patterns *ichidan* et *yodan* par le participant lors de la première phase et reproduit

exactement le même nombre d'utilisations dans ses propres réponses dans la deuxième phase.

- La condition « sur-régularisateur » : le robot produira davantage de formes *yodan* dans la deuxième phase que le nombre produit par le participant lors de la première phase.
- La condition « sous-régularisateur » : le robot produira moins de formes *yodan* dans la deuxième phase que le nombre produit par le participant lors de la première phase.

La stratégie retenue concernant la manière dont le robot doit augmenter ou réduire le nombre de formes *yodan* et la même que celle employée pour les formes régulières dans l'expérience ESP-CAL, pour les mêmes raisons qu'exposées en section 5.1.3. Les robots « sur-régularisateur » et « sous-régularisateur » appliquent donc des formules légèrement différentes. Le robot « sur-régularisateur » applique la formule suivante :

$$\text{round}((n + (0.3 * (1 - n)) * 40)$$

Dans cette formule, n représente la proportion de formes *yodan* produites par le participant en phase 1. En toute logique, $(1 - n)$ correspond donc au nombre de formes *ichidan* produites. Le robot « sur-régularisateur » prélève donc 30% des formes *ichidan* produites et les ajoute aux formes *yodan* à produire. En multipliant cette proportion par 40, puis en arrondissant, le robot obtient un nombre rond de formes *yodan* à produire sur les 40 items de la phase 2. Au maximum, le robot « sur-régularisateur » créera 12 formes *yodan* de plus que le participant en phase 1 si ce dernier n'a produit aucune forme *yodan*. L'importance de l'augmentation du nombre de formes *yodan* produites par le robot diminue à mesure que le nombre de formes *yodan* produites par le participant en phase 1 augmente.

Concernant le robot « sous-régularisateur », il appliquera la forme suivante :

$$\text{round}((n - (0.3 * n)) * 40)$$

Ce robot soustrait donc 30% des formes *yodan* produites par le participant en phase 1 et produira autant de formes *ichidan* à la place. Au maximum, il diminuera de 12 le nombre de formes *yodan* produites par le participant en phase 1, et ce dans le cas où ce dernier n'a produit que des formes *yodan*. À l'inverse du robot « sur-régularisateur », l'importance de la diminution du nombre de formes *yodan* produites par le robot augmente à mesure que le nombre de formes *yodan* produites par le participant en phase 1 augmente.

Contrairement au cas des CAL qui présentait plusieurs patterns irréguliers, le cas des flexions verbales ne présente que deux patterns différents. Ainsi il n'était pas nécessaire de programmer de stratégie particulière concernant l'emploi de patterns considérés comme « irréguliers ». Après avoir calculé le nombre de formes *yodan* qu'il devait produire, le robot produisait autant de formes *ichidan* que nécessaire pour arriver à un total de 40.

8.1.4 Participants

Le recrutement des participants, au nombre de 296, s'est déroulé sur la plateforme de crowdsourcing japonaise Crowdworks (www.crowdworks.jp). Comme pour les expériences précédentes, les participants sont restés anonymes, mais sont identifiables via leur pseudonyme sur le site de Crowdworks. Par ailleurs, les questions d'ordre démographique ont permis de collecter des informations sur : l'âge ; le sexe ; la profession ; le lieu de résidence actuel ; la région ou le pays d'origine ; les précédentes régions ou pays de résidence ; l'âge auquel le japonais est devenu une langue employée au quotidien ; les problèmes de vue ou de lecture ; le niveau de maîtrise des langues étrangères ; et le niveau d'éducation des participants. L'âge moyen des participants est de 35.99 ans (écart type = 10.86). Quant au sexe : 6 participants n'ont pas souhaité l'indiquer ; 171 étaient des femmes ; et 119 étaient des hommes.

8.2 Résultats

Afin d'exclure les participants visiblement non coopératifs, le même critère que celui utilisé pour l'expérience ESP sur les CAL a été utilisé : le log du temps de réponse médian dans les phases 1 et 3 uniquement. Cela a entraîné l'exclusion de quinze participants : quatorze ayant des temps de réponse trop courts, et un ayant des temps de réponse trop longs. Les résultats présentés dans cette section sont donc ceux des 281 participants restants. Parmi ceux-ci, 93 ont été exposés au robot « sans-changement » ; 98 au robot « sur-régularisateur » ; et 90 au robot « sous-régularisateur ».

Les principaux résultats de cette expérience portent sur l'effet du comportement du robot sur l'emploi des différents patterns de flexions verbales par les participants dans la création de nouvelles formes. La figure 24 ci-dessous présente l'évolution de la proportion de

formes *yodan* produites à mesure que le jeu progresse dans les différentes conditions expérimentales.

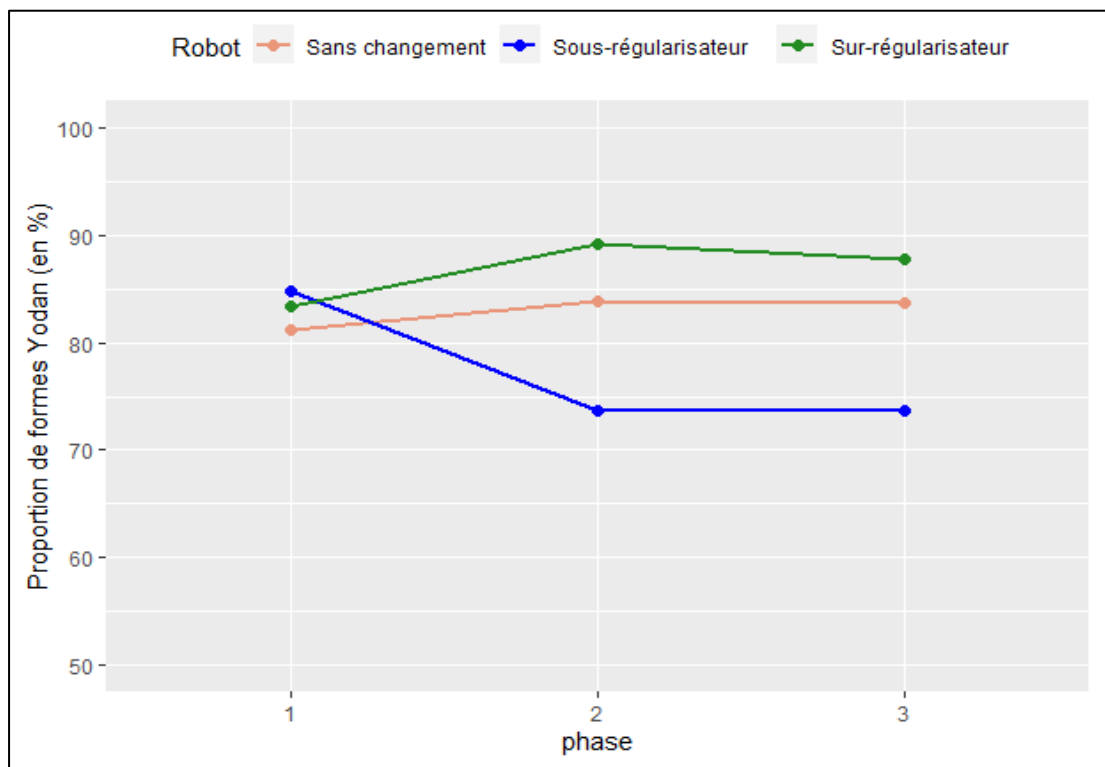


Fig.24 : proportions (en %) de formes *yodan* produites en fonction de la phase et du comportement du robot.

Comme précédemment, nous avons exclu la phase 2 de l'analyse statistique puisque seules les phases 1 et 3 (assimilables à une phase pré-test et post-test, respectivement) présentent des tâches expérimentales similaires dont les résultats sont directement comparables. Un modèle linéaire mixte généralisé a été appliqué à l'ensemble des données avec : le type de pattern utilisé (*yodan* vs *ichidan*) comme variable dépendante binaire ; le comportement du robot (« sans changement » ; « sur-régularisateur » ; « sous-régularisateur ») et la phase (1 vs 3) en facteurs fixes ; et le participant et le pseudo-verbe en effets aléatoires. Des tests du rapport de vraisemblance entre différents modèles ont été effectués pour évaluer la significativité des effets des différents facteurs. Les comparaisons de modèles ont révélé un effet significatif du comportement du robot dans le sens attendu, c-à-d que par rapport au robot « sans-changement », le robot « sur-régularisateur » entraîne une augmentation du nombre de formes *yodan* produites, tandis que le robot « sous-régularisateur » entraîne une diminution ($\chi^2=9.597$; $p<0.01$) ; de la phase ($\chi^2=4.5848$; $p<0.05$), en cela que les participants créent plus de formes

yodan à mesure que le jeu progresse ; et de l'interaction entre ces deux facteurs ($\chi^2=185.73$; $p<0.001$). Afin d'observer les différences de groupes d'une phase à l'autre ou au sein d'une même phase, toutes les comparaisons post hoc par paires ont été réalisées avec une correction de Tukey. Le tableau 17 ci-dessous présente les ratios z et les valeurs p de ces comparaisons. Les trois premières lignes, en vert, nous indiquent qu'aucune comparaison entre les différents comportements du robot en phase 1 ne montre d'effet significatif, ce qui signifie que les groupes de participants exposés aux différentes conditions sont homogènes et que les différences observées en phase 3 sont effectivement liées à des facteurs propres à l'expérience. Les trois lignes suivantes, en rouge, nous indiquent que la phase a un effet significatif pour toutes les conditions. Enfin, les trois lignes suivantes, en bleu, nous indiquent que l'effet du robot en phase 3 est principalement causé par la différence entre le robot « sous-régularisateur », qui entraîne une diminution de formes *yodan* créées par les participants, et les deux autres robots, qui entraînent une augmentation.

Tableau 17 : ratios z et valeurs p des comparaisons post hoc par paires avec correction de Tukey pour les données de l'expérience ESP – flexions verbales.

Comparaison					
Phase	Robot	Phase	Robot	Ratio z	Valeur p
1	1	1	3	-1.549	0.6323
1	1	1	2	-1.194	0.84
1	2	1	3	-0.385	0.9989
1	1	3	1	-3.461	<0.01
1	2	3	2	-5.882	<0.001
1	3	3	3	12.026	<0.001
3	1	3	2	-2.166	0.2536
3	1	3	3	3.933	<0.01
3	2	3	3	6.112	<0.001
3	1	1	3	-0.316	0.9996
3	1	1	2	0.066	1
1	2	3	3	3.916	<0.01
3	2	1	3	1.829	0.4471
1	1	3	3	2.703	0.0746
1	1	3	2	-3.422	<0.01

8.3 Discussion

Cette nouvelle expérience ESP se focalise sur l'effet de la convergence dans un processus morphologique auquel les locuteurs ont eu une forte exposition. En principe, si les locuteurs adoptent un comportement de probability-matching, les effets de la convergence devraient être atténués dans cette situation. En effet, si l'input auquel les locuteurs ont déjà été exposés est plus important, alors, les nouveaux inputs qu'ils reçoivent devraient être davantage noyés dans la masse et donc être proportionnellement moins important.

Pourtant, les résultats de l'expérience ne soutiennent pas cette hypothèse, en cela qu'ils sont assez similaires à ce que l'on observait dans l'expérience ESP portant sur les CAL, un processus morphologique pour lequel les locuteurs ont eu une plus faible exposition. On peut constater sur la figure 24 que le robot a l'effet attendu sur le comportement des participants, et ce même si le taux de formes *yodan* qu'ils produisent en phase 1 est supérieur à 80%, ce qui témoigne d'une variabilité assez faible dans ce processus morphologique. En comparant ces résultats à ceux des items /R/ de l'expérience ESP sur les CAL, pour lesquels le comportement du robot a un effet significatif (figure 16, page 97), on constate de fortes similitudes dans le comportement des participants.

Dans les deux situations, lorsque les participants sont exposés au robot « sans-changement », on observe une légère augmentation du nombre de formes régulières et *yodan* (pour les CAL et les flexions verbales respectivement). Cela semble dû au même effet d'habituation mentionné précédemment.

Lorsqu'ils sont exposés au robot « sur-régularisateur », on observe une augmentation plus importante du nombre de formes régulières et *yodan* que pour le robot « sans-changement ». Notons que dans le cas de l'expérience sur les verbes, les participants sont déjà proches de produire le nombre maximal de formes *yodan* en phase 1, et le robot « sur-régularisateur » ne se différencie que peu du « sans changement » dans cette situation. Cela peut expliquer que bien que l'augmentation soit plus importante que celle observée pour le robot « sans changement » de l'expérience sur les verbes, elle est toutefois moins importante que celle observée pour le robot « sur-régularisateur » de l'expérience sur les CAL.

Enfin, lorsque les participants sont confrontés au robot « sous-régularisateur », on observe une diminution du nombre de formes régulières et *yodan* produites. Il est

particulièrement intéressant de noter que cette diminution est plus importante pour les flexions verbales que pour les CAL, montrant que la convergence peut être aussi effective pour les deux processus morphologiques. Cette tendance a pu être encouragée par le fait que les participants soient proches du nombre maximal de formes *yodan* produite en phase 1, une situation dans laquelle le robot « sous-régularisateur » crée beaucoup moins de formes *yodan* en retour.

Cette expérience ESP sur les flexions verbales nous en apprend davantage sur le phénomène de convergence morphologique et nous apporte ainsi de nouveaux éléments de réponse à nos questions de recherche relatives à la construction des représentations qui guident la productivité des patterns morphologiques :

- 1) Les locuteurs reproduisent-ils simplement les probabilités de l'input dans leur output (c.-à-d. ont-ils un comportement de type « probability-matching ») ?
- 2) Sinon, quels autres facteurs entrent en jeu ?

Les situations dans lesquelles les locuteurs sont amenés à converger remettent nécessairement la notion de probability-matching en question, puisqu'un nouvel input rencontré le temps d'une conversation ou d'une tâche expérimentale ne devrait pas pouvoir modifier significativement les probabilités de l'input global auquel un locuteur a été exposé. Ainsi, les résultats de l'expérience ESP sur les flexions verbales apportent de nouveaux arguments en faveur d'une réponse négative à la question 1).

En réponse à la question 2), ces résultats montrent, au contraire, que les locuteurs peuvent considérablement s'éloigner des probabilités de l'input dans la création de nouvelles formes, et ce sous l'effet de la convergence morphologique. Cet effet, déjà observé pour le processus de formation des CAL largement soumis à la variation, se manifeste dans des proportions similaires dans le cas des flexions verbales, un processus morphologique beaucoup plus régulier. Cela indique que l'importance de l'exposition que les locuteurs ont à un certain processus morphologique n'a pas ou peu d'effet sur leur propension à converger vers leur partenaire d'interaction.

Chapitre 9 – Discussion générale

9.1 Réponses aux questions de recherche

À travers les différentes expériences présentées dans cette thèse, nous avons exploré le processus d'abstraction et l'évolution des représentations qui guident la productivité morphologique. La littérature sur le sujet regorge d'arguments soulignant l'importance de l'influence des probabilités de l'input auquel sont exposés les locuteurs sur celles de l'output qu'ils produisent. De fait, l'idée d'un comportement de probability-matching est une bonne approximation de ce qui motive les décisions des locuteurs, au moins dans des situations où ces derniers disposent d'un système stable et mature basé sur un input conséquent. Cependant les limites de cette approximation sont parfois clairement visibles. D'une part, on trouve des situations linguistiques qui ne reflètent pas ce qui serait attendu si les locuteurs suivaient un comportement de probability-matching (Becker et al., 2011 ; Dawdy-Hesterberg, 2014 ; Kapatsinski, 2010). D'autre part, sur le plan théorique, il est difficile de concilier l'idée du probability-matching avec celles de l'apparition et de l'évolution de patterns morphologiques, puisqu'une éternelle reproduction des probabilités d'un input donné ne laisse aucune place pour la variation nécessaire à l'émergence de ces phénomènes. Bien que l'on retrouve ces réflexions sur la nature approximative des explications en termes de probability-matching dès leurs origines, on trouve en revanche peu d'informations sur les facteurs qui mènent les locuteurs à dévier du probability-matching. Ce constat a motivé nos deux questions de recherche, la première étant focalisée sur les limites de cette approximation, et la seconde sur les facteurs à prendre en compte pour avoir une conception plus juste des représentations qui guident la productivité morphologique.

En réponse à notre première question de recherche, les résultats de toutes nos expériences témoignent d'une manière ou d'une autre d'un comportement de la part des participants ne correspondant pas à du probability-matching, mettant en avant autant de limites de cette approximation. Dans le cas des CAL, l'expérience baseline montre que les participants utilisent le pattern régulier d'une façon proche de ce qui serait attendu d'un comportement de probability-matching : dans la plupart des cas pour les items *sans conflit* ; presque aussi fréquemment que les patterns irréguliers pour les items /R/ ; et quasiment jamais pour les items

/Q/. Mais ces résultats, bien que proche de ce que l'on observe dans le lexique, n'en sont pas l'exact reflet pour autant. En effet, le pattern régulier y est employé globalement moins fréquemment que dans le lexique. Une divergence du probability-matching plus frappante se situe dans l'utilisation des différents patterns irréguliers. Tandis que dans le lexique, le pattern de Suppression est clairement plus fréquent que les deux autres, les participants utilisent indifféremment les trois patterns irréguliers. Dans le cas des flexions verbales, les participants utilisent le pattern *yodan* d'une manière qui n'est ni cohérente avec les probabilités observées dans l'ensemble du lexique, ni avec la productivité observée de ce pattern dans les verbes ayant récemment intégré le lexique japonais. Enfin, les expériences ESP montrent que les participants sont capables de converger avec un partenaire d'interaction et ce quelle que soit l'importance de l'input dont ils disposent. Cela implique que les probabilités dans l'output qu'ils produisent dans de telles situations n'est pas calqué sur celle de l'input total auquel ils ont été exposés, puisque les productions de leur partenaire d'interaction y jouent un rôle proportionnellement plus important.

Au-delà de simplement confirmer certaines limites des explications en termes de probability-matching, les résultats des différentes expériences apportent également des réponses concernant les différents facteurs qu'il faut prendre en compte afin de comprendre de quelle manière certains patterns morphologiques deviennent productifs. Les principales conclusions que l'on peut tirer de ces résultats sont les suivantes : les locuteurs possèdent des a priori sur la productivité de certains patterns ; ces a priori jouent un rôle d'autant plus important que l'input accessible aux locuteurs est réduit ; et il existe des situations linguistiques dans lesquelles les locuteurs peuvent rapidement et considérablement modifier leur comportement pour s'adapter à un nouvel input. Nous avons mentionné plusieurs fois qu'une approche bayésienne permettrait de décrire plus efficacement les mécanismes sous-jacents à l'abstraction des représentations qui guident la productivité morphologique. Cette approche nous semble particulièrement adaptée pour rendre compte de nos résultats pour plusieurs raisons. Tout d'abord, comme nous l'avons vu en section 1.3.2, avec un input important le comportement attendu d'un raisonnement bayésien et celui d'un comportement de probability-matching convergent, ce qui rend l'approche bayésienne a minima aussi efficace à décrire ces situations qui peuvent pourtant être assimilées à du probability-matching. Par ailleurs, l'un des principes d'un modèle d'apprentissage bayésien est que chaque nouvel input est pris en compte pour graduellement modifier l'état du système. C'est une caractéristique qu'il a en commun avec les

approches usage-based, telles que la théorie des exemplaires, qui s'avèrent particulièrement efficaces pour décrire l'évolution de catégories et de représentations linguistiques (Bybee, 2001, 2006 ; Pierrehumbert, 2001a ; Wedel, 2006). Ainsi, comme Shi, Griffiths, Feldman, & Sanborn (2010), nous ne considérons pas les raisonnements bayésiens comme une alternative aux modèles théoriques exemplaristes, mais nous considérons au contraire que les deux approches sont complémentaires. Appliqué à la linguistique, le théorème de Bayes peut s'interpréter comme une expression mathématique du processus par lequel les locuteurs modifient leurs représentations linguistiques par l'intégration de nouveaux exemplaires, et ce en fonction des exemplaires qu'ils ont déjà en mémoire. L'un des intérêts de l'approche bayésienne est qu'elle propose une manière convaincante de rendre compte des biais a priori des locuteurs, potentiellement avant même qu'ils ne soient exposés à un quelconque input. Cela s'avère particulièrement pertinent pour décrire nos résultats relatifs à des situations linguistiques dans lesquelles les locuteurs sont exposés à un input limité. Nous allons donc voir plus en détails comment ce cadre théorique s'applique aux résultats de nos expériences.

9.2 Modèle bayésien de l'apprentissage de la morphologie

Pour rappel (voir section 1.3.2), le théorème de Bayes nous indique que la probabilité a posteriori d'une hypothèse étant donné un certain jeu de données est proportionnelle à la probabilité a priori de cette hypothèse et la vraisemblance d'observer ces données en admettant cette hypothèse. Dans notre cas, nous cherchons à évaluer les probabilités d'hypothèses relatives à la fréquence d'utilisation de patterns linguistiques. Par exemple, dans le cas des CAL, on peut s'interroger sur la probabilité qu'un locuteur estime que le pattern régulier doit être utilisé dans une certaine proportion. Cette proportion peut varier entre 0% et 100% et on cherche à déterminer pour chacune des valeurs possibles ses chances d'être le reflet de l'état réel des préférences du locuteur. En quelques sortes, nous nous intéressons à des probabilités de probabilités. Par exemple, la probabilité que la probabilité d'apparition du pattern régulier soit de 0, ou de 0.1, ou de 1, etc. Cependant, nous devons évaluer l'ensemble des valeurs possibles entre 0 et 1 et nous avons donc un continuum d'hypothèses situé dans cet intervalle. Dans ce cas de figure, le théorème de Bayes prend la forme ci-dessous :

$$f(\theta|x) = \frac{p(x|\theta)f(\theta)}{\int_0^1 p(x|\theta)f(\theta)d\theta}$$

Dans cette formule, θ représente une valeur possible de l'hypothèse (soit une valeur donnée entre 0 et 1) et x représente les données observées. La fonction $f(\theta)$ est une loi de probabilité indiquant, pour chaque valeur possible de θ , sa probabilité a priori. Dans le numérateur, $p(x|\theta)$ est la vraisemblance d'observer les données x , étant donné la valeur de θ . Le dénominateur est égal à $p(x)$, c.-à-d. la probabilité totale, toutes hypothèses confondues, d'observer les données x . Dans le cas d'hypothèses continues, on obtient cette constante en calculant l'intégrale de la vraisemblance de x pour toutes les valeurs de θ . Enfin, $f(\theta/x)$ est une loi de probabilité représentant, pour chaque valeur de θ , sa probabilité a posteriori, c.-à-d. après avoir pris en compte les données x . Puisque le dénominateur reste une constante dans ce cas de figure, nous obtenons toujours des probabilités a posteriori (c.-à-d. $f(\theta/x)$) proportionnelles aux probabilités a priori (c.-à-d. $f(\theta)$) et à la vraisemblance de x pour chaque valeur de θ (c.-à-d. $p(x/\theta)$).

Nous allons voir à présent que cette formule contient tous les éléments nécessaires pour rendre compte des résultats de nos diverses expériences que ce soit dans les cas où ils sont semblables à du probability-matching, mais aussi et surtout, dans les cas où ils s'en éloignent. Notons toutefois qu'il ne s'agira pas ici de modéliser précisément les a priori des locuteurs japonais et leur évolution. En effet, dans bien des cas, nous ne disposons pas de données suffisantes nous permettant de définir avec précision les a priori des locuteurs. Nous tenterons toutefois d'illustrer de quelle manière de simples raisonnements bayésiens peuvent mener aux résultats que nous avons obtenus en partant d'a priori dont nous limiterons la subjectivité au maximum, et en les enrichissant par des données semblables à ce que l'on observe dans le lexique japonais. Par souci de simplification, nous réduirons les différentes fonctions considérées (qu'elles définissent les a priori, les a posteriori, ou la vraisemblance de données), à des lois normales dont la moyenne correspond à la valeur que l'on suppose être la probabilité d'emploi d'un certain pattern morphologique sur un type d'item en particulier, et dont l'écart type correspond à l'incertitude concernant cette valeur. Bien que cela ne reflète pas nécessairement l'entière complexité des processus morphologiques étudiés, où plus de deux patterns peuvent être en concurrence, ou le traitement d'un type d'items peut influencer sur le

traitement d'un autre type d'items, les principes essentiels du raisonnement bayésien sont conservés.

Concentrons-nous tout d'abord sur l'utilisation du pattern régulier dans le traitement des CAL *sans conflit*. Nous savons que dans le lexique de CAL, environ 90% des CAL *sans conflit* suivent le pattern régulier de maintien. Pourtant, dans l'expérience baseline ainsi que dans la première phase de l'expérience ESP, la proportion d'utilisation du pattern de maintien pour les items *sans conflit* est d'environ 70%. Dans cette situation, les participants ont un comportement proche mais quelque peu divergent du probability-matching. Dans un raisonnement bayésien, une manière d'obtenir un comportement très proche du probability-matching serait de supposer que les locuteurs ont un a priori uniforme (aussi appelé « a priori non informatif ») pour toutes les valeurs de θ . Autrement dit, un a priori tel que $f(\theta) = 1$. Explorons donc un cas de figure où les locuteurs auraient un a priori uniforme et seraient exposés à un jeu de données qui, en moyenne, présenterait 90% de formes régulières. Pour cela, la vraisemblance des données auxquelles sont exposés les locuteurs sera déterminée par une loi normale de moyenne 0.9 et un écart type de 0.3. Cet écart type correspond à la fiabilité des données observées. Plus il est faible et plus les données sont fiables, par exemple, parce qu'elles correspondent à une grande quantité de mots observés. Puisque les CAL sont issus d'un processus morphologique récent pour lesquels les locuteurs disposent d'un input restreint, nous choisissons un écart type élevé. Concrètement, ce cas de figure correspondrait à une situation dans laquelle le taux de régularisation a priori d'un locuteur pris au hasard dans la population aurait autant de chance d'être n'importe quelle valeur entre 0 et 1. Les locuteurs ne sont pas fréquemment confrontés à des CAL mais ceux qu'ils rencontrent sont à 90% réguliers en moyenne. Cela pourrait être dû au fait, par exemple, que par le fruit du hasard les locuteurs qui créent et utilisent des CAL sont ceux qui possèdent un fort taux de régularisation a priori.

La figure 25 présente les courbes des différentes fonctions relatives à ce cas de figure¹⁷. On peut constater qu'après la prise en compte des données (en bleu), l'a priori plat (en noir) devient un a posteriori (en rouge) qui correspond exactement aux données. Bien entendu, un résultat similaire se produirait avec n'importe quel autre jeu de données. C'est en cela que ce

¹⁷ Toutes les fonctions utilisées dans nos exemples sont divisées par leur intégrale afin de les transformer en loi de probabilité (dont l'intégrale est égale à 1), ce qui indispensable pour obtenir des valeurs proportionnelles entre a priori, a posteriori, et vraisemblance des données.

cas de figure est très proche du probability-matching. Or, dans cette situation, on observe que le pic de la loi de probabilité de l'a posteriori se situe à 0.9, comme dans les données. Concrètement, cela signifie que dans un tel cas de figure, les locuteurs convergeraient très rapidement vers un taux de régularisation de 90%. Cela n'est pas ce que nous observons dans nos résultats, ce qui nous permet donc d'exclure l'idée que les locuteurs abordent le traitement des CAL avec un a priori uniforme.

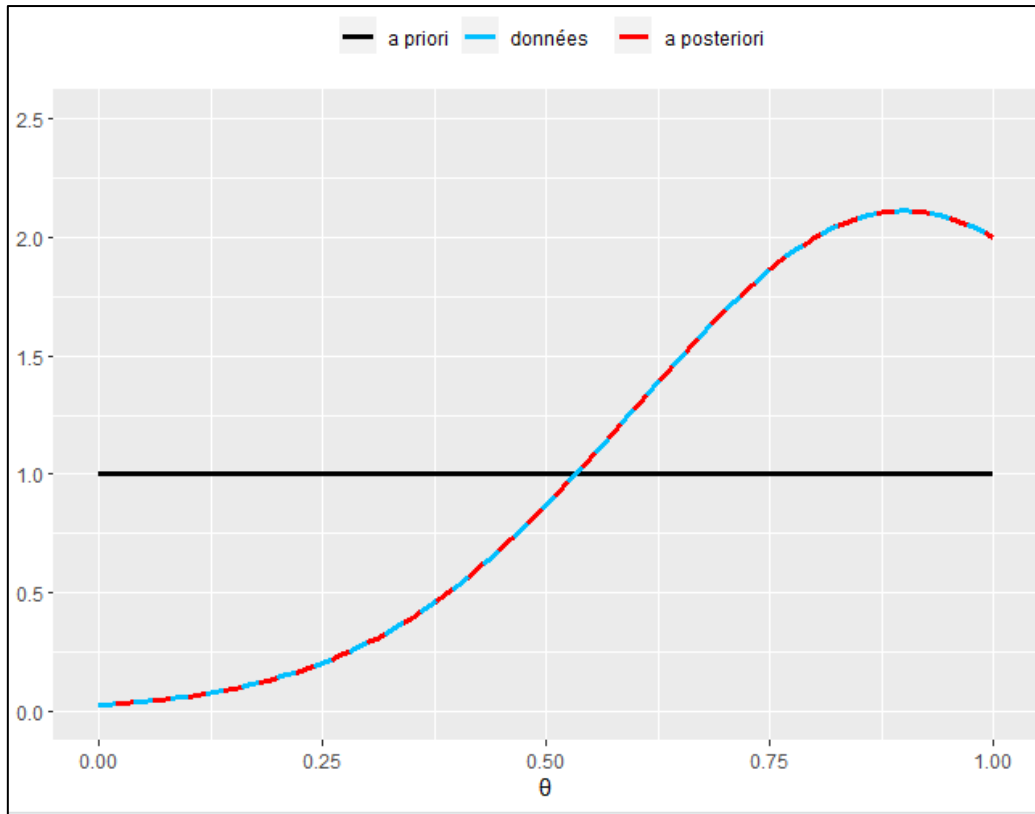


Fig.25 : apprentissage bayésien avec un a priori plat ($f(\theta) = 1$) et des données définies par une loi normale ($\mu=0.9$; $\sigma=0.3$).

Une autre manière de concevoir les a priori des locuteurs est d'envisager qu'ils ont une préférence pour une certaine valeur, mais sans grande certitude. Autrement dit, comme dans le cas des données dans notre cas de figure précédent, on peut conceptualiser les a priori comme suivant une loi normale centrée autour d'une certaine valeur. La moyenne de cette loi normale correspond à la valeur que l'on suppose être préférée par les locuteurs. À défaut d'arguments en faveur d'une valeur plutôt qu'une autre, on peut placer cette moyenne sur la valeur neutre de 0.5. Quant à l'écart type, il représente ici la certitude que les locuteurs ont à l'égard de cet a priori. Puisqu'il s'agit d'un a priori motivé par peu de données sur les CAL, nous n'avons pas

de raison de croire que cet a priori est particulièrement fort. Nous allons donc explorer un cas de figure dans lequel les locuteurs ont un a priori correspondant à une loi normale de moyenne 0.5 et d'écart type élevé de 0.3. Concernant les données, nous conservons la loi normale de moyenne 0.9 et d'écart type 0.3, pour les mêmes raisons que précédemment. Concrètement, ce cas de figure se différencie du précédent en cela que le taux de régularisation a priori d'un locuteur choisi au hasard a plus de chance de graviter autour de 0.5. Toutefois, l'a priori étant assez peu fiable, cette valeur peut varier assez fortement autour de 0.5 d'un locuteur à l'autre.

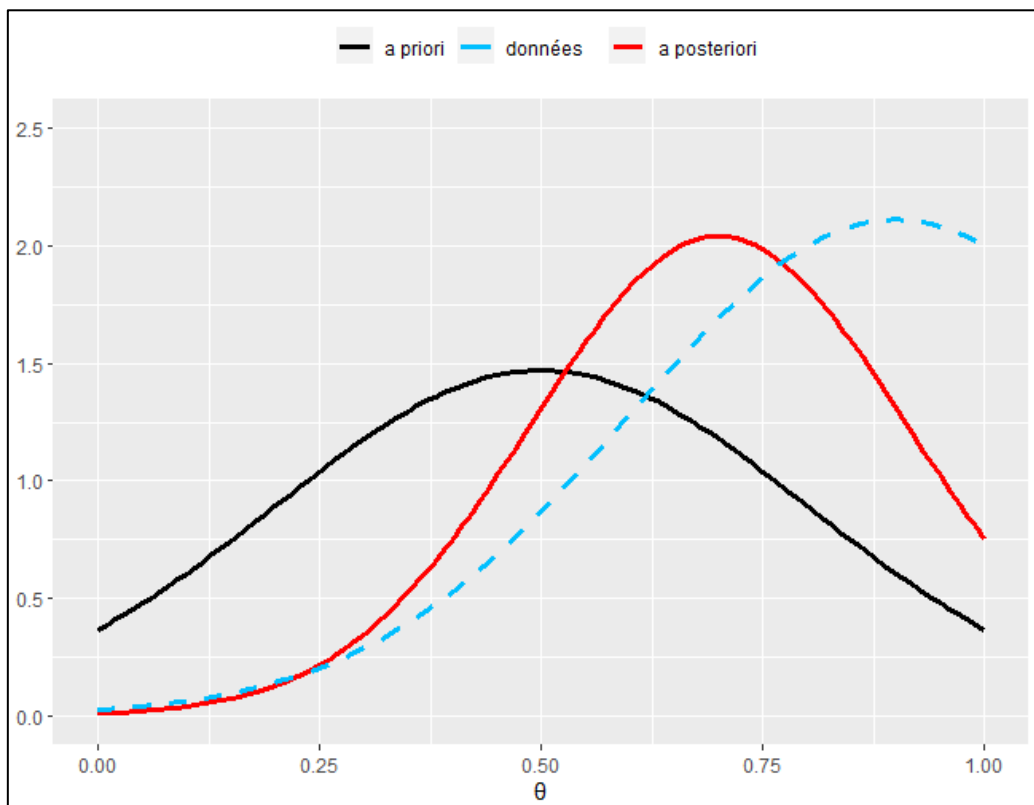


Fig.26 : apprentissage bayésien avec un a priori défini par une loi normale ($\mu=0.5$; $\sigma=0.3$) et des données définies par une loi normale ($\mu=0.9$; $\sigma=0.3$).

La figure 26 ci-dessus présente les courbes des fonctions relatives à ce cas de figure. On peut observer que dans ce cas de figure, le pic de la courbe a posteriori se situe à 0.7, à mi-chemin entre celui de la courbe de l'a priori (0.5) et celui de la courbe des données (0.9). Ce cas de figure est bien plus proche des résultats obtenus dans nos expériences et nous invite donc à penser que les locuteurs abordent l'acquisition de nouveaux patterns morphologiques avec un certain a priori qui fait dévier leur comportement du probability-matching. Bien entendu, à mesure que les données s'accumulent, cet a priori initial jouera un rôle de moins en moins

important dans les probabilités a posteriori. Un tel résultat pourrait être obtenu en réduisant l'écart type de la loi normale décrivant la vraisemblance des données, ce qui déplacerait et resserrerait le pic de la courbe a posteriori vers 0.9.

Dans le cas de l'utilisation du pattern régulier pour les CAL *sans conflit*, nous n'avions pas de raison de penser que l'a priori des locuteurs serait particulièrement fort, ni qu'il pencherait en faveur d'un pattern plutôt qu'un autre, et nous l'avons donc placé autour d'une valeur neutre de 0.5 avec un écart type élevé. Cependant, il peut arriver que nous ayons de bonnes raisons de croire que l'a priori des locuteurs est fortement biaisé dans une direction. Par exemple, si l'on s'intéresse à l'utilisation du pattern régulier dans les CAL /Q/, il paraît logique que les locuteurs partent avec un a priori très négatif contre le pattern régulier, puisqu'il enfreint une contrainte quasiment inviolable de la langue. Pour illustrer cette situation, représentons maintenant les a priori par une loi normale de moyenne 0 et avec un écart type très faible de 0.01. Concernant la vraisemblance des données il est plus difficile dans ce cas de décrire la loi normale qui décrirait au mieux les données observées par les participants. Si l'on s'en tient aux données du lexique concernant les CAL /Q/ uniquement, alors celle-ci devrait également suivre une loi normale de moyenne 0 et d'un écart type faible. Mais on peut également considérer que l'utilisation du pattern régulier en général a une influence, ce qui serait indispensable pour rendre compte de situation où un pattern général largement utilisé éclipse un pattern contextuel. Auquel cas, la vraisemblance des données pourrait suivre une loi normale plus proche d'une moyenne de 0.75, correspondant à la proportion d'utilisation du pattern régulier dans l'ensemble des CAL. Quoi qu'il en soit, nous savons que l'a priori des locuteurs est si fort qu'aucune des données auxquelles ils ont été exposés (dans leur emploi de la langue ou dans nos expériences) ne fait sensiblement évoluer leurs préférences. Avec l'a priori que nous avons défini ci-dessus, nous observons un résultat similaire même si on représente la vraisemblance des données par une loi normale avec une moyenne de 0.75 et le même écart type élevé de 0.3 utilisé précédemment. La figure 27 ci-dessous représente les courbes des fonctions relatives à cette situation. On observe cette fois-ci que les probabilités a posteriori sont quasiment identiques aux probabilités a priori, indiquant que les données ont eu un effet quasi nul sur ces derniers.

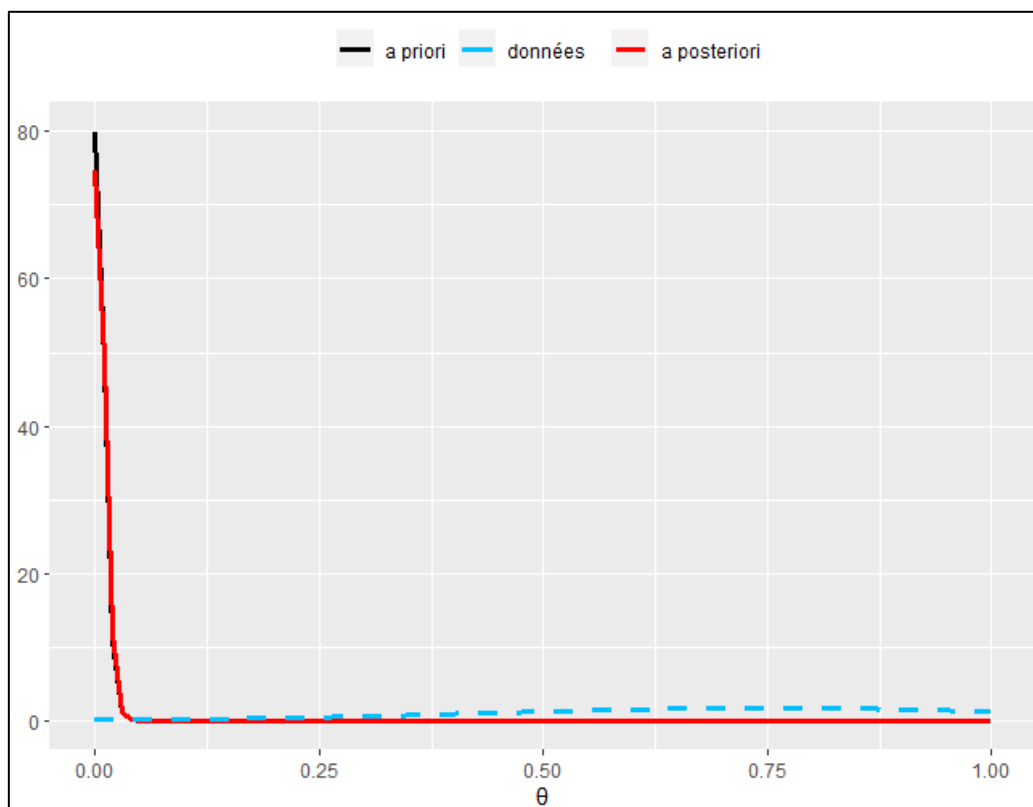


Fig.27 : apprentissage bayésien avec un a priori défini par une loi normale ($\mu=0$; $\sigma=0.01$) et des données définies par une loi normale ($\mu=0.75$; $\sigma=0.3$).

Les choses sont plus complexes lorsque l'on s'intéresse à l'utilisation du pattern régulier pour les CAL /R/. Pour rappel, dans le lexique, ce pattern est utilisé dans environ 35% des cas, tandis que dans l'expérience baseline et la première phase de l'expérience ESP, il est utilisé dans un peu plus de 40% des cas. On pourrait faire l'hypothèse, comme pour les CAL *sans conflit*, que les locuteurs ont un a priori autour de la valeur neutre de 0.5, et qu'ils sont exposés à des données centrées autour d'une moyenne de 0.35, mais cela ne serait pas très cohérent avec la situation linguistique pour plusieurs raisons. Premièrement, si les locuteurs ont le même a priori pour les CAL /R/ et *sans conflit*, comment pourrait-on justifier qu'ils prennent en compte des données radicalement opposées dans le traitement de ces deux types d'items ? Rien ne justifierait que les CAL *sans conflit* soient traités différemment des CAL /R/, et ces derniers devraient en tout logique être influencés par le mêmes données (représentées par une loi normales de moyenne 0.9). Deuxièmement, nous savons que le pattern régulier entraîne la violation d'une contrainte phonologique de la langue lorsqu'il est appliqué à un CAL /R/. Il semblerait donc plus cohérent que les locuteurs aient un a priori légèrement négatif sur l'utilisation du pattern régulier dans cette situation. Une manière de rester cohérent avec les

résultats de nos expériences serait alors de considérer que c'est l'a priori des locuteurs qui gravite autour d'un taux de régularisation moyen de 0.35, et les données qu'ils observent qui gravitent autour d'un taux de régularisation moyen de 0.5. Si les données du lexique indiquent que le taux de régularisation pour les CAL /R/ est bien de 0.35, on sait que le pattern régulier est également très souvent utilisé de manière générale. Cet emploi très fréquent du pattern régulier dans l'ensemble des CAL pourrait influencer sur la manière dont les locuteurs perçoivent la vraisemblance des données pour le cas spécifique des CAL /R/, favorisant légèrement l'emploi du pattern régulier pour ces derniers, ce qui serait cohérent avec les observations de Kapatsinski (2010). La figure 28 ci-dessous présente donc une situation dans laquelle les probabilités a priori suivent une loi normale avec une moyenne de 0.35 et la vraisemblance des données suit une loi normale avec une moyenne de 0.5. Les deux lois normales ont un écart type élevé de 0.3. On constate que le pic de la courbe des probabilités a posteriori est bien autour de 0.4.

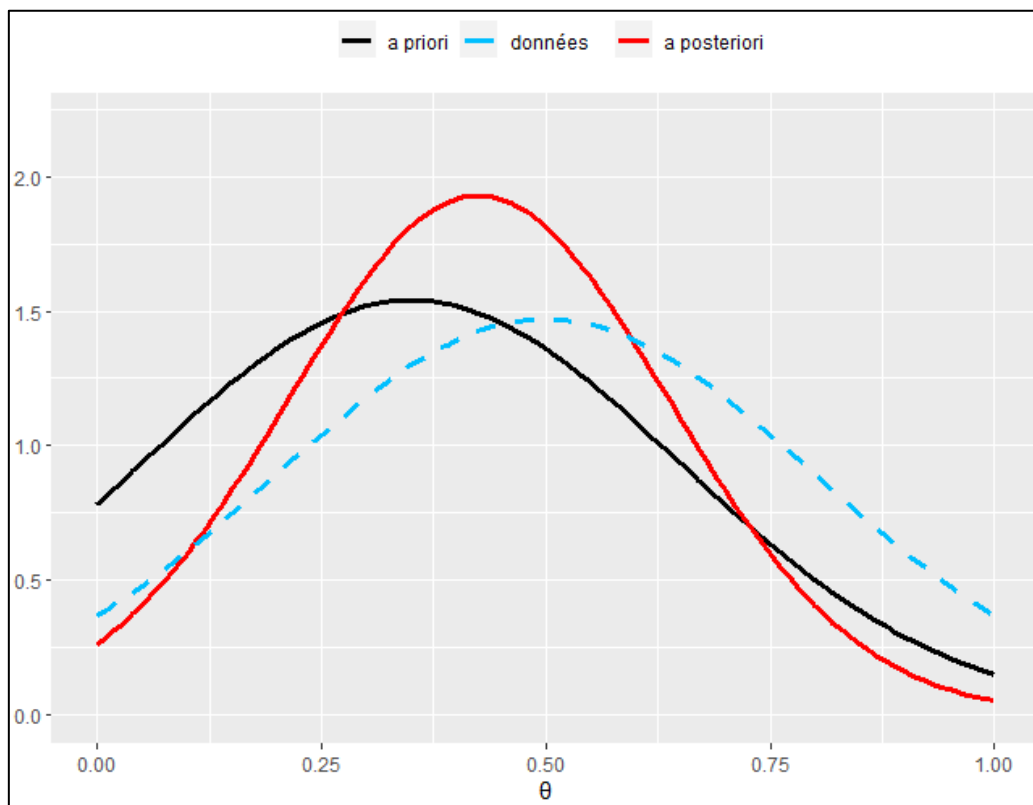


Fig.28 : apprentissage bayésien avec un a priori défini par une loi normale ($\mu=0.35$; $\sigma=0.3$) et des données définies par une loi normale ($\mu=0.5$; $\sigma=0.3$).

Voyons maintenant le cas de l'utilisation du pattern *yodan* dans la formation de nouveaux verbes. Ce cas est intéressant en cela que l'on peut constater que seuls des nouveaux verbes *yodan* sont effectivement créés en japonais, ce qui indique que les locuteurs ont un très fort a priori favorisant ce pattern. Pourtant, lorsqu'ils créent de nouvelles formes, leur comportement semble être influencé par les probabilités observées dans l'ensemble du lexique, où seulement 40% des verbes se terminant en *ru* sont des *yodan*. Toutefois, les résultats de l'expérience baseline et de la première phase de l'expérience ESP montrent un taux d'utilisation du pattern *yodan* de plus de 80% en moyenne, ce qui signifie que l'a priori des locuteurs joue un rôle majeur dans leurs décisions. Dans ce cas de figure, l'a priori des locuteurs sera représenté par une loi normale avec une moyenne de 1 et un écart type très faible de 0.05. Quant à la vraisemblance des données observées elle sera représentée par une loi normale avec une moyenne de 0.4, et un écart type de 0.07, également très faible, bien que légèrement supérieur à celui de l'a priori. La figure 29 ci-dessous présente les courbes des fonctions relatives à ce cas de figures. On peut constater que malgré un a priori dont le pic est très resserré sur la valeur de 1, les données influent sur celui-ci de telle manière que le pic de la courbe des a posteriori se situe autour de 80%.

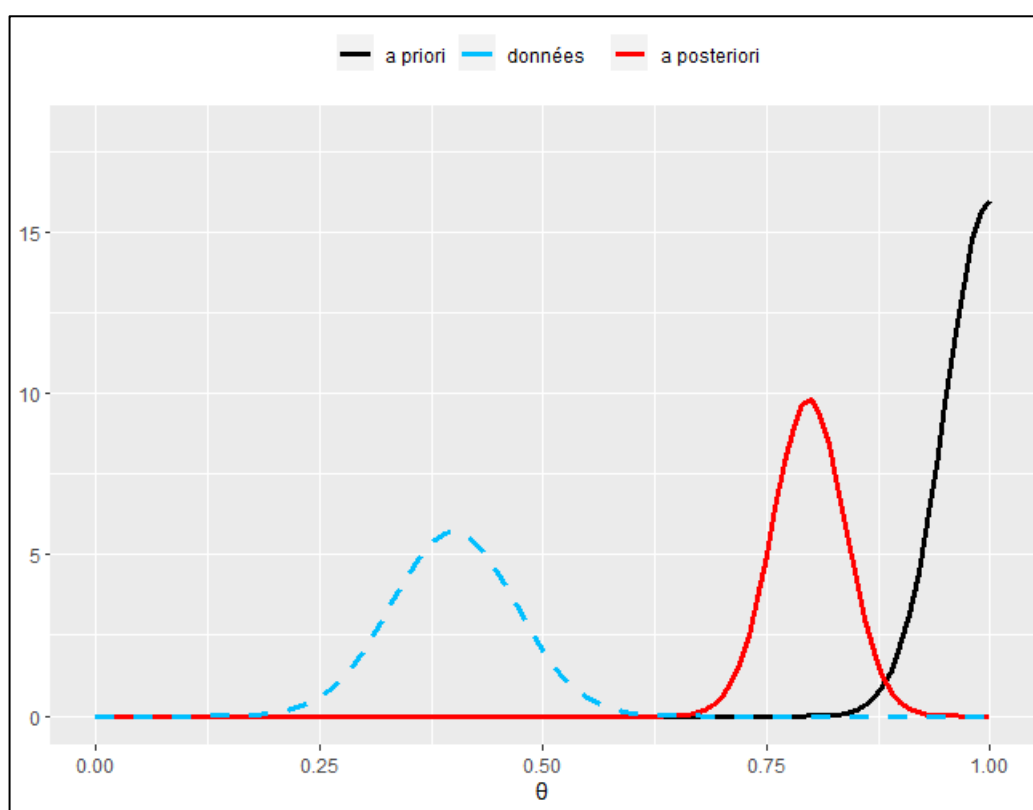


Fig.29 : apprentissage bayésien avec un a priori défini par une loi normale ($\mu=1$; $\sigma=0.05$) et des données définies par une loi normale ($\mu=0.4$; $\sigma=0.07$).

Enfin, intéressons-nous à présent à la manière dont la convergence morphologique peut être décrite dans ce cadre théorique. Lorsqu'un locuteur converge vers le comportement d'un partenaire d'interaction, il accorde une importance particulière au nouvel input que ce dernier lui fournit, par rapport à l'ensemble de ses expériences passées. En d'autres termes, il considère que les données fournies par le partenaire sont particulièrement fiables dans cette situation. Cependant, dans les expériences ESP que nous avons menées, le partenaire artificiel ne produit que 40 formes, ce qui ne devrait pas être d'une grande fiabilité, quel que soit l'a priori du locuteur. Si l'on prend en compte le principe de base du raisonnement bayésien, il serait plus cohérent que ce soit la fiabilité de l'a priori qui diminue dans les situations propices à la convergence. En effet, le but initial du théorème de Bayes est d'évaluer la probabilité de la véracité d'une certaine hypothèse. Dans une situation d'interaction, le locuteur ne cherche pas simplement à déterminer s'il doit utiliser un certain pattern, mais également la propension de son interlocuteur à utiliser ce pattern. Les hypothèses testées ne concernent plus « la probabilité d'utiliser tel pattern » mais plutôt « la probabilité que mon partenaire utilise tel pattern ». Dans le premier cas, l'a priori du locuteur est simplement basé sur sa propre expérience et ses propres préférences intrinsèques, mais dans le deuxième cas son a priori doit intégrer le comportement d'un autre locuteur ce qui apporte son lot d'incertitudes. Dans notre modèle d'apprentissage bayésien, la fiabilité est caractérisée par l'écart type des différentes lois normales utilisées. Ce changement dans la nature des hypothèses testées peut donc se traduire par la modification de l'écart type de l'a priori. Par ailleurs, en ce qui concerne la moyenne de l'a priori, on peut considérer que les locuteurs se basent sur leurs propres préférences pour formuler leurs estimations sur le comportement des autres locuteurs.

Reprenons l'exemple de l'utilisation du pattern *yodan* dans la création de nouveaux verbes. Nous savons que les participants à l'expérience ESP utilisent ce dernier dans environ 80% des cas lors de la phase 1. Nous connaissons donc la propension moyenne des locuteurs à utiliser le pattern *yodan* avant qu'ils ne soient confrontés à leur partenaire d'interaction. Cette valeur peut alors nous servir de moyenne pour l'a priori avec lequel les locuteurs abordent la situation d'interaction. Quant à l'écart type, nous utiliserons une valeur élevée de 0.3 pour refléter la difficulté d'évaluer a priori le comportement d'un autre locuteur¹⁸. Lorsqu'un

¹⁸ Rappelons que plusieurs études suggèrent que certains locuteurs sont naturellement plus enclins à la convergence que d'autres (Beckner et al., 2016; Sonderegger et al., 2017). Pour représenter cette variabilité, dans un modèle

participant produisant 80% de formes *yodan* en phase 1 est confronté au robot « sous-régularisateur », ce dernier produira 55% de formes *yodan* en phase 2. La moyenne pour la vraisemblance des données sera donc de 0.55. L'écart type aura une valeur élevée de 0.3 également, pour refléter la faible quantité de données auxquelles le participant est exposé (en comparaison à l'immensité de l'expérience de la langue au quotidien). Le figure 30 ci-dessous présente les courbes des fonctions relatives à ce cas de figure. On peut constater que le pic de la courbe des a posteriori se situe autour de 70%, ce qui correspond à la proportion moyenne de formes *yodan* produites en phase 3 par les participants exposés au robot « sous-régularisateur ».

Le même raisonnement est applicable dans tous les cas de convergence observés dans nos expériences. Du moment que l'écart type est identique pour les lois normales décrivant l'a priori et la vraisemblance des données, ce qui signifie que les participant accordent autant d'importance à leurs propres préférences qu'au nouvel input que leur fournit leur partenaire, la moyenne de l'a posteriori se situera à mi-chemin entre les moyennes de ces deux lois normales. Ainsi, le robot « sur-régularisateur » entraîne une augmentation de l'utilisation du pattern en question, le robot « sous-régularisateur » entraîne une diminution de son utilisation, et le robot « sans changement » resserre le pic de la courbe des a posteriori autour du pic commun de la courbe des a priori et de la vraisemblance des données.

plus complexe, on pourrait envisager que l'écart type de l'a priori concernant des hypothèses impliquant le comportement d'un partenaire d'interaction soit lui-même déterminé par une loi normale plutôt qu'une valeur fixe.

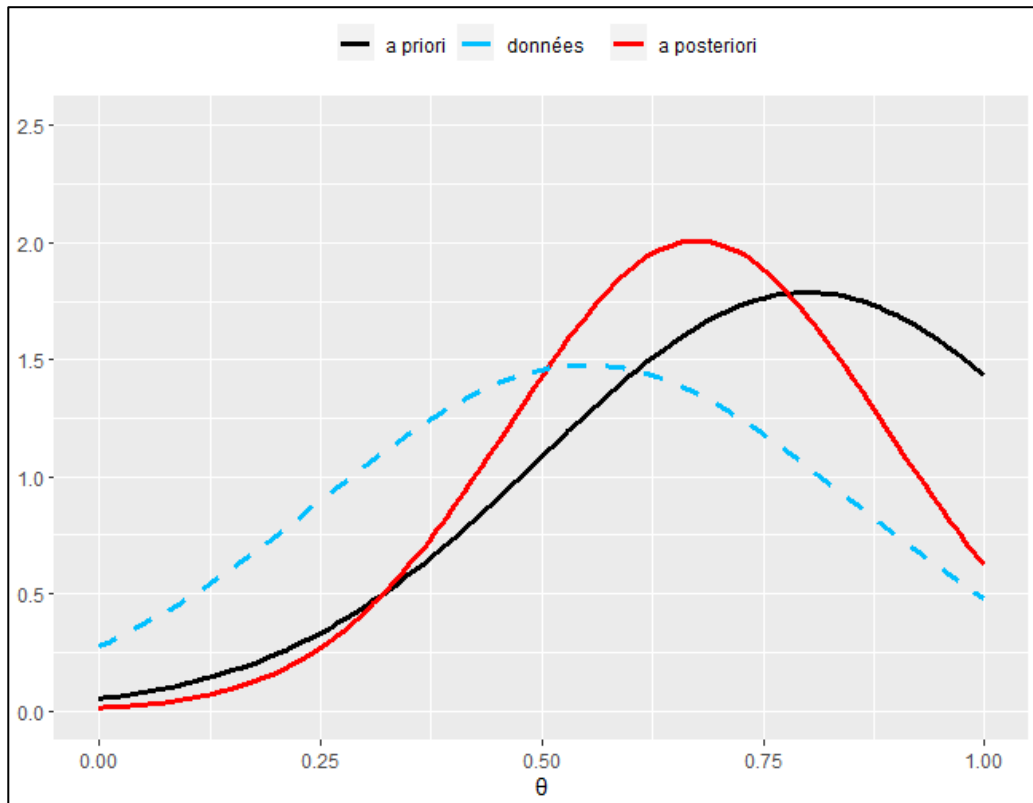


Fig.30 : apprentissage bayésien avec un a priori défini par une loi normale ($\mu=0.8$; $\sigma=0.3$) et des données définies par une loi normale ($\mu=0.55$; $\sigma=0.3$).

Notons toutefois que nous supposons dans ce cas de figure que lors de la phase 3, le comportement des participants suit les probabilités a posteriori obtenues par l'évaluation d'hypothèses impliquant leur partenaire d'interaction en phase 2, ce qui peut être surprenant étant donné que ce partenaire n'est plus présent lors de la phase 3. Cela soulève de nouveau l'épineuse question du lien entre convergence à court terme et évolution linguistique sur le long terme. Les conditions expérimentales encouragent probablement les participants à adopter un comportement cohérent, dans ce cas dépendant du partenaire artificiel, jusqu'à la fin de la tâche. Cela permettrait à l'influence du partenaire artificiel de perdurer après son départ. Cependant, en théorie, l'importance de l'input fournit par le partenaire artificiel devrait se fondre dans la masse de l'input total des locuteurs avec le temps, à partir du moment où ces derniers cessent d'incorporer leur partenaire dans leurs hypothèses. Ce fonctionnement permettrait d'observer des changements rapides et fréquents sur le court terme mais rares sur le long terme, comme cela a été observé par l'étude longitudinale de Sonderegger et al. (2017, voir section 1.3.3). Nous devons cependant souligner une nouvelle fois que les données longitudinales sur le sujet

sont trop rares pour pouvoir tirer des conclusions. Nous pourrions nous intéresser à cette question dans de futures recherches, par exemple, en faisant une nouvelle série de tâche de création de nouvelles formes quelques jours, semaines, ou mois après la passation de l'expérience ESP, afin de mesurer ce qui resterait alors de l'influence du partenaire au fil du temps. Notre hypothèse est qu'en l'absence de nouvelle interaction, l'effet de la convergence à court terme tel qu'observé dans nos expériences ESP ne serait plus visible après quelques jours ou semaines, les nouveaux inputs relatifs aux expériences ayant été absorbés dans la masse de l'input général observé par les participants. Toutefois, cet input relatif à l'expérience resterait présent en mémoire et pourrait contribuer à un changement de long terme s'il venait à être renforcé dans les futures interactions des participants.

Nous avons vu qu'une série de raisonnements bayésiens nous permet d'aboutir aux résultats de nos diverses expériences. Mais au-delà de la simple adéquation entre ce cadre théorique et nos résultats, il ressort des différents cas de figure passés en revue plusieurs observations qui nous permettent de combler les lacunes des vagues approximations formulées sur la base d'un comportement de probability-matching. Ces observations apportent des réponses à notre deuxième question de recherches concernant les facteurs qui entrent en jeu dans la construction des représentations qui guident la productivité morphologique, en dehors des probabilités observées dans les formes existantes.

Tout d'abord, même dans le cas de processus morphologiques récents, les locuteurs n'abordent pas la création de nouvelles formes avec un a priori uniforme. Ceci reviendrait peu ou prou à un comportement de probability-matching, qui s'avère insuffisant pour expliquer la productivité morphologique. Les locuteurs possèdent des a priori (non uniformes) qui jouent un rôle important notamment quand l'input auquel ils ont accès est limité.

Toutefois, un a priori non uniforme ne signifie pas nécessairement une préférence pour un pattern ou un autre. En effet, l'a priori peut-être centré autour d'une valeur neutre. Auquel cas, celui-ci aura pour effet d'éviter que le comportement du locuteur bascule dès la première occurrence d'un pattern rencontrée. Autrement dit, dans cette situation les locuteurs ont tout de même besoin d'une certaine quantité de données avant que l'on puisse observer un changement notable dans leur utilisation d'un pattern. Ce critère peut ne pas être rempli par certains patterns contextuels dans des processus morphologiques récents. C'est le cas pour le pattern de

Suppression dans la formation des CAL qui, par conséquent, s'avère peu productif malgré la proportion importante des formes irrégulières qu'il représente.

En revanche, un a priori peut être particulièrement fort en faveur (ou défaveur) d'un pattern. Certains patterns peuvent être naturellement préférés ou évités pour des raisons linguistiques. Par exemple, les locuteurs japonais ont un a priori extrêmement négatif sur l'emploi du pattern régulier pour la formation d'un CAL /Q/, puisque cela enfreindrait une contrainte phonologique très sévère de la langue.

Enfin, si la force d'un a priori peut être motivée par des raisons linguistiques, celle-ci peut également varier en fonction de la situation. Par exemple, dans les situations propices à la convergence, le fait que les locuteurs accordent plus d'importance au nouvel input reçu de leur partenaire d'interaction peut simplement découler du fait qu'ils doivent inclure le comportement d'une autre personne dans les informations à prendre en compte dans leurs prises de décision. Cette personne est nécessairement une source d'incertitude, par rapport aux simples préférences du locuteur lui-même, et amoindrit d'autant la fiabilité de l'a priori des locuteurs, rendant les nouvelles données plus importantes par contraste.

9.3 Perspectives et conclusion

L'approche bayésienne s'avère être un excellent cadre théorique pour décrire ces phénomènes car le théorème de Bayes contient tous les outils théoriques pour tenir compte à la fois des probabilités observées dans les formes existantes, représentées par la vraisemblance des données $p(x/\theta)$, et des préférences intrinsèques des locuteurs, représentées par l'a priori $f(\theta)$. De plus, le fait que cette approche nous renseigne sur les probabilités de différents continuums d'hypothèses permet de tenir compte de la variation qui existe à différents niveaux, que ce soit dans les comportements intra ou inter-individus ou dans les données. Ce modèle d'apprentissage est donc particulièrement efficace pour décrire des processus morphologiques soumis à une forte variation, que ce soit parce qu'ils sont apparus récemment dans la langue, ou employés dans des situations propices à la convergence.

Les nombreuses études soulignant l'importance des probabilités de l'input dans la productivité des patterns morphologiques témoignent de l'attention particulière qui a été portée jusqu'à présent sur les données disponibles pour les locuteurs. En revanche, les études portant

sur le fonctionnement des a priori des locuteurs sont plutôt rares. À l'avenir, nous souhaiterions explorer davantage cet aspect en étudiant d'autres situations dans lesquelles les a priori jouent un rôle important. Les processus morphologiques devenus récemment productifs sont des cas d'étude idéaux pour cela, puisque l'importance des données disponibles pour les locuteurs y est forcément limitée.

Un processus morphologique qui attire particulièrement notre attention est celui de la formation des mots de *verlan* en français. Ce processus de formation de mots est utilisé dans les banlieues depuis la fin de la Seconde Guerre mondiale et a été popularisé dans les années 1990 avec l'essor du mouvement hip-hop. Il permet de former de nouveaux mots argotiques à partir d'un mot de base par l'inversion de ses syllabes. Lors du processus de verlanisation, le mot peut subir d'autres modifications, telles que des épenthèses ou des élisions. Le mot *verlan*, est lui-même une verlanisation de la locution « à l'envers » avec une élision de la préposition « à ». L'exemple canonique d'un mot de *verlan* prend pour base un mot d'une structure de type $C_1V_1C_2V_2$, qui est remodelé selon une structure $C_2V_2C_1V_1$. Par exemple, 'choper' /ʃope/ devient 'pécho' [peʃo], et 'tomber' /tɔ̃be/ devient 'béton' [betɔ̃]. Les modifications que peuvent subir les mots de base lors de leur verlanisation tendent à les rapprocher de cette forme canonique. Ainsi, on peut voir apparaître des patterns spécifiques à certains contextes phonologiques. Par exemple, dans le cas où le mot de base est monosyllabique de structure $C_1V_1C_2$, il subit fréquemment l'épenthèse d'un schwa final avant l'inversion des syllabes, puis la voyelle V_1 maintenant en position finale est élidée. Par exemple 'mec' /mæk/ > mækə > kəme > [kəm] 'keum', ou bien 'sœur' /sœʁ/ > sœʁə > kəsœ > [kəs] 'reuss'. Bien que ce pattern soit très fréquent, il n'est pas le seul employé. Dans quelques cas, la voyelle V_1 n'est pas élidée, comme le montre les exemples 'louche' /luʃ/ > luʃə > [ʃəlu] 'cheulou' et 'rap' /ʁap/ > ʁapə > [pəʁa] 'peura'. En réalisant des tâches de création de nouvelles formes sur le *verlan* avec des locuteurs français, nous pourrions vérifier si le pattern canonique et les différents patterns contextuels ont la productivité qui serait attendue au vu de leur fréquence dans le lexique. Notre hypothèse est que, comme dans le cas des CAL japonais, les locuteurs français seraient capables de reproduire le pattern général pour la formation de mots *verlan* à partir de mots bisyllabiques, mais qu'ils ne montreraient en revanche aucune préférence pour les différents patterns contextuels pour la formation de mots *verlan* à partir de mots monosyllabiques, malgré la prédominance de l'un de ces patterns dans les formes existantes. Vérifier cette hypothèse nous permettrait de confirmer que les locuteurs n'ont pas un a priori uniforme même lorsqu'ils abordent un processus

morphologique récent. Mais qu'au contraire, leurs a priori, même s'ils sont neutres, guident leur comportement jusqu'à ce qu'une quantité suffisante de données leur soit présentées.

Le cas du *verlan*, comme celui des CAL japonais, n'est bien entendu qu'un exemple parmi d'autres de nouveau processus morphologique. D'autres études sur divers processus morphologiques pourraient apporter différents éclairages sur la nature des a priori des locuteurs concernant la productivité des patterns. Cependant, une autre piste à explorer davantage dans de futures recherches est celle de la convergence morphologique. Ces études nous permettraient d'approfondir nos connaissances sur la manière dont les locuteurs incorporent leurs partenaires dans leurs a priori sur le comportement à adopter dans des situations d'interaction. Mais elles pourraient aussi remédier à certaines des limitations des études présentées dans cette thèse.

Tout d'abord, dans le paradigme ESP utilisé dans nos travaux, le comportement des participants ne se manifeste que par leur sélection d'une réponse parmi un choix multiple. Bien que cela soit suffisant pour vérifier l'existence du phénomène de convergence et valider son intérêt en tant que sujet de recherche, cela pose la question du caractère écologique des données collectées. Plusieurs améliorations peuvent être envisagées pour rendre plus authentique l'interaction entre le participant et son partenaire ainsi que la manière dont la création de nouvelles formes est élicitée. La tâche de création de nouvelles formes pourrait être intégrée dans un jeu sérieux auquel plusieurs participants (humains) pourraient jouer en même temps. De tels protocoles sont déjà employés, que ce soit en dehors du cadre de la linguistique avec notamment le jeu de labélisation d'images qui sert de cadre à l'expérience ESP originale de Von Ahn & Dabbish (2004), mais également dans des tâches linguistiques telles que celle présentée par Lev-Ari & Peperkamp (2014) employant une variante du jeu des sept familles pour étudier l'adaptation phonologique de mots d'emprunt. Même dans le contexte de passations en ligne, les interactions entre les joueurs pourraient être plus libres en leur permettant de chatter (textuellement ou vocalement), du moment que le but du jeu les encourage à créer de nouveaux mots. Bien entendu, le jeu devrait être adapté au processus morphologique étudié. Par exemple, dans le cas des flexions verbales japonaises, nous pourrions apprendre aux joueurs des associations entre des actions et des pseudo-verbos (à la forme inaccomplie) dans une phase d'entraînement. Puis on présenterait à un joueur un personnage qui effectue une série d'actions et il serait demandé à ce joueur de décrire la suite d'action à son partenaire de jeu, ce dernier devant reproduire la suite d'actions avec son propre personnage pour que les deux joueurs gagnent la partie. En laissant les joueurs communiquer librement, cela devrait éliciter

des phrases avec des pseudo-verbos à la forme accomplie (du type « D'abord, le personnage a Xé, ensuite il a Yé, et à la fin il a Zé ») de manière plus écologique. On pourrait observer la manière dont les deux joueurs s'influencent en inversant les rôles des joueurs au cours de différentes parties et/ou en incluant une phase de pré-test et de post-test dans le protocole. Quoiqu'il en soit, tout effort dans le but d'obtenir des interactions plus authentiques ne pourrait que consolider les conclusions de ces potentielles futures études.

Une seconde limitation de nos expériences, que nous avons déjà mentionnée précédemment, est qu'elles ne nous renseignent que sur le phénomène de convergence à court terme. Le paradigme ESP a l'avantage de proposer une phase de post-test qui se déroule après le départ du partenaire artificiel et nous permet ainsi d'observer l'effet de la convergence quelques instants au-delà de la phase d'interaction. Cependant, nous ne pouvons qu'émettre des suppositions sur l'effet à moyen et long terme de la convergence. L'approche de Sonderegger et al. (2017) consistant à exploiter les enregistrements d'émissions de télé-réalité comme corpus pour étudier l'effet de la convergence sur plusieurs mois et semaines nous paraît particulièrement astucieuse. Malheureusement, si cette méthode peut s'avérer efficace pour l'étude de la variation de l'accent des locuteurs, il nous semble qu'elle ne serait pas aussi fructueuse pour l'étude de la morphologie. En effet, les locuteurs ne créent pas de nouveaux mots chaque jour, et même l'exploitation de mois d'enregistrement pourrait ne donner qu'une faible quantité de données pertinentes pour comprendre la convergence au niveau morphologique. Dans l'attente de solutions plus optimales, les futures études dans ce domaine devront probablement continuer de faire appel à l'élicitation dans un cadre expérimental. Notons que des protocoles expérimentaux incluant des réalisations de tâches étalées dans le temps ont déjà révélé que les locuteurs gardaient en mémoire des détails des épisodes passés relatifs à la passation de l'expérience influant, entre autres, sur leur reconnaissance et leur prononciation de mots dans le même contexte deux semaines plus tard (Goldinger, 1996; Goldinger & Azuma, 2004). Il est donc envisageable d'étudier la convergence morphologique dans des études longitudinales dans lesquelles les participants effectueraient des tâches de création de nouvelles formes, sous la forme de sessions de jeu permettant divers types d'interactions, à différents intervalles de temps.

Pour conclure, nous souhaiterions souligner les implications de nos résultats pour notre compréhension du fonctionnement du langage dans son ensemble. Le théorème de Bayes a de multiples applications et est, entre autres, employés pour modéliser divers processus cognitifs

humains (Griffiths et al., 2008). Le fait que l'émergence de patterns morphologiques repose sur un jeu d'a priori continuellement et graduellement modifiés avec chaque nouvel input rencontré par les locuteurs montre que ce processus suit également un raisonnement bayésien. Si les représentations qui guident la productivité morphologique sont formées sur la base de ce raisonnement qui joue par ailleurs un rôle important dans la cognition en général, il ne serait pas étonnant qu'il joue également un rôle dans la formation des représentations à d'autres niveaux linguistiques. La prise en compte des a priori des locuteurs pourrait donc considérablement modifier notre conception des représentations linguistiques en général. Or, l'une des critiques récurrentes du théorème de Bayes est que l'a priori est souvent déterminé de manière plus ou moins subjective, selon les données disponibles (Sprenger, 2018). Pour réduire cette part de subjectivité, il est donc crucial d'améliorer notre compréhension des a priori linguistiques afin de les intégrer efficacement dans les modèles théoriques décrivant les représentations linguistiques à différents niveaux.

Références :

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses : A computational/experimental study. *Cognition*, 90(2), 119-161.
- Anttila, A. (1997a). Deriving variation from grammar : A study of Finnish genitives. *Variation, change and phonological theory*, 35-68.
- Anttila, A. (1997b). *Variation in Finnish phonology and morphology*. Ph.D dissertation, Stanford CA: Stanford University.
- Arvaniti, A. (2009). Rhythm, Timing and the Timing of Rhythm. *Phonetica*, 66(1-2), 46-63. <https://doi.org/10.1159/000208930>
- Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. Dans H. Guetzhaw (Éd.), *Groups, leadership, and men* (pp. 177-190). Pittsburgh, PA: Carnegie Press.
- Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, 193, 31-35.
- Asch, S. E. (1956). Studies of independence and conformity : I. A minority of one against a unanimous majority. *Psychological Monographs : General and Applied*, 70(9), 1-70. <https://doi.org/10.1037/h0093718>
- Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, 39(4), 437-456. <https://doi.org/10.1017/S0047404510000400>
- Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40(1), 177-189. <https://doi.org/10.1016/j.wocn.2011.09.001>
- Babel, M., & Bulatov, D. (2012). The Role of Fundamental Frequency in Phonetic Accommodation. *Language and Speech*, 55(2), 231-248. <https://doi.org/10.1177/0023830911417695>
- Batchelder, E. O. (1999). Rule or rote ? Native-speaker knowledge of Japanese verb inflection. Dans *Proceedings of the second international conference on cognitive science* (p. 6).
- Batchelder, E. O., & Ohta, A. (2000). Rule vs. Rote in Japanese verb inflection. Dans M. Alan & L. Arle (Éds), *LACUS* (Vol. 26, pp. 55-66). Fullerton, CA: LACUS.
- Bauer, L. (2006). *Morphological Productivity* (Vol. 95). Cambridge: Cambridge University Press.
- Becker, M., Ketrez, N., & Nevins, A. (2011). The Surfeit of the Stimulus : Analytic Biases Filter Lexical Statistics in Turkish Laryngeal Alternations. *Language*, 87(1), 84-125. <https://doi.org/10.1353/lan.2011.0016>

- Beckman, M. (1982). Segment Duration and the ‘Mora’ in Japanese. *Phonetica*, 39(2-3), 113-135.
- Beckner, C., Pierrehumbert, J. B., & Hay, J. (2017). The emergence of linguistic structure in an online iterated learning task. *Journal of Language Evolution*, 2(2), 160-176. <https://doi.org/10.1093/jole/lzx001>
- Beckner, C., Rácz, P., Hay, J., Brandstetter, J., & Bartneck, C. (2016). Participants Conform to Humans but Not to Humanoid Robots in an English Past Tense Formation Task. *Journal of Language and Social Psychology*, 35(2), 158-179. <https://doi.org/10.1177/0261927X15584682>
- Berko, J. (1958). The Child’s Learning of English Morphology. *WORD*, 14(2-3), 150-177. <https://doi.org/10.1080/00437956.1958.11659661>
- Bloch, B. (1946). Studies in Colloquial Japanese I Inflection. *Journal of the American Oriental Society*, 66(2), 97. <https://doi.org/10.2307/596327>
- Bloch, B. (1950). Studies in colloquial Japanese IV phonemics. *Language*, 26(1), 86-125.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18(3), 355-387. [https://doi.org/10.1016/0010-0285\(86\)90004-6](https://doi.org/10.1016/0010-0285(86)90004-6)
- Boersma, P. (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam 21*, 43-58.
- Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic inquiry*, 32(1), 45-86.
- Boersma, P., & Weenink, D. (2020). Praat : Doing phonetics by computer (version 6.1.09) [Logiciel].
- Boulis, C., & Ostendorf, M. (2005). A quantitative analysis of lexical differences between genders in telephone conversations. Dans *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics—ACL ’05* (pp. 435-442). Ann Arbor, Michigan: Association for Computational Linguistics. <https://doi.org/10.3115/1219840.1219894>
- Brandstetter, J., Beckner, C., Sandoval, E. B., & Bartneck, C. (2017). Persistent Lexical Entrainment in HRI. Dans *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 63-72). Vienna, Austria: ACM. <https://doi.org/10.1145/2909824.3020257>
- Breen, J. (2004). JMdict : A Japanese-multilingual dictionary. Dans *Proceedings of the Workshop on Multilingual Linguistic Ressources—MLR ’04* (p. 8). Geneva, Switzerland: Association for Computational Linguistics. <https://doi.org/10.3115/1706238.1706248>
- Brennan, S. E., & Clark, H. H. (1996). Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22(6), 1482-1493.
- Bybee, J. (1985). *Morphology : A study of the relation between meaning and form*. Amsterdam ; Philadelphia: J. Benjamins.

- Bybee, J. (1995). Regular Morphology and the Lexicon. *Language and Cognitive Processes*, 10, 425-455.
- Bybee, J. (2001). *Phonology and language use*. Cambridge, England ; New York, NY: Cambridge University Press.
- Bybee, J. (2006). From usage to grammar : The mind's response to repetition. *Language*, 82(4), 711-733.
- Campbell, N. (1999). A Study of Japanese Speech Timing from the Syllable Perspective. *Journal of the Phonetic Society of Japan*, 3(2), 29-39.
- Clark, H. H. (1996). *Using language*. New York: Cambridge University Press.
- Crepaldi, D., Rastle, K., Coltheart, M., & Nickels, L. (2010). 'Fell' primes 'fall', but does 'bell' prime 'ball'? Masked priming with irregularly-inflected primes. *Journal of Memory and Language*, 63(1), 83-99. <https://doi.org/10.1016/j.jml.2010.03.002>
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25(4), 385-400. [https://doi.org/10.1016/0749-596X\(86\)90033-1](https://doi.org/10.1016/0749-596X(86)90033-1)
- Cutler, A., & Otake, T. (1994). Mora or Phoneme ? Further Evidence for Language-Specific Listening. *Journal of Memory and Language*, (33), 824-844.
- Dal, G., & Namer, F. (2018). Playful nonce-formations in French : Creativity and productivity. Dans S. Arndt-Lappe, A. Braun, C. Moulin, & E. Winter-Froemel (Éds), *Dynamics of wordplay : Expanding the Lexicon* (Vol. 5, pp. 203-228). Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9783110501933-205>
- Daland, R., & Pierrehumbert, J. B. (2011). Learning Diphone-Based Segmentation. *Cognitive Science*, 35(1), 119-155. <https://doi.org/10.1111/j.1551-6709.2010.01160.x>
- Davis, S. (2011). Quantity. Dans *The handbook of phonological theory (2nd ed.)* (pp. 103-140). (S.I.): Wiley-Blackwell.
- Dawdy-Hesterberg, L. G. (2014). *The structural and statistical basis of morphological generalization in Arabic* (PhD Thesis). Northwestern University.
- Dawdy-Hesterberg, L. G., & Pierrehumbert, J. B. (2014). Learnability and generalisation of Arabic broken plural nouns. *Language, Cognition and Neuroscience*, 29(10), 1268-1282. <https://doi.org/10.1080/23273798.2014.899377>
- De Chene, B. (1982). The Segmentation of Japanese Verbs : Experimental Evidence. *Journal of Japanese Linguistics*, 8(1-2), 29-62. <https://doi.org/10.1515/jjl-1982-1-203>
- Estival, D. (1985). Syntactic priming of the passive in English. *Text—Interdisciplinary Journal for the Study of Discourse*, 5(1-2), 7-22. <https://doi.org/10.1515/text.1.1985.5.1-2.7>
- Fehringer, C. (2004). How Stable are Morphological Doublets? A Case Study of [ə]/[ø] Variants in Dutch and German. *Journal of Germanic Linguistics*, 16(04). <https://doi.org/10.1017/S1470542704040425>

- Foulkes, P., & Hay, J. B. (2015). The Emergence of Sociophonetic Structure. Dans B. MacWhinney & W. O'Grady (Éds), *The Handbook of Language Emergence* (pp. 292-313). Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118346136.ch13>
- Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, 8(1), 8-11. <https://doi.org/10.1016/j.tics.2003.10.016>
- Giles, H., & Coupland, N. (1991). *Language : Contexts and consequences*. Pacific Grove, CA: Thomson Brooks/Cole Publishing Co.
- Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory : Communication, context, and consequence. Dans H. Giles, N. Coupland, & J. Coupland (Éds), *Contexts of accommodation : Developments in applied sociolinguistics* (pp. 1-68). New York, NY: Cambridge University Press.
- Goldinger, S. D. (1996). Words and Voices : Episodic Traces in Spoken Word Identification and Recognition Memory. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 22(5), 1166-1183.
- Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review*, 11(4), 716-722. <https://doi.org/10.3758/BF03196625>
- Goldsmith, J. A. (1976). *Autosegmental Phonology*. Ph.D dissertation, Massachusetts Institute of Technology.
- Grainger, J., Colé, P., & Segui, J. (1991). Masked morphological priming in visual word recognition. *Journal of Memory and Language*, 30(3), 370-384. [https://doi.org/10.1016/0749-596X\(91\)90042-I](https://doi.org/10.1016/0749-596X(91)90042-I)
- Gregory, S., Webster, S., & Huang, G. (1993). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language & Communication*, 13, 195-217.
- Gries, S. Th. (2005). Syntactic Priming: A Corpus-based Approach. *Journal of Psycholinguistic Research*, 34(4), 365-399. <https://doi.org/10.1007/s10936-005-6139-3>
- Griffiths, T. L., & Kalish, M. L. (2007). Language Evolution by Iterated Learning With Bayesian Agents. *Cognitive Science*, 31(3), 441-480. <https://doi.org/10.1080/15326900701326576>
- Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. Dans R. Sun (Éd.), *Cambridge Handbook of Computational Cognitive Modeling* (pp. 59-100). Cambridge, MA: Cambridge University Press.
- Guillaume, P. (1927)[1973]. The development of formal elements in the child's speech. Dans C. A. Ferguson & D. I. Slobin (Éds), *Studies of child language development* (pp. 240-251). New York: Holt, Rinehart and Winston.
- Gussenhoven, C., & Jacobs, H. (2011). *Understanding Phonology* (Third). Great Britain: Hodder Education.
- Haber, L. R. (1976). Leaped and Leapt : A Theoretical Account of Linguistic Variation. *Foundations of Language*, 14(2), 211-238.

- Hall, M. L., Ferreira, V. S., & Mayberry, R. I. (2015). Syntactic Priming in American Sign Language. *PLOS ONE*, 10(3), 19. <https://doi.org/10.1371/journal.pone.0119611>
- Haspelmath, M. (2002). *Understanding morphology*. London : New York: Arnold ; Oxford University Press.
- Hay, J., & Baayen, H. (2002). Parsing and productivity. Dans G. Booij & J. Van Marle (Éds), *Yearbook of Morphology 2001* (pp. 203-235). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-3726-5_8
- Hayes, B., & Londe, Z. C. (2006). Stochastic phonological knowledge : The case of Hungarian vowel harmony. *Phonology*, 23(1), 59-104. <https://doi.org/10.1017/S0952675706000765>
- Hayes, B., Zuraw, K., Siptár, P., & Londe, Z. (2009). Natural and Unnatural Constraints in Hungarian Vowel Harmony. *Language*, 85(4), 822-863. <https://doi.org/10.1353/lan.0.0169>
- Hepburn, J. C. (1903). *A Japanese-English and English-Japanese Dictionary* (7th éd.). Tokyo: Z. P Maruya & Co.
- Horton, W. S. (2007). The influence of partner-specific memory associations on language production : Evidence from picture naming. *Language and Cognitive Processes*, 22(7), 1114-1139. <https://doi.org/10.1080/01690960701402933>
- Horton, W. S., & Brennan, S. E. (2016). The Role of Metarepresentation in the Production and Resolution of Referring Expressions. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01111>
- Ibarra, A., & Tanenhaus, M. K. (2016). The Flexibility of Conceptual Pacts : Referring Expressions Dynamically Shift to Accommodate New Conceptualizations. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00561>
- Itô, J., & Mester, A. (1995a). Japanese Phonology. Dans J. Goldsmith (Éd.), *The Handbook of Phonological Theory* (pp. 817-838). Oxford: Blackwell.
- Itô, J., & Mester, A. (1995b). The core-periphery structure of the lexicon and constraints on reranking. Dans J. Beckman, S. Urbanczyk, & L. Walsh (Éds), *University of Massachusetts Occasional Papers in Linguistics [UMOP] : Papers in Optimality Theory* (Vol. 18, pp. 181-209). (S.l.): Amherst: GLSA.
- Itô, J., & Mester, A. (1999). The Phonological Lexicon. Dans N. Tsujimura (Éd.), *The Handbook of Japanese Linguistics* (pp. 62-100). (S.l.): Malden, Massachusetts, and Oxford, U.K.: Blackwell Publishers.
- Itô, J., & Mester, A. (2003). *Japanese morphophonemics : Markedness and word structure*. Cambridge, Mass. ; London: MIT Press.
- Itô, J., & Mester, A. (2008). Lexical classes in phonology. Dans S. Miyagawa & M. Saito (Éds), *Handbook of Japanese Linguistics* (pp. 84-106). (S.l.): Oxford: Oxford University Press.
- Kapatsinski, V. (2010). Velar palatalization in Russian and artificial grammar : Constraints on models of morphophonology. *Laboratory Phonology*, 1(2). <https://doi.org/10.1515/labphon.2010.019>

- Kay, G. (1995). English loanwords in Japanese. *World Englishes*, 14(1), 67-76. <https://doi.org/10.1111/j.1467-971X.1995.tb00340.x>
- Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Journal of Laboratory Phonology*, 2, 125-156.
- Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2), 102-110. <https://doi.org/10.1109/4235.918430>
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory : An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31), 10681-10686. <https://doi.org/10.1073/pnas.0707835105>
- Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences of the United States of America*, (104), 5.
- Kirby, S., & Hurford, J. R. (2002). The Emergence of Linguistic Structure. Dans A. Cangelosi & D. Parisi (Éds), *Simulating the Evolution of Language* (pp. 121-147). London: Springer-Verlag.
- Kokuritsu Kokugo Kenkyūjo. (1984). *Goi no kenkyū to kyōiku* (Vol. 1). Tokyo: Kokuritsu Kokugo Kenkyūjo.
- Kubozono, H. (1989). The mora and syllable structure in Japanese : Evidence from speech errors. *Language and Speech*, 32(3), 249-278.
- Kubozono, H. (1999). Mora and Syllable. Dans N. Tsujimura (Éd.), *The Handbook of Japanese Linguistics* (pp. 31-61). Oxford, UK: Blackwell Publishing Ltd. <https://doi.org/10.1002/9781405166225.ch2>
- Labov, W. (1989). The child as linguistic historian. *Language Variation and Change*, 1(1), 85-97. <https://doi.org/10.1017/S0954394500000120>
- Labrune, L. (2005). Autour de la syllabe : Les constituants prosodiques mineurs en phonologie. Dans N. Nguyen, S. Wauquiers, & J. Durand (Éds), *Phonétique et phonologie, approches contemporaines* (pp. 95-116). (S.l.): Hermès.
- Labrune, L. (2006a). *La phonologie du japonais*. Leuven: Peeters.
- Labrune, L. (2006b). Patterns of Phonemic Preferences in Japanese non-headed Binary Compounds : What waa-puro, are-kore and mecha-kucha Have in Common. *Gengo Kenkyu*, 129, 3-41.
- Labrune, L. (2007). Morphophonologie des emprunts abrégés en japonais : Aspects prosodiques et mélodiques. Dans E. Delais-Roussarie & L. Labrune (Éds), *Des sons et des sens : Données et modèles en phonologie et en morphologie*. Paris: Hermès.
- Labrune, L. (2008). Principes d'organisation phonémique des emprunts occidentaux composés abrégés. *Revue d'Etudes Japonaises*, 107-121.

- Labrune, L. (2012a). Questioning the universality of the syllable : Evidence from Japanese. *Phonology*, 29(01), 113-152. <https://doi.org/10.1017/S095267571200005X>
- Labrune, L. (2012b). *The phonology of Japanese*. Oxford ; New York: Oxford University Press.
- Ladefoged, P. (1975). *A course in phonetics*. New York: Harcourt Brace Jovanovich.
- Lamarque, R. (2015). *Morphophonologie des emprunts abrégés japonais dans les jeux vidéo : Évolution du processus d'abréviation*. Master's dissertation, under the supervision of Vittrant, Alice, Aix-Marseille Université, 132p.
- Lamarque, R. (2016). *La variation en phonologie : Le cas des emprunts abrégés composés en japonais*. Master's dissertation, under the supervision of German, James, Aix-Marseille Université, 74p.
- Lamarque, R., & German, J. (2017). Community-and individual-level variation in Japanese compound loanword formation. Dans *The Forty-Third Annual Meeting of the Berkeley Linguistics Society* (pp. 211-234). Berkeley: CA.
- Lenth, R. V. (2016). Least-Squares Means : The R Package lsmeans. *Journal of Statistical Software*, 69(1). <https://doi.org/10.18637/jss.v069.i01>
- Lev-Ari, S., & Peperkamp, S. (2014). An experimental study of the role of social factors in language change : The case of loanword adaptations. *Laboratory Phonology*, 5(3). <https://doi.org/10.1515/lp-2014-0013>
- Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., ... Den, Y. (2014). Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation*, 48(2), 345-371. <https://doi.org/10.1007/s10579-013-9261-0>
- McCawley, J. D. (1968). *The phonological component of a grammar of Japanese*. The Hague: Mouton.
- Meunier, F., & Segui, J. (1999). Morphological Priming Effect : The Role of Surface Frequency. *Brain and Language*, 68(1-2), 54-60. <https://doi.org/10.1006/brln.1999.2098>
- Moder, C. L. (1992). *Productivity and categorization in morphological classes*. State University of New York at Buffalo.
- Moreton, E. (1997). Phonotactic rules in speech perception. Abstract 2aSC4. Communication présentée au 134th Meeting of the Acoustical Society of America, San Diego, CA, Dec. 1-5.
- Moulin-Frier, C., Diard, J., Schwartz, J.-L., & Bessière, P. (2015). COSMO (“Communicating about Objects using Sensory–Motor Operations”) : A Bayesian modeling framework for studying speech communication and the emergence of phonological systems. *Journal of Phonetics*, 53, 5-41. <https://doi.org/10.1016/j.wocn.2015.06.001>
- Nguyen, N., & Delvaux, V. (2015). Role of imitation in the emergence of phonological systems. *Journal of Phonetics*, (53), 46-54.
- Nosofsky, R. M. (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology*, 34(4), 393-418.

- Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable ? Speech Segmentation in Japanese. *Journal of Memory and Language*, (32), 258-278.
- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4), 2382-2393. <https://doi.org/10.1121/1.2178720>
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02). <https://doi.org/10.1017/S0140525X04000056>
- Pickering, M. J., & Garrod, S. (2006). Alignment as the Basis for Successful Communication. *Research on Language and Computation*, 4(2-3), 203-228. <https://doi.org/10.1007/s11168-006-9004-0>
- Pierrehumbert, J. B. (à paraître). 70+ years of probabilistic phonology. Dans B. E. Dresher & H. van der Hulst (Éds), *Oxford Handbook on the the History of Phonology* (p. 47). Oxford: Oxford University Press.
- Pierrehumbert, J. B. (2001a). Exemplar dynamics : Word frequency, lenition and contrast. Dans J. Bybee & P. J. Hopper (Éds), *Typological Studies in Language* (Vol. 45, p. 137). Amsterdam, Netherlands: John Benjamins Publishing Company. <https://doi.org/10.1075/tsl.45.08pie>
- Pierrehumbert, J. B. (2001b). Stochastic phonology. *GLoT*, 5(6), 1-13.
- Pierrehumbert, J. B. (2001c). Why phonological constraints are so coarse-grained. *Language and Cognitive Processes*, 16(5-6), 691-698. <https://doi.org/10.1080/01690960143000218>
- Pike, K. L. (1945). *The intonation of American English*. Ann Arbor, Michigan: University of Michigan Press.
- Prasada, S., & Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes*, 8(1), 1-56. <https://doi.org/10.1080/01690969308406948>
- Prince, A., & Smolensky, P. (1993). *Optimality Theory : Constraint interaction in generative grammar*. New Jersey: Rutgers University Center for Cognitive Science.
- R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing (version 3.5.2) [Logiciel]. Vienna, Austria: R Core Team. Repéré à <https://www.R-project.org/>
- Rácz, P., Beckner, C., Hay, J. B., & Pierrehumbert, J. B. (2020). Morphological convergence as on-line lexical analogy. *Language*, 96(4), 735-770. <https://doi.org/10.1353/lan.2020.0061>
- Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues : A study based on speech resynthesis. *Journal of the Acoustical Society of America*, 105(1), 512-521.
- Roberts, G. (2010). An experimental study of social selection and frequency of interaction in linguistic diversity. *Interaction Studies*, 11(1), 138-159. <https://doi.org/10.1075/is.11.1.06rob>

- Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. Dans *Parallel distributed processing (volume 2) : Psychological and biological models* (pp. 216-271). Cambridge, MA: MIT Press.
- Säily, T. (2011). Variation in morphological productivity in the BNC : Sociolinguistic and methodological considerations. *Corpus Linguistics and Linguistic Theory*, 7(1). <https://doi.org/10.1515/cllt.2011.006>
- Scirra. (2020). Construct 2 (version r279) [Logiciel]. Scirra ltd.
- Shi, L., Griffiths, T. L., Feldman, N. H., & Sanborn, A. N. (2010). Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review*, 17(4), 443-464. <https://doi.org/10.3758/PBR.17.4.443>
- Shimizu Han, M. (1962). The feature of duration in Japanese. *Onsei kenkyuu 音声研究 (research in phonetics)*, (10), 65-80.
- Skousen, R. (1989). *Analogical Modeling of Language*. Dordrecht: Kluwer.
- Skousen, R. (1993). *Analogy and Structure*. Dordrecht: Kluwer.
- Smith, K. (2009). Iterated learning in populations of Bayesian agents. Dans *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 31, p. 7).
- Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, 116(3), 444-449. <https://doi.org/10.1016/j.cognition.2010.06.004>
- Smolensky, P., & Legendre, G. (2006). *The harmonic mind : From neural computation to optimality-theoretic grammar (Cognitive architecture)* (Vol. 1). Cambridge, MA: MIT Press.
- Soliz, J., & Giles, H. (2014). Relational and identity processes in communication : A contextual and meta-analytical review of communication accommodation theory. Dans E. Cohen (Éd.), *Communication yearbook* (Vol. 38, pp. 106-143). Thousand oaks, CA: Sage.
- Sonderegger, M., Bane, M., & Graff, P. (2017). The medium-term dynamics of accents on reality television. *Language*, 93(3), 598-640. <https://doi.org/10.1353/lan.2017.0038>
- Sprenger, J. (2018). The objectivity of Subjective Bayesianism. *European Journal for Philosophy of Science*, 8(3), 539-558. <https://doi.org/10.1007/s13194-018-0200-1>
- Szmrecsanyi, B. (2005). Language users as creatures of habit : A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory*, 1(1). <https://doi.org/10.1515/cllt.2005.1.1.113>
- Szmrecsanyi, B. (2006). *Morphosyntactic persistence in spoken English : A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Berlin: Walter de Gruyter.
- Thornton, A. M. (2012). Overabundance in Italian verb morphology and its interactions with other non-canonical phenomena. Dans H. Otsuka, A. Urdze, J. van der Auwera, & T. Stolz (Éds), *Irregularity in Morphology (and beyond)*. Berlin: Akademie Verlag. <https://doi.org/10.1524/9783050059587.251>

- Vance, T. J. (1987). *An Introduction to Japanese Phonology*. Albany: State University of New York Press. Repéré à <http://www.jstor.org/stable/489314?origin=crossref>
- Vance, T. J. (1991). A New Experimental Study of Japanese Verb Morphology. *Journal of Japanese Linguistics*, 13(1). <https://doi.org/10.1515/jjl-1991-0107>
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth Edition). New York, NY: Springer. Repéré à <http://www.stats.ox.ac.uk/pub/MASS4>
- Veríssimo, J., & Clahsen, H. (2009). Morphological priming by itself : A study of Portuguese conjugations. *Cognition*, 112(1), 187-194. <https://doi.org/10.1016/j.cognition.2009.04.003>
- Von Ahn, L., & Dabbish, L. (2004). Labeling images with a computer game. Dans *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 319-326). Paris.
- Warner, N., & Arai, T. (2001). Japanese Mora-Timing : A Review. *Phonetica*, 58(1-2), 1-25. <https://doi.org/10.1159/000028486>
- Webb, J. (1972). Interview synchrony : An investigation of two speech rate measures in an automated standardized interview. Dans A. W. Siegman & B. Pope (Éds), *Studies in dyadic communication* (pp. 115-133). New York: Pergamon Press.
- Wedel, A. B. (2006). Exemplar models, evolution and language change. *The Linguistic Review*, 23(3). <https://doi.org/10.1515/TLR.2006.010>
- Wilson, C., & Davidson, L. (2013). Bayesian Analysis of Non-native Cluster Production. Dans *Proceedings of NELS* (Vol. 40).

Annexes

Annexe 1 : instructions de l'expérience baseline-CAL

En japonais :

実験に入る前に実験方法についての説明をさせていただきます。

アルファベットとカタカナで、英語からの二つの外来語の単語が表示されます。この二つの単語をもとにした略語の選択肢も四つ表示されますので、あなたにとって一番自然だと思ふ略語を選んでください。

意味をなさない組み合わせもいくつかありますが、それは考慮しなくて結構です。

それでは実験を始めます。

よろしくお願いいたします。

Traduction :

Avant de démarrer l'expérience, voici quelques explications concernant son déroulement :

Deux mots d'emprunt à l'anglais vont vous être présentés en alphabet ainsi qu'en *katakana*.

Quatre propositions d'abréviations basées sur ces deux mots vous seront également présentées.

Veillez choisir l'abréviation que vous pensez être la plus naturelle.

Il se peut que certaines combinaisons n'aient pas beaucoup de sens, mais veuillez ne pas prendre cela en considération.

L'expérience peut commencer.

Merci de votre participation.

Annexe 2 : instructions de l'expérience ESP-CAL

En japonais :

Instructions présentées en début d'expérience :

ゲームを始める前に方法についての説明をさせていただきます。

このゲームは三つのパーツで分けています。

パート 1 (シングルプレイヤー) : アルファベットとカタカナで、英語からの二つの外来語の単語が表示されます。この二つの単語をもとにした略語の選択肢も四つ表示されますので、あなたにとって一番自然だと思える略語を選んでください。意味をなさない組み合わせもいくつかありますが、それは考慮しなくて結構です。

パート 2 (マルチプレイヤー) : パート 1 の同じように新しい略語を作るんですが、あなたはオンラインで他のプレイヤーに接続されています。目的は二人のプレイヤーは談話しないで同じ答えを選ぶのです。接続上で、誰がプレイヤー1やプレイヤー2が自動的に決めています。

プレイヤー1は最初に選びます。

プレイヤー2はプレイヤー1が答えた後で選びます。

パート 3 (シングルプレイヤー) : 他のプレイヤーがディスコネクトされています。あなたはもう一回パート 1 のように略語を作るんです。

それでは「パート 1 (シングルプレイヤー)」を始めます。

よろしく願いいたします。"

Instructions présentées entre la phase 1 et la phase 2 :

「パート 1 (シングルプレイヤー)」が終わりました。これから、「パート 2 (マルチプレイヤー)」がはじめるのです。

パート1と同じように新しい略語を作りますが、あなたはオンラインで他のプレイヤーに接続されています。

このパートでは、二人のプレイヤーが相談せずに同じ答えを選択することを目的としています。

コネクション上で、誰がプレイヤー1で誰がプレイヤー2かが自動的に決められています。

プレイヤー1は最初に選択します。

プレイヤー2はプレイヤー1が答えた後で選択します。

プレイヤー2がプレイヤー1と同じ答えを選択すれば、1ポイントもらうことができます。

ハイスコア狙いましょう。

Instructions présentées entre la phase 2 et la phase 3 :

「パート2(マルチプレイヤー)」が終わりました。あなたのスコアはXポイントです。

これから、「パート3(シングルプレイヤー)」がはじめるのです。

他のプレイヤーがディスコネクトされています。あなたはもう一回パート1のように略語を作るんです。

Traduction :

Instructions présentées en début d'expérience :

Avant de démarrer le jeu, voici quelques explications sur son fonctionnement :

Ce jeu est divisé en trois parties.

Partie 1 (un joueur) : deux mots d'emprunt à l'anglais vont vous être présentés en alphabet ainsi qu'en *katakana*. Quatre propositions d'abréviations basées sur ces deux mots vous seront également présentées. Veuillez choisir l'abréviation que vous pensez être la plus naturelle.

Il se peut que certaines combinaisons n'aient pas beaucoup de sens, mais veuillez ne pas prendre cela en considération.

Partie 2 (multijoueur) : vous allez créer de nouvelles abréviations de la même manière que lors de la partie 1, cependant, vous serez connectés en ligne à un autre joueur. Le but est que les deux joueurs choisissent les mêmes réponses sans se concerter. Au moment de la connexion, les rôles de premier et deuxième joueur sont attribués aléatoirement.

Le premier joueur choisit en premier.

Le deuxième joueur choisit après que le premier joueur a répondu.

Partie 3 (un joueur) : vous serez déconnecté de l'autre joueur. Une fois de plus, vous créerez de nouvelles abréviations de la même manière que dans la partie 1.

La « Partie 1 (un joueur) » peut commencer.

Merci de votre participation.

Instructions présentées entre la phase 1 et la phase 2 :

La « Partie 1 (un joueur) » est terminée. Maintenant, la « Partie 2 (multijoueur) » va commencer.

Vous allez créer de nouvelles abréviations de la même manière que lors de la partie 1, cependant, vous serez connectés en ligne à un autre joueur. Dans cette partie, le but des deux joueurs est de choisir les mêmes réponses sans se concerter.

Au moment de la connexion, les rôles de premier et deuxième joueur sont attribués aléatoirement.

Le premier joueur choisit en premier.

Le deuxième joueur choisit après que le premier joueur a répondu.

Si le deuxième joueur choisit la même réponse que le premier joueur, vous obtenez 1 point.

Visez un high score !

Instructions présentées entre la phase 2 et la phase 3 :

La « Partie 2 (multijoueur) » est terminée. Votre score est : X points.

Maintenant, la « Partie 3 (un joueur) » va commencer.

Vous serez déconnecté de l'autre joueur. Une fois de plus, vous créez de nouvelles abréviations de la même manière que dans la partie 1.

Annexe 3 : instructions de l'expérience baseline-flexions verbales

En japonais :

調査に入る前に調査方法についての説明をさせていただきます。

これは日本語の動詞についての調査です。二人の友人（萌、翔太）の短い会話が提示されます。一人が最初に、ある動詞を含んだ文章を話します。次にそれに対して二人目が同じ動詞を違った形にして答えます。しかし、二人目の文章では動詞の位置に「_____」が表示されます。あなたは二人目の答える動詞のもっとも自然な形を二つの選択肢から選んで、「_____」に記入して下さい。会話で使われる動詞は見慣れないものですが、意味には注意を払わず、あなたにとってより自然と思われる形を選んで下さい。

それでは調査を始めます。

よろしく願いいたします。

Traduction :

Avant de démarrer l'expérience, voici quelques explications concernant son déroulement :

Cette expérience porte sur les verbes japonais. De courts dialogues entre deux amis (Moe et Shōta) vous seront présentés. Le premier personnage dit une phrase contenant un verbe. Ensuite, le second personnage répond en utilisant le même verbe sous une autre forme. Cependant, il y aura un blanc « _____ » à la place du verbe dans la phrase du second personnage. Vous devrez remplir ce blanc « _____ » en choisissant parmi deux formes proposées celle qui vous semble la plus naturelle pour le verbe dans la réponse du deuxième personnage. Les verbes employés dans les dialogues sont des verbes que vous n'avez pas l'habitude de voir. Cependant, ne prêtez pas trop attention au sens et choisissez la forme qui vous semble la plus naturelle.

L'expérience peut commencer.

Merci de votre participation.

Annexe 4 : instructions de l'expérience ESP-CAL

En japonais :

Instructions présentées en début d'expérience :

ゲームを始める前に方法についての説明をさせていただきます。

このゲームは三つのパートに分かれています。

パート 1 (シングルプレイヤー) : 二人の友人 (萌、翔太) の短い会話が提示されます。一人が最初に、ある動詞を含んだ文章を話します。次にそれに対して二人目が 同じ動詞を違った形にして答えます。しかし、二人目の文章では動詞の位置に「_____」が表示されます。あなたは二人目が 答える動詞としてもっとも自然な形を二つの選択肢から選んで、「_____」に記入して下さい。会話で使われる動詞は見慣れないものですが、意味には注意を払わず、あなたにとってより自然と思われる形を選んで下さい。

パート 2 (マルチプレイヤー) : パート 1 と同じように動詞のもっとも自然な形を選んでください。あなたはオンラインで他のプレイヤーに接続されています。

このパートの目標は二人のプレイヤーが相談せずに同じ答えを選ぶことです。 コネクション上で、誰がプレイヤー1 で誰がプレイヤー2 になるかは自動的に割り当てられます。

プレイヤー1 は最初に選びます。プレイヤー2 はプレイヤー1 が答えた後で選びます。

パート 3 (シングルプレイヤー) : 他のプレイヤーはディスコネクトされています。もう一回パート 1 のように動詞のもっとも自然な形を選んでください。

それでは「パート 1 (シングルプレイヤー)」を始めます。

よろしく願いいたします。

Instructions présentées entre la phase 1 et la phase 2 :

「パート 1 (シングルプレイヤー)」が終わりました。これから、「パート 2 (マルチプレイヤー)」がはじめるのです。

パート 1 と同じように動詞のもっとも自然な形を選んでください。あなたはオンラインで他のプレイヤーに接続されています。

このパートの目標は二人のプレイヤーが相談せずに同じ答えを選ぶことです。

コネクション上で、誰がプレイヤー 1 で誰がプレイヤー 2 になるかは自動的に割り当てられます。

プレイヤー 1 は最初に選択します。

プレイヤー 2 はプレイヤー 1 が答えた後で選択します。

プレイヤー 2 がプレイヤー 1 と同じ答えを選択すれば、1 ポイントもらうことができます。

ハイスコアを狙いましょう。

Instructions présentées entre la phase 2 et la phase 3 :

「パート 2 (マルチプレイヤー)」が終わりました。あなたのスコアは X ポイントです。

これから、「パート 3 (シングルプレイヤー)」がはじめるのです。

他のプレイヤーはディスコネクトされています。

もう一回パート 1 のように動詞のもっとも自然な形を選んでください。

Traduction :

Instructions présentées en début d'expérience :

Avant de démarrer le jeu, voici quelques explications sur son fonctionnement :

Ce jeu est divisé en trois parties.

Partie 1 (un joueur) : de courts dialogues entre deux amis (Moe et Shōta) vous seront présentés. Le premier personnage dit une phrase contenant un verbe. Ensuite, le second personnage répond en utilisant le même verbe sous une autre forme. Cependant, il y aura un blanc « _____ » à la place du verbe dans la phrase du second personnage. Vous devrez remplir ce blanc « _____ » en choisissant parmi deux formes proposées celle qui vous semble la plus naturelle pour le verbe dans la réponse du deuxième personnage. Les verbes employés dans les dialogues sont des verbes que vous n'avez pas l'habitude de voir. Cependant, ne prêtez pas trop attention au sens et choisissez la forme qui vous semble la plus naturelle.

Partie 2 (multijoueur) : vous allez choisir la forme la plus naturelle pour des verbes de la même manière que lors de la partie 1, cependant, vous serez connectés en ligne à un autre joueur. Le but de cette partie est que les deux joueurs choisissent les mêmes réponses sans se concerter. Au moment de la connexion, les rôles de premier et deuxième joueur sont attribués aléatoirement.

Le premier joueur choisit en premier.

Le deuxième joueur choisit après que le premier joueur a répondu.

Partie 3 (un joueur) : vous serez déconnecté de l'autre joueur. Une fois de plus, vous choisirez la forme la plus naturelle pour des verbes de la même manière que dans la partie 1.

La « Partie 1 (un joueur) » peut commencer.

Merci de votre participation.

Instructions présentées entre la phase 1 et la phase 2 :

La « Partie 1 (un joueur) » est terminée. Maintenant, la « Partie 2 (multijoueur) » va commencer.

Veillez choisir la forme la plus naturelle pour des verbes de la même manière que lors de la partie 1. Vous allez être connectés en ligne à un autre joueur. Dans cette partie, le but des deux joueurs est de choisir les mêmes réponses sans se concerter.

Au moment de la connexion, les rôles de premier et deuxième joueur sont attribués aléatoirement.

Le premier joueur choisit en premier.

Le deuxième joueur choisit après que le premier joueur a répondu.

Si le deuxième joueur choisit la même réponse que le premier joueur, vous obtenez 1 point.

Visez un high score !

Instructions présentées entre la phase 2 et la phase 3 :

La « Partie 2 (multijoueur) » est terminée. Votre score est : X points.

Maintenant, la « Partie 3 (un joueur) » va commencer.

Vous serez déconnecté de l'autre joueur. Une fois de plus, veuillez choisir la forme la plus naturelle pour des verbes de la même manière que dans la partie 1.

Annexe 5 : liste des premiers constituants utilisés pour les items des expériences sur les CAL

Description des colonnes :

- Numéro : numéro arbitrairement donné aux constituants pour les identifier.
- Set : le set d'appartenance du constituant. Pour former les items, les C1 sont appairés aléatoirement avec un C2 possédant le même numéro de set.
- Anglais : mot emprunté dans la langue d'origine.
- Katakana : transcription du mot en *katakana*.
- Alphabet : translittération selon la méthode de romanisation Hepburn (avec voyelles longues dédoublées).
- Abréviation : forme abrégée (en *katakana*) du constituant telle qu'elle apparaissait dans les différentes proposition de réponses.

Numéro	Set	Anglais	Katakana	Alphabet	Abréviation
1	1	drive	ドライブ	doraibu	ドラ
2	2	temptation	テンプテーション	teNputeeshoN	テン
3	3	craft	クラフト	kurafuto	クラ
4	4	challenge	チャレンジ	chareNji	チャレ
5	5	rival	ライバル	raibaru	ライ
6	6	licence	ライセンス	raiseNsu	ライ
7	7	culture	カルチャー	karuchaa	カル
8	8	sponsor	スポンサー	supoNsaa	スポ
9	9	seminar	セミナー	seminaa	セミ
10	10	text	テキスト	tekisuto	テキ
11	11	delicacy	デリカシー	derikashii	デリ
12	12	ranking	ランキング	raNkiNgu	ラン
13	13	original	オリジナル	orijinaru	オリ

14	14	country	カントリー	kaNtorii	カン
15	15	captain	キャプテン	kyaputeN	キャプ
16	16	design	デザイン	dezaiN	デザ
17	17	spectre	スペクトル	supekutoru	スペ
18	18	digital	デジタル	dejitaru	デジ
19	19	concourse	コンコース	koNkoosu	コン
20	20	contrast	コントラスト	koNtorasuto	コン
21	21	document	ドキュメント	dokyumeNto	ドキュ
22	22	style	スタイル	sutairu	スタ
23	23	system	システム	shisutemu	シス
24	24	tunnel	トンネル	toNneru	トン
25	13	cosmos	コスモス	kosumosu	コス
26	14	deflator	デフレーター	defureetaa	デフ
27	15	diet	ダイエット	daietto	ダイ
28	16	colorful	カラフル	karafuru	カラ
29	17	jumper	ジャンパー	jaNpaa	ジャン
30	18	template	テンプレート	teNpureeto	テン
31	19	slider	スライダー	suraidaa	スラ
32	20	straw	ストロー	sutoroo	スト
33	21	engine	エンジン	eNjiN	エン
34	22	clinic	クリニック	kurinikku	クリ
35	23	trunk	トランク	toraNku	トラ
36	24	cemetary	セメタリー	semetarii	セメ
37	13	spiritual	スピリチュアル	supirichuaru	スピ
38	14	center	センター	sentaa	セン
39	15	single	シングル	shinguru	シン
40	16	gasoline	ガソリン	gasoriN	ガソ

41	17	aluminium	アルミニウム	aruminiumu	アル
42	18	sponge	スポンジ	supoNji	スポ
43	19	spray	スプレー	supuree	スプ
44	20	sample	サンプル	saNpuru	サン
45	21	trouble	トラブル	toraburu	トラ
46	22	catalogue	カタログ	katarogu	カタ
47	23	episode	エピソード	episoodo	エピ
48	24	credit	クレジット	kurejitto	クレ
49	25	command	コマンド	komaNdo	コマ
50	26	steroid	ステロイド	suteroido	ステ
51	27	sexy	セクシー	sekushii	セク
52	28	taxi	タクシー	takushii	タク
53	29	stress	ストレス	sutoresu	スト
54	30	advice	アドバイス	adobaisu	アド
55	31	escort	エスコート	esukooto	エス
56	32	title	タイトル	taitoru	タイ
57	33	talent	タレント	tareNto	タレ
58	34	stretch	ストレッチ	sutorecchi	スト
59	35	dynamite	ダイナマイト	dainamaito	ダイ
60	36	temple	テンプル	teNpuru	テン
61	25	cement	セメント	semeNto	セメ
62	26	standard	スタンダード	sutaNdaado	スタ
63	27	regular	レギュラー	regyuraa	レギュ
64	28	senator	セネター	senetaa	セネ
65	29	specialist	スペシャリスト	supesharisuto	スペ
66	30	gradation	グラデーション	guradeeshoN	グラ
67	31	clarinet	クラリネット	kurarinetto	クラ

68	32	trust	トラスト	torasuto	トラ
69	33	syndrome	シンドローム	shiNdoroomu	シン
70	34	adult	アダルト	adaruto	アダ
71	35	orange	オレンジ	oreNji	オレ
72	36	synchro	シンクロ	shiNkuro	シン
73	37	comical	コミカル	komikaru	コミ
74	38	client	クライアント	kuraiaNto	クラ
75	39	salary	サラリー	sararii	サラ
76	40	company	カンパニー	kaNpanii	カン
77	41	negative	ネガティブ	negatibu	ネガ
78	42	security	セキュリティー	sekyuritii	セキュ
79	43	spring	スプリング	supuriNgu	スプ
80	44	ground	グラウンド	guruNdo	グラ
81	45	sensor	センサー	seNsaa	セン
82	46	stage	ステージ	suteeji	ステ
83	47	action	アクション	akushoN	アク
84	48	calorie	カロリー	karorii	カロ
85	37	domain	ドメイン	domeiN	ドメ
86	38	dilemma	ジレンマ	jireNma	ジレ
87	39	state	ステート	suteeto	ステ
88	40	trend	トレンド	toreNdo	トレ
89	41	amateur	アマチュア	amachua	アマ
90	42	android	アンドロイド	aNdoroido	アン
91	43	smooth	スムーズ	sumuuzu	スム
92	44	chemical	ケミカル	kemikaru	ケミ
93	45	studio	スタジオ	sutajio	スタ
94	46	graphic	グラフィック	gurafikku	グラ

95	47	library	ライブラリー	raiburarii	ライ
96	48	subtitle	サブタイトル	sabutaitoru	サブ
97	37	stoic	ストイック	sutoikku	スト
98	38	reception	レセプション	resepushoN	レセ
99	39	adapter	アダプター	adaputa	アダ
100	40	stand	スタンド	sutaNdo	スタ
101	41	screen	スクリーン	sukuriiN	スク
102	42	establishment	エスタブリッシュメント	esutaburisshumeNto	エス
103	43	section	セクション	sekushoN	セク
104	44	timing	タイミング	taimiNgu	タイ
105	45	cycle	サイクル	saikuru	サイ
106	46	economic	エコノミック	ekonomikku	エコ
107	47	desktop	デスクトップ	desukutoppu	デス
108	48	connector	コネクター	konekutaa	コネ
109	37	analog	アナログ	anarogu	アナ
110	38	symbol	シンボル	shiNboru	シン
111	39	calendar	カレンダー	kareNdaa	カレ
112	40	experience	エクスペリアンス	ekusuperiaNsu	エク
113	41	drink	ドリンク	doriNku	ドリ
114	42	report	レポート	repooto	レポ
115	43	gypsy	ジプシー	jipushii	ジプ
116	44	cream	クリーム	kuriimu	クリ
117	45	sandal	サンダル	saNdaru	サン
118	46	synapse	シナプス	shinapusu	シナ
119	47	charisma	カリスマ	karisuma	カリ
120	48	silhouette	シルエット	shiruetto	シル


Annexe 6 : liste n°1 de deuxièmes constituants utilisés pour les items de l'expérience baseline-CAL


Description des colonnes :


- Numéro : numéro arbitrairement donné aux constituants pour les identifier.
- Set : le set d'appartenance du constituant. Pour former les items, les C2 sont appairés aléatoirement avec un C1 possédant le même numéro de set.
- /Q/ : la deuxième more du constituant est-elle /Q/ ? (O = oui ; X = non).
- /R/ : la deuxième more du constituant est-elle /R/ ? (X = non ; - H = oui, et la voyelle allongée est non haute ; + H = oui, et la voyelle allongée est haute).
- Bilab. : la consonne initiale est-elle une consonne bilabiale ? Le phonème /h/ est considéré comme une consonne bilabiale pour des raisons historiques et phonologiques le liant au phonème /p/ en japonais archaïque (Labruno, 2006b). (O = oui ; X = non)
- Long. : longueur du constituant en more (3 μ = trois mores ; >3 μ = plus de trois mores).
- Anglais : mot emprunté dans la langue d'origine.
- Katakana : transcription du mot en *katakana*
- Hepburn : translittération selon la méthode de romanisation Hepburn (avec voyelles longues dédoublées).
- Maintien : forme abrégée selon le pattern de Maintien du constituant, telle qu'elle apparaissait dans la proposition de réponse correspondante.
- Suivi : forme abrégée selon le pattern de Suivi du constituant, telle qu'elle apparaissait dans la proposition de réponse correspondante.
- Suppr. : forme abrégée selon le pattern de Suppression du constituant, telle qu'elle apparaissait dans la proposition de réponse correspondante.


- Remp. : forme abrégée selon le pattern de Remplacement du constituant, telle qu'elle apparaissait dans la proposition de réponse correspondante.

Note : les items possédant les mêmes caractéristiques phonologiques (dans les limites décrites en section 4.1.1) sont regroupés par couleur. De plus, les types d'items formant des sets sont regroupés par dégradés de couleur.

 = C2 *sans conflit* dans des sets d'un seul item

 = C2 *sans conflit* dans des sets de trois items

 = C2 /Q/

 = C2 /R/

Numéro	Set	/Q/	/R/	Bilab.	Long.	Anglais	Katakana	Hepburn	Maintien	Suivi	Suppr.	Remp.
1	1	X	X	X	>3μ	cancel	キャンセル	kyaNseru	キャン	キャンセ	キャ	キャセ
2	2	X	X	X	>3μ	crown	クラウン	kurauN	クラ	クラウ	ク	クウ
3	3	X	X	X	>3μ	royal	ロイヤル	roiyaru	ロイ	ロイヤ	ロ	ロヤ
4	4	X	X	X	>3μ	legend	レジェンド	regeNdo	レジェ	レジェン	レ	レン
5	5	X	X	X	>3μ	slime	スライム	suraimu	スラ	スライ	ス	スイ
6	6	X	X	X	>3μ	dragon	ドラゴン	doragoN	ドラ	ドラゴ	ド	ドゴ
7	7	X	X	X	>3μ	quest	クエスト	kuesuto	クエ	クエス	ク	クス
8	8	X	X	X	>3μ	gunner	ガンナー	gaNnaa	ガン	ガンナ	ガ	ガナ
9	9	X	X	X	>3μ	crystal	クリスタル	kurisutaru	クリ	クリス	ク	クス
10	10	X	X	X	>3μ	strike	ストライク	sutoraiku	スト	ストラ	ス	スラ
11	11	X	X	X	>3μ	tiger	タイガー	taigaa	タイ	タイガ	タ	タガ
12	12	X	X	X	>3μ	stealth	ステルス	suterusu	ステ	ステル	ス	スル
13	13	X	X	X	>3μ	slide	スライド	suraido	スラ	スライ	ス	スイ

14	14	X	X	X	>3μ	relief	リリーフ	ririifu	リリ	リリー	リ	リー
15	15	X	X	X	>3μ	series	シリーズ	shiriizu	シリ	シリー	シ	シー
16	16	X	X	X	>3μ	trans	トランス	toraNsu	トラ	トラン	ト	トン
17	17	X	X	X	>3μ	slice	スライス	suraisu	スラ	スライ	ス	スイ
18	18	X	X	X	>3μ	drastic	ドラスチック	dorasuchikku	ドラ	ドラス	ド	ドス
19	19	X	X	X	>3μ	wind	ウインド	uiNdo	ウイ	ウイン	ウ	ウン
20	20	X	X	X	>3μ	single	シングル	shiNguru	シン	シング	シ	シグ
21	21	X	X	X	>3μ	sound	サウンド	sauNdo	サウ	サウン	サ	サン
22	22	X	X	X	>3μ	rounder	ラウンダー	rauNdaa	ラウ	ラウン	ラ	ラン
23	23	X	X	X	>3μ	cantor	カントル	kaNtoru	カン	カント	カ	カト
24	24	X	X	X	>3μ	saloon	サルーン	saruuN	サル	サルー	サ	サー
25	13	X	X	X	3μ	dry	ドライ	dorai	ドラ	ドライ	ド	ドイ
26	14	X	X	X	3μ	rally	ラリー	rarii	ラリ	ラリー	ラ	ラー
27	15	X	X	X	3μ	tree	ツリー	tsurii	ツリ	ツリー	ツ	ツー
28	16	X	X	X	3μ	clan	クラン	kuraN	クラ	クラン	ク	クン
29	17	X	X	X	3μ	cry	クライ	kurai	クラ	クライ	ク	クイ
30	18	X	X	X	3μ	class	クラス	kurasu	クラ	クラス	ク	クス
31	19	X	X	X	3μ	coin	コイン	koiN	コイ	コイン	コ	コン
32	20	X	X	X	3μ	ring	リング	riNgu	リン	リング	リ	リグ
33	21	X	X	X	3μ	town	タウン	tauN	タウ	タウン	タ	タン
34	22	X	X	X	3μ	down	ダウン	dauN	ダウ	ダウン	ダ	ダン
35	23	X	X	X	3μ	tent	テント	teNto	テン	テント	テ	テト
36	24	X	X	X	3μ	crew	クルー	kuruu	クル	クルー	ク	クー
37	13	X	X	O	>3μ	bride	ブライド	buraido	ブラ	ブライ	ブ	ブイ
38	14	X	X	O	>3μ	belief	ビリーフ	biriifu	ビリ	ビリー	ビ	ビー
39	15	X	X	O	>3μ	freeze	フリーズ	furiizu	フリ	フリー	フ	フー
40	16	X	X	O	>3μ	balance	バランス	baraNsu	バラ	バラン	バ	バン

41	17	X	X	O	>3μ	price	プライス	puraisu	プラ	プライ	プ	パイ
42	18	X	X	O	>3μ	plastic	プラスチック	purasuchikku	プラ	プラス	プ	プス
43	19	X	X	O	>3μ	mind	マインド	maiNdo	マイ	メイン	マ	マン
44	20	X	X	O	>3μ	bangle	バングル	baNguru	バン	バング	バ	バグ
45	21	X	X	O	>3μ	bound	バウンド	bauNdo	バウ	バウン	バ	バン
46	22	X	X	O	>3μ	founder	ファウンダー	fauNdaa	ファウ	ファウン	ファ	ファン
47	23	X	X	O	>3μ	mantle	マントル	maNtoru	マン	マント	マ	マト
48	24	X	X	O	>3μ	balloon	バルーン	baruuN	バル	バルー	バ	バー
49	25	O	X	X	>3μ	cup	カップ	kappu	カッ	カップ	カ	カプ
50	26	O	X	X	>3μ	rich	リッチ	ricchi	リッ	リッチ	リ	リチ
51	27	O	X	O	>3μ	hack	ハック	hakku	ハッ	ハック	ハ	ハク
52	28	O	X	O	>3μ	mash	マッシュ	masshu	マッ	マッシュ	マ	マシュ
53	29	O	X	O	>3μ	pick	ピック	pikku	ピッ	ピック	ピ	ピク
54	30	O	X	X	>3μ	dog	ドッグ	doggu	ドッ	ドッグ	ド	ドグ
55	31	O	X	X	>3μ	sack	サック	sakku	サッ	サック	サ	サク
56	32	O	X	O	>3μ	bat	バット	batto	バッ	バット	バ	バト
57	33	O	X	X	>3μ	neck	ネック	nekku	ネッ	ネック	ネ	ネク
58	34	O	X	X	>3μ	net	ネット	netto	ネッ	ネット	ネ	ネト
59	35	O	X	O	>3μ	book	ブック	bukku	ブッ	ブック	ブ	ブク
60	36	O	X	O	>3μ	hedge	ヘッジ	hejji	ヘッ	ヘッジ	ヘ	ヘジ
61	25	O	X	X	3μ	couple	カップル	kappuru	カッ	カップ	カ	カプ
62	26	O	X	X	3μ	richness	リッチネス	ricchinesu	リッ	リッチ	リ	リチ
63	27	O	X	O	3μ	hackle	ハックル	hakuru	ハッ	ハック	ハ	ハク
64	28	O	X	O	3μ	mushroom	マッシュルーム	masshuruumu	マッ	マッシュ	マ	マシュ
65	29	O	X	O	3μ	pickle	ピクル	pikkuru	ピッ	ピック	ピ	ピク
66	30	O	X	X	3μ	toggle	トグル	togguru	トッ	トッグ	ト	トグ
67	31	O	X	X	3μ	sax	サククス	sakkusu	サッ	サック	サ	サク

68	32	O	X	O	3μ	batter	バッテリー	battaa	バッ	バッタ	バ	バタ
69	33	O	X	X	3μ	necklace	ネックレス	nekkuresu	ネッ	ネック	ネ	ネク
70	34	O	X	X	3μ	network	ネットワーク	nettowaaku	ネッ	ネット	ネ	ネト
71	35	O	X	O	3μ	booklet	ブックレット	bukkuretto	ブッ	ブック	ブ	ブク
72	36	O	X	O	3μ	hedgehog	ヘッジホッグ	hejjihoggu	ヘッ	ヘッジ	へ	ヘジ
73	37	X	-H	X	>3μ	taper	テーパー	teepaa	テー	テーパ	テ	テパ
74	38	X	-H	X	3μ	kart	カート	kaato	カー	カート	カ	カト
75	39	X	-H	X	>3μ	server	サーバー	saabaa	サー	サーバ	サ	サバ
76	40	X	-H	X	3μ	road	ロード	roodo	ロー	ロード	ロ	ロド
77	41	X	-H	X	>3μ	gamer	ゲーマー	geemaa	ゲー	ゲーマ	ゲ	ゲマ
78	42	X	-H	X	>3μ	rating	レーティング	reetiNgu	レー	レーティ	レ	レティ
79	43	X	-H	X	3μ	girl	ガール	gaaru	ガー	ガール	ガ	ガル
80	44	X	-H	X	3μ	sauce	ソース	soosu	ソー	ソース	ソ	ソス
81	45	X	-H	X	>3μ	girdle	ガードル	gaadoru	ガー	ガード	ガ	ガド
82	46	X	-H	X	3μ	name	ネーム	neemu	ネー	ネーム	ネ	ネム
83	47	X	-H	X	>3μ	tones	トーンズ	tooNzu	トー	トーン	ト	トン
84	48	X	-H	X	3μ	zone	ゾーン	zooN	ゾー	ゾーン	ゾ	ゾン
85	37	X	+H	X	>3μ	super	スーパー	suupaa	スー	スーパ	ス	スパ
86	38	X	+H	X	3μ	seat	シート	shiito	シー	シート	シ	シト
87	39	X	+H	X	>3μ	louver	ルーバー	ruubaa	ルー	ルーバ	ル	ルバ
88	40	X	+H	X	3μ	seed	シード	shiido	シー	シード	シ	シド
89	41	X	+H	X	>3μ	zoomer	ズーマー	zuumaa	ズー	ズーマ	ズ	ズマ
90	42	X	+H	X	>3μ	routing	ルーティング	ruutiNgu	ルー	ルーティ	ル	ルティ
91	43	X	+H	X	3μ	tool	ツール	tsuuru	ツー	ツール	ツ	ツル
92	44	X	+H	X	3μ	lease	リース	riisu	リー	リース	リ	リス
93	45	X	+H	X	>3μ	noodle	ヌードル	nuudoru	ヌー	ヌード	ヌ	ヌド
94	46	X	+H	X	3μ	team	チーム	chiimu	チー	チーム	チ	チム

95	47	X	+ H	X	>3μ	jeans	ジーンズ	jiiNzu	ジー	ジーン	ジ	ジン
96	48	X	+ H	X	3μ	scene	シーン	shiiN	シー	シーン	シ	シン
97	37	X	- H	O	>3μ	paper	ペーパー	peepaa	ペー	ペーパ	ペ	ペパ
98	38	X	- H	O	3μ	part	パート	paato	パー	パート	パ	パト
99	39	X	- H	O	>3μ	harbor	ハーバー	haabaa	ハー	ハーバ	ハ	ハバ
100	40	X	- H	O	3μ	board	ボード	boodo	ボー	ボード	ボ	ボド
101	41	X	- H	O	>3μ	homer	ホーム	hoomaa	ホー	ホーム	ホ	ホマ
102	42	X	- H	O	>3μ	porting	ポータィング	pootiNgu	ポー	ポーティ	ポ	ポティ
103	43	X	- H	O	3μ	ball	ボール	booru	ボー	ボール	ボ	ボル
104	44	X	- H	O	3μ	base	ベース	peesu	ベー	ベース	ベ	ベス
105	45	X	- H	O	>3μ	hurdle	ハードル	haadoru	ハー	ハード	ハ	ハド
106	46	X	- H	O	3μ	home	ホーム	hoomu	ホー	ホーム	ホ	ホム
107	47	X	- H	O	>3μ	bones	ボーンズ	booNzu	ボー	ボーン	ボ	ボン
108	48	X	- H	O	3μ	burn	バーン	baaN	バー	バーン	バ	バン
109	37	X	+ H	O	>3μ	Hooper	フーパー	huupaa	フー	フーパ	フ	フパ
110	38	X	+ H	O	3μ	beat	ビート	biito	ビー	ビート	ビ	ビト
111	39	X	+ H	O	>3μ	beaver	ビーバー	biibaa	ビー	ビーバ	ビ	ビバ
112	40	X	+ H	O	3μ	mood	ムード	muudo	ムー	ムード	ム	ムド
113	41	X	+ H	O	>3μ	boomer	ブーマー	boomaa	ブー	ブーマ	ブ	ブマ
114	42	X	+ H	O	>3μ	meeting	ミーティング	miitiNgu	ミー	ミーティ	ミ	ミティ
115	43	X	+ H	O	3μ	beer	ビール	biiru	ビー	ビール	ビ	ビル
116	44	X	+ H	O	3μ	booth	ブース	buusu	ブー	ブース	ブ	ブス
117	45	X	+ H	O	>3μ	poodle	プードル	puudoru	プー	プード	プ	プド
118	46	X	+ H	O	3μ	beam	ビーム	biimu	ビー	ビーム	ビ	ビム
119	47	X	+ H	O	>3μ	beans	ビーンズ	biiNzu	ビー	ビーン	ビ	ビン
120	48	X	+ H	O	3μ	mean	ミーン	miiN	ミー	ミーン	ミ	ミン


Annexe 7 : liste n°2 de deuxièmes constituants utilisés pour les items de l'expérience baseline-CAL


Description des colonnes :


- Numéro : numéro arbitrairement donné aux constituants pour les identifier.
- Set : le set d'appartenance du constituant. Pour former les items, les C2 sont appairés aléatoirement avec un C1 possédant le même numéro de set.
- /Q/ : la deuxième more du constituant est-elle /Q/ ? (O = oui ; X = non).
- /R/ : la deuxième more du constituant est-elle /R/ ? (X = non ; - H = oui, et la voyelle allongée est non haute ; + H = oui, et la voyelle allongée est haute).
- Bilab. : la consonne initiale est-elle une consonne bilabiale ? Le phonème /h/ est considéré comme une consonne bilabiale pour des raisons historiques et phonologiques le liant au phonème /p/ en japonais archaïque (Labrune, 2006b). (O = oui ; X = non)
- Long. : longueur du constituant en more (3 μ = trois mores ; >3 μ = plus de trois mores).
- Anglais : mot emprunté dans la langue d'origine.
- Katakana : transcription du mot en *katakana*
- Hepburn : translittération selon la méthode de romanisation Hepburn (avec voyelles longues dédoublées).
- Maintien : forme abrégée selon le pattern de Maintien du constituant, telle qu'elle apparaissait dans la proposition de réponse correspondante.
- Suivi : forme abrégée selon le pattern de Suivi du constituant, telle qu'elle apparaissait dans la proposition de réponse correspondante.
- Suppr. : forme abrégée selon le pattern de Suppression du constituant, telle qu'elle apparaissait dans la proposition de réponse correspondante.


- Remp. : forme abrégée selon le pattern de Remplacement du constituant, telle qu'elle apparaissait dans la proposition de réponse correspondante.

Note : les items possédant les mêmes caractéristiques phonologiques (dans les limites décrites en section 4.1.1) sont regroupés par couleur. De plus, les types d'items formant des sets sont regroupés par dégradés de couleur.

 = C2 *sans conflit* dans des sets d'un seul item

 = C2 *sans conflit* dans des sets de trois items

 = C2 /Q/

 = C2 /R/

Numéro	Set	/Q/	/R/	Bilab.	Long.	Anglais	Katakana	Hepburn	Maintien	Suivi	Suppr.	Remp.
1	1	X	X	X	>3μ	runner	ランナー	raNnaa	ラン	ランナ	ラ	ラナ
2	2	X	X	X	>3μ	classic	クラシック	kurashikku	クラ	クラシ	ク	クシ
3	3	X	X	X	>3μ	custom	カスタム	kasutamu	カス	カスタ	カ	カタ
4	4	X	X	X	>3μ	contact	コンタクト	koNtakuto	コン	コンタ	コ	コタ
5	5	X	X	X	>3μ	correction	コレクション	korekushoN	コレ	コレク	コ	コク
6	6	X	X	X	>3μ	stream	ストリーム	sutoriimu	スト	ストリ	ス	スリ
7	7	X	X	X	>3μ	drift	ドリフト	dorifuto	ドリ	ドリフ	ド	ドフ
8	8	X	X	X	>3μ	landing	ランディング	raNdiNgu	ラン	ランディ	ラ	ラディ
9	9	X	X	X	>3μ	snipe	スナイプ	sunaipu	スナ	スナイ	ス	スイ
10	10	X	X	X	>3μ	technique	テクニック	tekunikku	テク	テクニ	テ	テニ
11	11	X	X	X	>3μ	thunder	サンダー	saNdaa	サン	サンダ	サ	サダ

12	12	X	X	X	>3μ	special	スペシャル	supesharu	スペ	スペシャ	ス	スシャ
13	13	X	X	X	>3μ	lizard	リザード	rizaado	リザ	リザー	リ	リー
14	14	X	X	X	>3μ	saint	セイント	seiNto	セイ	セイン	セ	セン
15	15	X	X	X	>3μ	layer	レイヤー	reiyaa	レイ	レイヤ	レ	レヤ
16	16	X	X	X	>3μ	creep	クリープ	kuriipu	クリ	クリー	ク	クー
17	17	X	X	X	>3μ	collage	コラージュ	koraaju	コラ	コラー	コ	コー
18	18	X	X	X	>3μ	reward	リワード	riwaado	リワ	リワー	リ	リー
19	19	X	X	X	>3μ	galleon	ガリオン	garioN	ガリ	ガリオ	ガ	ガオ
20	20	X	X	X	>3μ	winter	ウインター	uiNtaa	ウイ	ウイン	ウ	ウン
21	21	X	X	X	>3μ	assist	アシスト	ashisuto	アシ	アシス	ア	アス
22	22	X	X	X	>3μ	number	ナンバー	naNbaa	ナン	ナンバ	ナ	ナバ
23	23	X	X	X	>3μ	count	カウント	kauNto	カウ	カウン	カ	カン
24	24	X	X	X	>3μ	texture	テクスチャー	tekusuchaa	テク	テクス	テ	テス
25	13	X	X	X	3μ	leather	レザー	rezaa	レザ	レザー	レ	レー
26	14	X	X	X	3μ	line	ライン	raiN	ライ	ライン	ラ	ラン
27	15	X	X	X	3μ	tire	タイヤ	taiya	タイ	タイヤ	タ	タヤ
28	16	X	X	X	3μ	collie	コリー	korii	コリ	コリー	コ	コー
29	17	X	X	X	3μ	seller	セラー	seraa	セラ	セラー	セ	セー
30	18	X	X	X	3μ	tower	タワー	tawaa	タワ	タワー	タ	ター
31	19	X	X	X	3μ	trio	トリオ	torio	トリ	トリオ	ト	トオ
32	20	X	X	X	3μ	gain	ゲイン	geiN	ゲイ	ゲイン	ゲ	ゲン
33	21	X	X	X	3μ	cassis	カシス	kashisu	カシ	カシス	カ	カス
34	22	X	X	X	3μ	samba	サンバ	saNba	サン	サンバ	サ	サバ
35	23	X	X	X	3μ	gown	ガウン	gauN	ガウ	ガウン	ガ	ガン
36	24	X	X	X	3μ	locus	ロクス	rokusu	ロク	ロクス	ロ	ロス
37	13	X	X	O	>3μ	hazard	ハザード	hazaado	ハザ	ハザー	ハ	ハー
38	14	X	X	O	>3μ	point	ポイント	poiNto	ポイ	ポイン	ポ	ポン
39	15	X	X	O	>3μ	hire	ハイヤー	haiyaa	ハイ	ハイヤ	ハ	ハヤ

40	16	X	X	O	>3μ	polyp	ポリープ	poriipu	ポリ	ポリー	ポ	ポー
41	17	X	X	O	>3μ	mirage	ミラージュ	miraaju	ミラ	ミラー	ミ	ミー
42	18	X	X	O	>3μ	forward	フォワード	fowaado	フォワ	フォワー	フォ	フォー
43	19	X	X	O	>3μ	million	ミリオン	mirioN	ミリ	ミリオ	ミ	ミオ
44	20	X	X	O	>3μ	painter	ペインター	peiNtaa	ペイ	ペイン	ペ	ペン
45	21	X	X	O	>3μ	fascist	ファシスト	fashisuto	ファシ	ファシス	ファ	ファス
46	22	X	X	O	>3μ	member	メンバー	meNbaa	メン	メンバ	メ	メバ
47	23	X	X	O	>3μ	mount	マウント	mauNto	マウ	マウン	マ	マン
48	24	X	X	O	>3μ	mixture	ミクスチャー	mikusuchaa	ミク	ミクス	ミ	ミス
49	25	O	X	O	>3μ	head	ヘッド	heddo	ヘッ	ヘッド	ヘ	ヘド
50	26	O	X	X	>3μ	nip	ニップ	nippu	ニッ	ニップ	ニ	ニプ
51	27	O	X	O	>3μ	buffer	バッファ	baffa	バッ	バッファ	バ	バファ
52	28	O	X	O	>3μ	back	バック	bakku	バッ	バック	バ	バク
53	29	O	X	X	>3μ	tap	タップ	tappu	タッ	タップ	タ	タップ
54	30	O	X	X	>3μ	tack	タック	takku	タッ	タック	タ	タク
55	31	O	X	O	>3μ	hook	HOOK	hokku	ホッ	HOOK	ホ	ホク
56	32	O	X	O	>3μ	mat	マット	matto	マッ	マット	マ	マト
57	33	O	X	X	>3μ	wreck	レック	rekku	レッ	レック	レ	レク
58	34	O	X	X	>3μ	sick	シック	shikku	シッ	シック	シ	シク
59	35	O	X	O	>3μ	patch	パッチ	pacchi	パッ	パッチ	パ	パチ
60	36	O	X	O	>3μ	pop	ポップ	poppu	ポッ	ポップ	ポ	ポップ
61	25	O	X	O	3μ	headline	ヘッドライン	heddoraiN	ヘッ	ヘッド	ヘ	ヘド
62	26	O	X	X	3μ	nipple	ニップル	nippuru	ニッ	ニップ	ニ	ニプ
63	27	O	X	O	3μ	buffalo	バッファロー	baffaroo	バッ	バッファ	バ	バファ
64	28	O	X	O	3μ	buckle	バックル	bakkuru	バッ	バック	バ	バク
65	29	O	X	X	3μ	dapple	ダップル	dappuru	タッ	タップ	タ	タップ
66	30	O	X	X	3μ	tax	タックス	takkusu	タッ	タック	タ	タク
67	31	O	X	O	3μ	box	ボックス	bokkusu	ボッ	ボック	ボ	ボク

68	32	O	X	O	3μ	mattress	マットレス	mattoresu	マツ	マット	マ	マト
69	33	O	X	X	3μ	rex	レックス	rekkesu	レッ	レック	レ	レク
70	34	O	X	X	3μ	six	シックス	shikkusu	シッ	シック	シ	シク
71	35	O	X	O	3μ	patchwork	パッチワーク	pacchiwaaku	パッ	パッチ	パ	パチ
72	36	O	X	O	3μ	popcorn	ポップコーン	poppukooN	ポッ	ポップ	ポ	ポップ
73	37	X	-H	X	3μ	loan	ローン	roon	ロー	ローン	ロ	ロン
74	38	X	-H	X	>3μ	sweater	セーター	seetaa	セー	セータ	セ	セタ
75	39	X	-H	X	3μ	date	デート	deeto	デー	デート	デ	デト
76	40	X	-H	X	>3μ	curler	カーラー	kaaraa	カー	カーラ	カ	カラ
77	41	X	-H	X	3μ	code	コード	koodo	コー	コード	コ	コド
78	42	X	-H	X	3μ	sale	セール	seeru	セー	セール	セ	セル
79	43	X	-H	X	>3μ	coaching	コーチング	koochiNgu	コー	コーチ	コ	コチ
80	44	X	-H	X	>3μ	sailing	セーリング	seeringu	セー	セーリ	セ	セリ
81	45	X	-H	X	3μ	notes	ノーツ	nootsu	ノー	ノーツ	ノ	ノツ
82	46	X	-H	X	>3μ	goggle	ゴーグル	googuru	ゴー	ゴーグ	ゴ	ゴグ
83	47	X	-H	X	3μ	safe	セーフ	seefu	セー	セーフ	セ	セフ
84	48	X	-H	X	3μ	curve	カーブ	kaabu	カー	カーブ	カ	カブ
85	37	X	+H	X	3μ	gene	ジーン	jiiN	ジー	ジーン	ジ	ジン
86	38	X	+H	X	>3μ	router	ルーター	ruutaa	ルー	ルータ	ル	ルタ
87	39	X	+H	X	3μ	cheat	チート	chiito	チー	チート	チ	チト
88	40	X	+H	X	>3μ	cooler	クーラー	kuuraa	クー	クーラ	ク	クラ
89	41	X	+H	X	3μ	lead	リード	riido	リー	リード	リ	リド
90	42	X	+H	X	3μ	rule	ルール	ruuru	ルー	ルール	ル	ルル
91	43	X	+H	X	>3μ	sheeting	シーティング	shiitiNgu	シー	シーティ	シ	シティ
92	44	X	+H	X	>3μ	sealing	シーリング	shiiriNgu	シー	シーリ	シ	シリ
93	45	X	+H	X	3μ	suits	スーツ	suutsu	スー	スーツ	ス	スツ
94	46	X	+H	X	>3μ	google	グーグル	guuguru	グー	グーグ	グ	ググ
95	47	X	+H	X	3μ	leaf	リーフ	riifu	リー	リーフ	リ	リフ

96	48	X	+H	X	3μ	leave	リーブ	riibu	リー	リーブ	リ	リブ
97	37	X	-H	O	3μ	main	メイン	meeN	メー	メイン	メ	メン
98	38	X	-H	O	>3μ	porter	ポーター	pootaa	ポー	ポータ	ポ	ポタ
99	39	X	-H	O	3μ	mate	メート	meeto	メー	メート	メ	メト
100	40	X	-H	O	>3μ	mailer	メーラー	meeraa	メー	メーラ	メ	メラ
101	41	X	-H	O	3μ	mode	モード	moodo	モー	モード	モ	モド
102	42	X	-H	O	3μ	hall	ホール	hooru	ホー	ホール	ホ	ホル
103	43	X	-H	O	>3μ	marching	マーチング	maachiNgu	マー	マーチ	マ	マチ
104	44	X	-H	O	>3μ	bowling	ボーリング	booriNgu	ボー	ボーリ	ボ	ポリ
105	45	X	-H	O	3μ	hearts	ハーツ	haatsu	ハー	ハーツ	ハ	ハツ
106	46	X	-H	O	>3μ	bagel	ベーグル	beeguru	ベー	ベーグ	ベ	ベグ
107	47	X	-H	O	3μ	half	ハーフ	haafu	ハー	ハーフ	ハ	ハフ
108	48	X	-H	O	3μ	mauve	モーブ	moobu	モー	モーブ	モ	モブ
109	37	X	+H	O	3μ	moon	ムーン	muuN	ムー	ムーン	ム	ムン
110	38	X	+H	O	>3μ	heater	ヒーター	hiitaa	ヒー	ヒータ	ヒ	ヒタ
111	39	X	+H	O	3μ	meat	ミート	miito	ミー	ミート	ミ	ミト
112	40	X	+H	O	>3μ	healer	ヒーラー	hiiraa	ヒー	ヒーラ	ヒ	ヒラ
113	41	X	+H	O	3μ	food	フード	fuudo	フー	フード	フ	フド
114	42	X	+H	O	3μ	fool	フール	fuuru	フー	フール	フ	フル
115	43	X	+H	O	>3μ	footing	フーティング	fuutiNgu	フー	フーティ	フ	フティ
116	44	X	+H	O	>3μ	feeling	フィーリング	fiiriNgu	フィー	フィーリ	フィ	フィリ
117	45	X	+H	O	3μ	boots	ブーツ	buutsu	ブー	ブーツ	ブ	ブツ
118	46	X	+H	O	>3μ	beagle	ビークル	biiguru	ビー	ビーク	ビ	ビク
119	47	X	+H	O	3μ	beef	ビーフ	biifu	ビー	ビーフ	ビ	ビフ
120	48	X	+H	O	3μ	move	ムーブ	muubu	ムー	ムーブ	ム	ムブ

Annexe 8 : liste de deuxièmes constituants utilisés pour les items de l'expérience ESP-CAL

Description des colonnes :

- N° : numéro arbitrairement donné aux constituants pour les identifier.
- Trio : le trio d'appartenance du constituant (voir section 5.1.1 pour le processus de formation des trios). Les trois nombres indiqués représentent les trois constituants membres du trio par leur N°.
- Score : le score du trio indiquant la proximité dans le traitement des différents membres du trio dans l'expérience baseline (voir section 5.1.1 pour le processus de calcul du score).
- Lis. : la liste d'appartenance d'origine du C2 lors de l'expérience baseline.
- S. Lis. : la sous-liste d'appartenance du C2. Un membre de chaque trio est réparti aléatoirement dans chaque sous-liste. Une sous-liste (sélectionnée aléatoirement) apparaît dans chaque phase de l'expérience ESP.
- Set : le set d'appartenance du constituant. Pour former les items, les C2 sont appairés aléatoirement avec un C1 possédant le même numéro de set. Pour chaque C2, ce numéro est identique à celui qu'il avait dans l'expérience baseline.
- Rang Main. : position (entre 1^{er} et 240^{ème}) dans le classement en termes de proportion de pattern Maintien reçu lors de l'expérience baseline. Le 1^{er} étant celui ayant la proportion la plus élevée. Cette information aide le partenaire artificiel à choisir sur quel constituant utiliser quelle stratégie (voir section 5.1.3).
- Rang Suivi : idem pour le pattern de Suivi.
- Rang Sup. : idem pour le pattern de Suppression.
- Rang Remp. : idem pour le pattern de Remplacement.
- /Q/ : la deuxième more du constituant est-elle /Q/ ? (O = oui ; X = non).

- /R/ : la deuxième more du constituant est-elle /R/ ? (X = non ; - H = oui, et la voyelle allongée est non haute ; + H = oui, et la voyelle allongée est haute).
- Bilab. : la consonne initiale est-elle une consonne bilabiale ? Le phonème /h/ est considéré comme une consonne bilabiale pour des raisons historiques et phonologiques le liant au phonème /p/ en japonais archaïque (Labrune, 2006b). (O = oui ; X = non)
- Long. : longueur du constituant en more (3μ = trois mores ; >3μ = plus de trois mores).
- Anglais : mot emprunté dans la langue d'origine.
- Katakana : transcription du mot en *katakana*
- Hepburn : translittération selon la méthode de romanisation Hepburn (avec voyelles longues dédoublées).
- Main. : forme abrégée selon le pattern de Maintien du constituant, telle qu'elle apparaissait dans la proposition de réponse correspondante.
- Suivi : forme abrégée selon le pattern de Suivi du constituant, telle qu'elle apparaissait dans la proposition de réponse correspondante.
- Sup. : forme abrégée selon le pattern de Suppression du constituant, telle qu'elle apparaissait dans la proposition de réponse correspondante.
- Remp. : forme abrégée selon le pattern de Remplacement du constituant, telle qu'elle apparaissait dans la proposition de réponse correspondante.

N°	Trio	Score	Lis.	S. Lis.	Set	Rang Main.	Rang Suivi	Rang Sup.	Rang Remp.	/Q/	/R/	Bilab.	Long.	Anglais	Katakana	Hepburn	Main.	Suivi	Sup.	Remp.
1	1/12/17	19.1	1	1	1	14	234	197	202	X	X	X	>3μ	cancel	キャンセル	kyaNseru	キャン	キャンセ	キャ	キャセ
2	2/16/21	12.14	1	3	2	15	212	210	221	X	X	X	>3μ	crown	クラウン	kurauN	クラ	クラウ	ク	クウ
3	15/20/23	39.38	2	2	3	29	153	219	234	X	X	X	>3μ	custom	カスタム	kasutamu	カス	カスタ	カ	カタ
4	2/16/21	12.14	2	1	4	19	206	232	215	X	X	X	>3μ	contact	コンタクト	koNtakuto	コン	コンタ	コ	コタ
5	1/12/17	19.1	2	2	5	7	239	217	214	X	X	X	>3μ	correction	コレクション	korekushoN	コレ	コレク	コ	コク
6	7/18/22	24.98	2	1	6	8	187	237	237	X	X	X	>3μ	stream	ストリーム	sutoriimu	スト	ストリ	ス	スリ
7	7/18/22	24.98	1	2	7	6	223	235	209	X	X	X	>3μ	quest	クエスト	kuesuto	クエ	クエス	ク	クス

8	15/20/23	39.38	2	1	8	30	207	184	212	X	X	X	>3μ	landing	ランディング	raNdiNgu	ラン	ランディ	ラ	ラディ
9	2/16/21	12.14	2	2	9	13	205	231	228	X	X	X	>3μ	snipe	スナイプ	sunaiipu	スナ	スナイ	ス	スイ
10	7/18/22	24.98	2	3	10	2	230	229	235	X	X	X	>3μ	technique	テクニック	tekunikku	テク	テクニ	テ	テニ
11	15/20/23	39.38	2	3	11	43	161	186	190	X	X	X	>3μ	thunder	サンダー	saNdaa	サン	サンダ	サ	サダ
12	1/12/17	19.1	1	3	12	16	235	227	166	X	X	X	>3μ	stealth	ステルス	suterusu	ステ	ステル	ス	スル
13	13/16/19	18.2	2	2	13	31	188	183	224	X	X	X	>3μ	lizard	リザード	rizaado	リザ	リザー	リ	リー
14	2/5/12	144.22	1	1	14	100	139	101	116	X	X	X	>3μ	relief	リリーフ	ririifu	リリ	リリー	リ	リー
15	3/8/18	65.34	1	1	15	49	229	200	107	X	X	X	>3μ	series	シリーズ	shiriizu	シリ	シリー	シ	シー
16	13/16/19	18.2	2	3	16	35	170	204	213	X	X	X	>3μ	creep	クリーブ	kuriipu	クリ	クリー	ク	クー
17	2/5/12	144.22	1	3	17	11	225	225	206	X	X	X	>3μ	slice	スライス	suraisu	スラ	スライ	ス	スイ
18	3/8/18	65.34	2	2	18	59	191	124	138	X	X	X	>3μ	reward	リワード	riwaado	リワ	リワー	リ	リー
19	13/16/19	18.2	2	1	19	36	160	185	231	X	X	X	>3μ	galleon	ガリオン	garioN	ガリ	ガリオ	ガ	ガオ
20	3/8/18	65.34	1	3	20	64	180	213	83	X	X	X	>3μ	single	シングル	shiNguru	シン	シング	シ	シグ
21	11/21/22	33.58	2	1	21	20	222	238	176	X	X	X	>3μ	assist	アシスト	ashisuto	アシ	アシス	ア	アス
22	11/21/22	33.58	2	3	22	37	189	220	175	X	X	X	>3μ	number	ナンバー	naNbaa	ナン	ナンバ	ナ	ナバ
23	11/21/22	33.58	1	2	23	33	217	181	186	X	X	X	>3μ	cantor	カントル	kaNtoru	カン	カント	カ	カト
24	2/5/12	144.22	1	2	24	44	179	150	199	X	X	X	>3μ	saloon	サルーン	saruuN	サル	サルー	サ	サー
25	13/15/16	30.32	2	1	13	88	93	110	188	X	X	X	3μ	leather	レザー	rezaa	レザ	レザー	レ	レー
26	7/14/20	47.2	2	2	14	121	81	112	129	X	X	X	3μ	line	ライン	raiN	ライ	ライン	ラ	ラン
27	13/15/16	30.32	2	3	15	73	135	78	168	X	X	X	3μ	tire	タイヤ	taiya	タイ	タイヤ	タ	タヤ
28	13/15/16	30.32	2	2	16	74	117	107	169	X	X	X	3μ	collie	コリー	korii	コリ	コリー	コ	コー
29	11/17/18	145.72	2	3	17	60	80	233	192	X	X	X	3μ	seller	セラー	seraa	セラ	セラー	セ	セー
30	11/17/18	145.72	2	2	18	122	7	175	211	X	X	X	3μ	tower	タワー	tawaa	タワ	タワー	タ	ター
31	7/14/20	47.2	1	1	19	140	74	55	143	X	X	X	3μ	coin	コイン	koiN	コイ	コイン	コ	コン
32	7/14/20	47.2	2	3	20	152	56	70	127	X	X	X	3μ	gain	ゲイン	geiN	ゲイ	ゲイン	ゲ	ゲン
33	21/22/24	33.34	2	2	21	57	172	191	118	X	X	X	3μ	cassis	カシス	kashisu	カシ	カシス	カ	カス
34	21/22/24	33.34	2	3	22	71	173	156	98	X	X	X	3μ	samba	サンバ	saNba	サン	サンバ	サ	サバ

35	11/17/18	145.72	1	1	23	52	198	86	193	X	X	X	3μ	tent	テント	teNto	テン	テント	テ	テト
36	21/22/24	33.34	2	1	24	61	162	192	119	X	X	X	3μ	locus	ロクス	rokusu	ロク	ロクス	ロ	ロス
37	1/4/6	22.96	1	3	13	26	227	166	217	X	X	O	>3μ	bride	ブライド	buraido	ブラ	ブライ	ブ	ブイ
38	14/15/21	45.44	2	2	14	89	210	24	167	X	X	O	>3μ	point	ポイント	poiNto	ポイ	ポイン	ポ	ボン
39	14/15/21	45.44	2	1	15	98	175	51	126	X	X	O	>3μ	hire	ハイヤー	haiyaa	ハイ	ハイヤ	ハ	ハヤ
40	1/4/6	22.96	1	2	16	28	197	180	219	X	X	O	>3μ	balance	バランス	baraNsu	バラ	バラン	バ	バン
41	8/11/17	97.1	2	2	17	47	125	187	232	X	X	O	>3μ	mirage	ミラージュ	miraaju	ミラ	ミラー	ミ	ミー
42	1/4/6	22.96	1	1	18	18	236	179	218	X	X	O	>3μ	plastic	プラスチック	purasuchikku	ブラ	プラス	ブ	プス
43	12/19/22	36.24	2	1	19	48	171	174	172	X	X	O	>3μ	million	ミリオン	mirioN	ミリ	ミリオ	ミ	ミオ
44	8/11/17	97.1	1	3	20	65	219	87	132	X	X	O	>3μ	bangle	バングル	baNguru	バン	バング	バ	バグ
45	14/15/21	45.44	2	3	21	75	174	60	158	X	X	O	>3μ	fascist	ファシスト	fashisuto	ファシ	ファシス	ファ	ファス
46	12/19/22	36.24	2	2	22	39	233	173	161	X	X	O	>3μ	member	メンバー	meNbaa	メン	メンバ	メ	メバ
47	8/11/17	97.1	1	1	23	42	228	116	195	X	X	O	>3μ	mantle	マントル	maNtoru	マン	マント	マ	マト
48	12/19/22	36.24	1	3	24	54	166	168	154	X	X	O	>3μ	balloon	バルーン	baruun	バル	バルー	バ	バー
49	1/3/4	19.68	1	2	25	213	66	4	62	O	X	X	>3μ	cup	カップ	kappu	カッ	カップ	カ	カブ
50	14/17/21	27.28	2	1	26	230	38	129	8	O	X	X	>3μ	nip	ニップ	nippu	ニッ	ニップ	ニ	ニブ
51	1/3/4	19.68	1	1	27	214	67	5	63	O	X	X	>3μ	hack	ハック	hakku	ハッ	ハック	ハ	ハク
52	1/3/4	19.68	1	3	28	206	53	14	64	O	X	O	>3μ	mash	マッシュ	masshu	マッ	マッシュ	マ	マシュ
53	14/17/21	27.28	2	3	29	232	13	163	12	O	X	X	>3μ	tap	タップ	tappu	タッ	タップ	タ	タブ
54	18/20/23	84.86	2	2	30	233	6	147	20	O	X	X	>3μ	tack	タック	takku	タッ	タック	タ	タク
55	7/10/24	34.22	1	1	31	226	11	34	47	O	X	X	>3μ	sack	サック	sakku	サッ	サック	サ	サク
56	18/20/23	84.86	2	3	32	234	10	72	29	O	X	O	>3μ	mat	マット	matto	マッ	マット	マ	マト
57	14/17/21	27.28	2	2	33	235	32	164	7	O	X	X	>3μ	wreck	レック	rekku	レッ	レック	レ	レク
58	7/10/24	34.22	1	2	34	208	15	21	66	O	X	X	>3μ	net	ネット	netto	ネッ	ネット	ネ	ネト
59	18/20/23	84.86	2	1	35	236	95	73	2	O	X	O	>3μ	patch	パッチ	pacchi	パッ	パッチ	パ	パチ
60	7/10/24	34.22	2	3	36	202	25	25	53	O	X	O	>3μ	pop	ポップ	poppu	ポッ	ポップ	ポ	ポブ
61	4/13/20	99.02	2	1	25	198	20	99	30	O	X	O	3μ	headline	ヘッドライン	heddoraiN	ヘッ	ヘッド	ヘ	ヘド

62	14/23/24	39.38	2	2	26	237	33	85	18	O	X	X	3μ	nipple	ニップル	nippuru	ニッ	ニップ	ニ	ニブ
63	3/5/19	33.28	1	2	27	217	99	12	42	O	X	O	3μ	hackle	ハックル	hakkuru	ハッ	ハック	ハ	ハク
64	4/13/20	99.02	1	2	28	218	40	33	38	O	X	O	3μ	mushroom	マッシュルーム	masshuruumu	マッ	マッシュ	マ	マシュ
65	3/5/19	33.28	1	3	29	209	152	3	37	O	X	O	3μ	pickle	ピッケル	pikkuru	ピッ	ピック	ピ	ピク
66	18/21/22	30.32	2	1	30	223	26	177	10	O	X	X	3μ	tax	タックス	takkusu	タッ	タック	タ	タク
67	3/5/19	33.28	2	1	31	224	121	6	39	O	X	O	3μ	box	ボックス	bokkusu	ボッ	ボック	ボ	ボク
68	4/13/20	99.02	2	3	32	204	2	115	59	O	X	O	3μ	mattress	マットレス	mattoresu	マッ	マット	マ	マト
69	18/21/22	30.32	2	2	33	238	46	178	1	O	X	X	3μ	rex	レックス	rekusu	レッ	レック	レ	レク
70	18/21/22	30.32	2	3	34	212	37	161	9	O	X	X	3μ	six	シックス	shikkusu	シッ	シック	シ	シク
71	14/23/24	39.38	2	3	35	239	14	148	15	O	X	O	3μ	patchwork	パッチワーク	pacchiwaaku	パッ	パッチ	パ	パチ
72	14/23/24	39.38	2	1	36	225	12	84	26	O	X	O	3μ	popcorn	ポップコーン	poppukooN	ポッ	ポップ	ポ	ポブ
73	11/13/21	80.82	2	2	37	185	19	160	55	X	-H	X	3μ	loan	ローン	roon	ロー	ローン	ロ	ロン
74	2/4/8	52.48	1	2	38	101	131	43	180	X	-H	X	3μ	kart	カート	kaato	カー	カート	カ	カト
75	7/12/15	33.38	2	1	39	145	54	61	159	X	-H	X	3μ	date	デート	deeto	デー	デート	デ	デト
76	2/4/8	52.48	1	1	40	126	88	45	181	X	-H	X	3μ	road	ロード	roodo	ロー	ロード	ロ	ロド
77	5/6/22	22.6	1	2	41	79	148	121	133	X	-H	X	>3μ	gamer	ゲーマー	geemaa	ゲー	ゲーマ	ゲ	ゲマ
78	5/6/22	22.6	1	1	42	92	149	133	106	X	-H	X	>3μ	rating	レーティング	reetiNgu	レー	レーティ	レ	レティ
79	7/12/15	33.38	1	2	43	158	75	40	142	X	-H	X	3μ	girl	ガール	gaaru	ガー	ガール	ガ	ガル
80	2/4/8	52.48	1	3	44	103	63	75	229	X	-H	X	3μ	sauce	ソース	soosu	ソー	ソース	ソ	ソス
81	11/13/21	80.82	2	1	45	146	35	125	117	X	-H	X	3μ	notes	ノーツ	nootsu	ノー	ノーツ	ノ	ノツ
82	5/6/22	22.6	2	3	46	76	154	108	128	X	-H	X	>3μ	goggle	ゴーグル	googuru	ゴー	ゴーク	ゴ	ゴグ
83	11/13/21	80.82	1	3	47	177	27	202	58	X	-H	X	>3μ	tones	トーンズ	toonZu	トー	トーン	ト	トン
84	7/12/15	33.38	1	3	48	159	89	47	122	X	-H	X	3μ	zone	ゾーン	zooN	ゾー	ゾーン	ゾ	ゾン
85	12/13/21	90.9	2	2	37	138	100	207	48	X	+H	X	3μ	gene	ジーン	jiiN	ジー	ジーン	ジ	ジン
86	4/14/18	45	2	1	38	147	69	144	88	X	+H	X	>3μ	router	ルーター	ruutaa	ルー	ルータ	ル	ルタ
87	10/11/15	43.02	2	1	39	107	106	95	136	X	+H	X	3μ	cheat	チート	chiito	チー	チート	チ	チト
88	4/14/18	45	1	2	40	135	50	103	146	X	+H	X	3μ	seed	シード	shiido	シー	シード	シ	シド

89	5/8/19	31.98	1	2	41	118	141	123	81	X	+H	X	>3μ	zoomer	ズーマー	zuumaa	ズー	ズーマ	ズ	ズマ
90	4/14/18	45	2	3	42	148	55	126	101	X	+H	X	3μ	rule	ルール	ruuru	ルー	ルール	ル	ルル
91	5/8/19	31.98	2	1	43	112	119	194	77	X	+H	X	>3μ	sheeting	シーティング	shiitiNgu	シー	シーティ	シ	シティ
92	5/8/19	31.98	1	3	44	115	132	135	91	X	+H	X	3μ	lease	リース	riisu	リー	リース	リ	リス
93	12/13/21	90.9	2	1	45	113	41	221	120	X	+H	X	3μ	suits	スーツ	suutsu	スー	スーツ	ス	スツ
94	10/11/15	43.02	1	3	46	80	96	120	198	X	+H	X	3μ	team	チーム	chiimu	チー	チーム	チ	チム
95	10/11/15	43.02	1	2	47	105	87	134	149	X	+H	X	>3μ	jeans	ジーンズ	jiiNzu	ジー	ジーン	ジ	ジン
96	12/13/21	90.9	1	3	48	163	22	215	84	X	+H	X	3μ	scene	シーン	shiiN	シー	シーン	シ	シン
97	1/2/9	114.76	1	1	37	131	202	8	162	X	-H	O	>3μ	paper	ペーパー	peepaa	ペー	ペーパー	ペ	ペパ
98	1/2/9	114.76	1	2	38	93	182	37	153	X	-H	O	3μ	part	パート	paato	パー	パート	パ	パト
99	11/12/15	36.78	2	3	39	183	45	63	96	X	-H	O	3μ	mate	メート	meeto	メー	メート	メ	メト
100	4/7/10	52.46	1	3	40	119	72	44	216	X	-H	O	3μ	board	ボード	boodo	ボー	ボード	ボ	ボド
101	8/17/18	41.88	2	3	41	171	103	18	124	X	-H	O	3μ	mode	モード	moodo	モー	モード	モ	モド
102	8/17/18	41.88	2	1	42	161	102	28	125	X	-H	O	3μ	hall	ホール	hooru	ホー	ホール	ホ	ホル
103	4/7/10	52.46	1	1	43	132	73	54	163	X	-H	O	3μ	ball	ボール	booru	ボー	ボール	ボ	ボル
104	8/17/18	41.88	1	2	44	136	115	20	178	X	-H	O	3μ	base	ベース	peesu	ベー	ベース	ベ	ベス
105	1/2/9	114.76	1	3	45	110	184	122	68	X	-H	O	>3μ	hurdle	ハードル	haadoru	ハー	ハード	ハ	ハド
106	4/7/10	52.46	1	2	46	164	34	48	182	X	-H	O	3μ	home	ホーム	hoomu	ホー	ホーム	ホ	ホム
107	11/12/15	36.78	1	2	47	172	51	104	89	X	-H	O	>3μ	bones	ボーンズ	booNzu	ボー	ボーン	ボ	ボン
108	11/12/15	36.78	1	1	48	182	28	105	90	X	-H	O	3μ	burn	バーン	baaN	バー	バーン	バ	バン
109	1/5/20	36.92	1	3	37	111	185	65	99	X	+H	O	>3μ	Hooper	フーパー	huupaa	フー	フーパー	フ	フパ
110	14/15/21	87.86	2	1	38	108	107	79	151	X	+H	O	>3μ	heater	ヒーター	hiitaa	ヒー	ヒーター	ヒ	ヒタ
111	14/15/21	87.86	2	2	39	149	8	157	152	X	+H	O	3μ	meat	ミート	miito	ミー	ミート	ミ	ミト
112	4/10/19	27.92	1	1	40	87	123	131	148	X	+H	O	3μ	mood	ムード	muudo	ムー	ムード	ム	ムド
113	1/5/20	36.92	1	1	41	106	159	90	104	X	+H	O	>3μ	boomer	ブーマー	boomaa	ブー	ブーマ	ブ	ブマ
114	12/18/23	28.94	2	1	42	144	94	142	76	X	+H	O	3μ	fool	フール	fuuru	フー	フール	フ	フル
115	4/10/19	27.92	2	3	43	81	136	109	137	X	+H	O	>3μ	footing	フーティング	fuutiNgu	フー	フーティ	フ	フティ

116	1/5/20	36.92	2	2	44	82	209	68	108	X	+H	O	>3μ	feeling	フィーリング	fiiriNgu	フィー	フィーリ	フィ	フィリ
117	14/15/21	87.86	2	3	45	99	60	139	171	X	+H	O	3μ	boots	ブーツ	buutsu	ブー	ブーツ	ブ	ブツ
118	4/10/19	27.92	1	2	46	94	150	74	144	X	+H	O	3μ	beam	ビーム	biimu	ビー	ビーム	ビ	ビム
119	12/18/23	28.94	2	2	47	150	83	158	72	X	+H	O	3μ	beef	ビーフ	biifu	ビー	ビーフ	ビ	ビフ
120	12/18/23	28.94	1	3	48	169	65	137	69	X	+H	O	3μ	mean	ミーン	miiN	ミー	ミーン	ミ	ミン

Annexe 9 : liste de pseudo-verbos utilisés pour les items des expériences sur les flexions verbales

Description des colonnes :

- N° : numéro arbitrairement donné aux pseudo-verbos pour les identifier.
- Term. : la terminaison du pseudo-verbe.
- Long. : longueur du pseudo-verbe, en mores.
- F1 : forme n°1 du verbe (voir section 7.1.1).
- F2 : forme n°2 du verbe (voir section 7.1.1).
- Innac. Rom. : forme inaccomplie du pseudo-verbe dans une romanisation proche d'une transcription phonologique. Les apostrophes « ' » représentent une attaque consonantique vide. Le N représente la nasale homorganique.
- Innac. Hir. : forme inaccomplie du pseudo-verbe en *hiragana*, telle qu'elle apparaissait dans les dialogues présentés aux participants.
- Ichidan Acc. Rom. : romanisation de la forme accomplie *ichidan* du pseudo-verbe.
- Ichidan Acc. Hir. : forme accomplie *ichidan* du pseudo-verbe en *hiragana*, telle qu'elle apparaissait dans la proposition de réponse correspondante.
- Yodan Acc. Rom. : romanisation de la forme accomplie *yodan* du pseudo-verbe.
- Yodan Acc. Hir. : forme accomplie *yodan* du pseudo-verbe en *hiragana*, telle qu'elle apparaissait dans la proposition de réponse correspondante.
- Ichidan Nég. Rom. : romanisation de la forme négative *ichidan* du pseudo-verbe.
- Ichidan Nég. Hir. : forme négative *ichidan* du pseudo-verbe en *hiragana*, telle qu'elle apparaissait dans la proposition de réponse correspondante.
- Yodan Nég. Rom. : romanisation de la forme négative *yodan* du pseudo-verbe.

- Yodan Nég. Hir. : forme négative *yodan* du pseudo-verbe en *hiragana*, telle qu'elle apparaissait dans la proposition de réponse correspondante.

N°	Term.	Long.	F1	F2	Innac. Rom.	Innac. Hir.	Ichidan Acc. Rom.	Ichidan Acc. Hir	Yodan Acc. Rom.	Yodan Acc. Hir.	Ichidan Nég. Rom.	Ichidan Nég. Hir	Yodan Nég. Rom.	Yodan Nég. Hir.
1	IRU	3	Acc.	Nég.	hutiru	ふちる	hutita	ふちた	hutitta	ふちった	hutinai	ふちない	hutiranai	ふちらない
2	IRU	3	Acc.	Nég.	sodiru	そじる	sodita	そじた	soditta	そじった	sodinai	そじない	sodiranai	そじらない
3	IRU	3	Acc.	Nég.	'emiru	えみる	'emita	えみた	'emitta	えみった	'eminai	えみない	'emiranai	えみらない
4	IRU	3	Acc.	Nég.	metiru	めちる	metita	めちた	metitta	めちった	metinai	めちない	metiranai	めちらない
5	IRU	3	Acc.	Nég.	satiru	さちる	satita	さちた	satitta	さちった	satinai	さちない	satiranai	さちらない
6	IRU	3	Acc.	Nég.	muziru	むじる	muzita	むじた	muzitta	むじった	muzinai	むじない	muziranai	むじらない
7	IRU	3	Acc.	Nég.	nakiru	なきる	nakita	なきた	nakitta	なきった	nakinai	なきない	nakiranai	なきらない
8	IRU	3	Nég.	Acc.	saniru	さなる	sanita	さにた	sanitta	さになった	saninai	さにない	saniranai	さにらない
9	IRU	3	Nég.	Acc.	naniru	なる	nanita	なにた	nanitta	なにになった	naninai	なにない	naniranai	なにらない
10	IRU	3	Nég.	Acc.	sakiru	さきる	sakita	さきた	sakitta	さきった	sakinai	さきない	sakiranai	さきらない
11	IRU	3	Nég.	Acc.	sariru	さる	sarita	さりた	saritta	さりった	sarinai	さりない	sariranai	さりらない
12	IRU	3	Nég.	Acc.	sigiru	しきる	sigita	しぎた	sigitta	しぎった	siginai	しぎない	sigiranai	しぎらない
13	IRU	3	Nég.	Acc.	hotiru	ほちる	hotita	ほちた	hotitta	ほちった	hotinai	ほちない	hotiranai	ほちらない
14	IRU	3	Nég.	Acc.	mamiru	まみる	mamita	まみた	mamitta	まみった	maminai	まみない	mamiranai	まみらない
15	IRU	3	Nég.	Acc.	muriru	むりる	murita	むりた	muritta	むりった	murinai	むりない	muriranai	むりらない
16	IRU	4	Nég.	Acc.	mobu'iru	もぶいる	mobu'ita	もぶいた	mobu'itta	もぶいった	mobu'inai	もぶいない	mobu'iranai	もぶいらない
17	IRU	4	Nég.	Acc.	kigikiru	きぎきる	kigikita	きぎきた	kigikitta	きぎきった	kigikinai	きぎきない	kigikiranai	きぎきらない
18	IRU	4	Nég.	Acc.	'akoniru	あこなる	'akonita	あこにた	'akonitta	あこにった	'akoninai	あこにない	'akoniranai	あこにらない
19	IRU	4	Nég.	Acc.	kuzukiru	くずきる	kuzukita	くずきた	kuzukitta	くずきった	kuzukinai	くずきない	kuzukiranai	くずきらない
20	IRU	4	Nég.	Acc.	ka'ohiru	かおひる	ka'ohita	かおひた	ka'ohitta	かおひった	ka'ohinai	かおひない	ka'ohiranai	かおひらない
21	IRU	4	Nég.	Acc.	kuguriru	くぐりる	kugurita	くぐりた	kuguritta	くぐりった	kugurinai	くぐりない	kuguriranai	くぐりらない

22	IRU	4	Nég.	Acc.	kimo'iru	きもいる	kimo'ita	きもいた	kimo'itta	きもいった	kimo'inai	きもいない	kimo'iranai	きもいらない
23	IRU	4	Nég.	Acc.	hamogiru	はもぎる	hamogita	はもぎた	hamogitta	はもぎった	hamoginai	はもぎない	hamogiranai	はもぎらない
24	IRU	4	Acc.	Nég.	hinokiru	ひのきる	hinokita	ひのきた	hinokitta	ひのきった	hinokinai	ひのきない	hinokiranai	ひのきらない
25	IRU	4	Acc.	Nég.	koyokiru	こよきる	koyokita	こよきた	koyokitta	こよきった	koyokinai	こよきない	koyokiranai	こよきらない
26	IRU	4	Acc.	Nég.	'omemiru	おめみる	'omemita	おめみた	'omemitta	おめみった	'omeminai	おめみない	'omemiranai	おめみらない
27	IRU	4	Acc.	Nég.	sikusiru	しくしる	sikusita	しくした	sikusitta	しくしった	sikusinai	しくしない	sikusiranai	しくしらない
28	IRU	4	Acc.	Nég.	togebiru	とげびる	togebita	とげびた	togebitta	とげびった	togebinai	とげびない	togebiranai	とげびらない
29	IRU	4	Acc.	Nég.	nirosiru	にろしる	nirosita	にろした	nirositta	にろしった	nirosinai	にろしない	nirosiranai	にろしらない
30	IRU	4	Acc.	Nég.	notisiru	のちしる	notisita	のちした	notisitta	のちしった	notisinai	のちしない	notisiranai	のちしらない
31	ERU	3	Nég.	Acc.	koberu	こべる	kobeta	こべた	kobetta	こべった	kobenai	こべない	koberanai	こべらない
32	ERU	3	Nég.	Acc.	heteru	へてる	heteta	へてた	hetetta	へてった	hetenai	へてない	heteranai	へてらない
33	ERU	3	Nég.	Acc.	'oneru	おねる	'oneta	おねた	'onetta	おねった	'onenai	おねない	'oneranai	おねらない
34	ERU	3	Nég.	Acc.	neseru	ねせる	neseta	ねせた	nesetta	ねせった	nesenai	ねせない	neseranai	ねせらない
35	ERU	3	Nég.	Acc.	kideru	きでる	kideta	きでた	kidetta	きでった	kidenai	きでない	kideranai	きでらない
36	ERU	3	Nég.	Acc.	heneru	へねる	heneta	へねた	henetta	へねった	henenai	へねない	heneranai	へねらない
37	ERU	3	Nég.	Acc.	nuberu	ぬべる	nubeta	ぬべた	nubetta	ぬべった	nubenai	ぬべない	nuberanai	ぬべらない
38	ERU	3	Acc.	Nég.	'omeru	おめる	'ometa	おめた	'ometta	おめった	'omenai	おめない	'omeranai	おめらない
39	ERU	3	Acc.	Nég.	koderu	こでる	kodeta	こでた	kodetta	こでった	kodenai	こでない	koderanai	こでらない
40	ERU	3	Acc.	Nég.	toneru	とねる	toneta	とねた	tonetta	とねった	tonenai	とねない	toneranai	とねらない
41	ERU	3	Acc.	Nég.	mekeru	めける	meketa	めけた	meketta	めけった	mekenai	めけない	mekeranai	めけらない
42	ERU	3	Acc.	Nég.	nineru	にねる	nineta	にねた	ninetta	にねった	ninenai	にねない	nineranai	にねらない
43	ERU	3	Acc.	Nég.	'ekeru	えける	'eketa	えけた	'eketta	えけった	'ekenai	えけない	'ekeranai	えけらない
44	ERU	3	Acc.	Nég.	zuteru	ずてる	zuteta	ずてた	zutetta	ずてった	zutenai	ずてない	zuteranai	ずてらない
45	ERU	3	Acc.	Nég.	nuderu	ぬでる	nudeta	ぬでた	nudetta	ぬでった	nudenai	ぬでない	nuderanai	ぬでらない
46	ERU	4	Acc.	Nég.	mari'eru	まりえる	mari'eta	まりえた	mari'etta	まりえった	mari'enai	まりえない	mari'eranai	まりえらない
47	ERU	4	Acc.	Nég.	hazoseru	はぞせる	hazoseta	はぞせた	hazosetta	はぞせった	hazosenai	はぞせない	hazoseranai	はぞせらない
48	ERU	4	Acc.	Nég.	bonameru	ぼなめる	bonameta	ぼなめた	bonametta	ぼなめった	bonamenai	ぼなめない	bonameranai	ぼなめらない

49	ERU	4	Acc.	Nég.	ga'akeru	があける	ga'aketa	があけた	ga'aketta	があけった	ga'akenai	があけない	ga'akeranai	があけらない
50	ERU	4	Acc.	Nég.	gemameru	げまめる	gemameta	げまめた	gemametta	げまめった	gemamenai	げまめない	gemameranai	げまめらない
51	ERU	4	Acc.	Nég.	sunazeru	すなぜる	sunazeta	すなぜた	sunazetta	すなぜった	sunazenai	すなぜない	sunazeranai	すなぜらない
52	ERU	4	Acc.	Nég.	hunoneru	ふのねる	hunoneta	ふのねた	hunonetta	ふのねった	hunonenai	ふのねない	hunoneranai	ふのねらない
53	ERU	4	Acc.	Nég.	karokeru	かろける	karoketa	かろけた	karoketta	かろけった	karokenai	かろけない	karokeranai	かろけらない
54	ERU	4	Nég.	Acc.	tamateru	たまてる	tamateta	たまたた	tamatetta	たまたった	tamatenai	たまたない	tamateranai	たまたらない
55	ERU	4	Nég.	Acc.	'asimeru	あしめる	'asimeta	あしめた	'asimetta	あしめった	'asimenai	あしめない	'asimeranai	あしめらない
56	ERU	4	Nég.	Acc.	na'ameru	なあめる	na'ameta	なあめた	na'ametta	なあめった	na'amenai	なあめない	na'ameranai	なあめらない
57	ERU	4	Nég.	Acc.	hunureru	ふぬれる	hunureta	ふぬれた	hunuretta	ふぬれった	hunurenai	ふぬれない	hunureranai	ふぬれらない
58	ERU	4	Nég.	Acc.	tadakeru	ただける	tadaketa	ただけた	tadaketta	ただけった	tadakenai	ただけない	tadakeranai	ただけらない
59	ERU	4	Nég.	Acc.	tusekeru	つせける	tuseketa	つせけた	tuseketta	つせけった	tusekenai	つせけない	tusekeranai	つせけらない
60	ERU	4	Nég.	Acc.	'izaseru	いざせる	'izaseta	いざせた	'izasetta	いざせった	'izasenai	いざせない	'izaseranai	いざせらない
61	ARU	3	Acc.	Nég.	nanaru	ななる	nanata	ななた	nanatta	ななった	nananai	ななない	nanaranai	なならない
62	ARU	3	Acc.	Nég.	setaru	せたる	setata	せたた	setatta	せたった	setanai	せたない	setaranai	せたらない
63	ARU	3	Acc.	Nég.	hasaru	はさる	hasata	はさた	hasatta	はさった	hasanai	はさない	hasaranai	はさらない
64	ARU	3	Acc.	Nég.	kasaru	かさる	kasata	かさた	kasatta	かさった	kasandai	かさない	kasaranai	かさらない
65	ARU	3	Acc.	Nég.	tokaru	とかる	tokata	とかた	tokatta	とかった	tokanai	とかない	tokaranai	とからない
66	ARU	3	Nég.	Acc.	horaru	ほらる	horata	ほらた	horatta	ほらった	horanai	ほらない	horaranai	ほららない
67	ARU	3	Nég.	Acc.	suraru	すらる	surata	すらた	suratta	すらった	suranai	すらない	suraranai	すららない
68	ARU	3	Nég.	Acc.	heraru	へらる	herata	へらた	heratta	へらった	heranai	へらない	heraranai	へららない
69	ARU	3	Nég.	Acc.	ka'aru	かある	ka'ata	かあた	ka'atta	かあった	ka'anai	かあない	ka'aranai	かあらない
70	ARU	3	Nég.	Acc.	si'aru	しある	si'ata	しあた	si'atta	しあった	si'anai	しあない	si'aranai	しあらない
71	ARU	4	Acc.	Nég.	kukataru	くかたる	kukatata	くかたた	kukatatta	くかたった	kukatanai	くかたない	kukataranai	くかたらない
72	ARU	4	Acc.	Nég.	korekaru	これかる	korekata	これかた	korekatta	これかった	korekanai	これかない	korekaranai	これからない
73	ARU	4	Acc.	Nég.	'utibaru	うちばる	'utibata	うちばた	'utibatta	うちばった	'utibanai	うちばない	'utibaranai	うちばらない
74	ARU	4	Acc.	Nég.	humamaru	ふままる	humamata	ふままた	humamatta	ふままた	humamanai	ふままない	humamaranai	ふままらない
75	ARU	4	Acc.	Nég.	'ururaru	うるらる	'ururata	うるらた	'ururatta	うるらった	'ururanai	うるらない	'ururaranai	うるららない

76	ARU	4	Nég.	Acc.	nemokaru	ねもかる	nemokata	ねもかた	nemokatta	ねもかった	nemokanai	ねもかない	nemokaranai	ねもからない
77	ARU	4	Nég.	Acc.	humomaru	ふもまる	humomata	ふもまた	humomatta	ふもまった	humomanai	ふもまない	humomaranai	ふもまらない
78	ARU	4	Nég.	Acc.	tokutaru	とくたる	tokutata	とくたた	tokutatta	とくたった	tokutanai	とくたない	tokutaranai	とくたらない
79	ARU	4	Nég.	Acc.	kariharu	かりはる	karihata	かりはた	karihatta	かりはった	karihanai	かりはない	kariharanai	かりはらない
80	ARU	4	Nég.	Acc.	misataru	みさたる	misatata	みさたた	misatatta	みさたった	misatanai	みさたない	misataranai	みさたらない
81	ORU	3	Acc.	Nég.	higoru	ひごる	higota	ひごた	higotta	ひごった	higonai	ひごない	higoranai	ひごらない
82	ORU	3	Acc.	Nég.	waboru	わぼる	wabota	わぼた	wabotta	わぼった	wabonai	わぼない	waboranai	わぼらない
83	ORU	3	Acc.	Nég.	zororu	ぞろる	zorota	ぞろた	zorotta	ぞろった	zoronai	ぞろない	zororanai	ぞろらない
84	ORU	3	Acc.	Nég.	mokoru	もこる	mokota	もこた	mokotta	もこった	mokonai	もこない	mokoranai	もこらない
85	ORU	3	Acc.	Nég.	'unoru	うのる	'unota	うのた	'unotta	うのった	'unonai	うのない	'unoranai	うのらない
86	ORU	3	Nég.	Acc.	sagoru	さごる	sagota	さごた	sagotta	さごった	sagonai	さごない	sagoranai	さごらない
87	ORU	3	Nég.	Acc.	'uroru	うろる	'urota	うろた	'urotta	うろった	'uronai	うろない	'uroranai	うろらない
88	ORU	3	Nég.	Acc.	ne'oru	ねおる	ne'ota	ねおた	ne'otta	ねおった	ne'onai	ねおない	ne'oranai	ねおらない
89	ORU	3	Nég.	Acc.	mozoru	もぞる	mozota	もぞた	mozotta	もぞった	mozonai	もぞない	mozoranai	もぞらない
90	ORU	3	Nég.	Acc.	haroru	はろる	harota	はろた	harotta	はろった	haronai	はろない	haroranai	はろらない
91	ORU	4	Acc.	Nég.	tukamoru	つかもる	tukamota	つかもた	tukamotta	つかもった	tukamonai	つかもない	tukamoranai	つかもらない
92	ORU	4	Acc.	Nég.	honotoru	ほのとる	honotota	ほのとた	honototta	ほのとった	honotonai	ほのとない	honotoranai	ほのとらない
93	ORU	4	Acc.	Nég.	meritoru	めりとる	meritota	めりとた	meritotta	めりとった	meritonai	めりとない	meritoranai	めりとらない
94	ORU	4	Acc.	Nég.	ziramoru	じらもる	ziramota	じらもた	ziramotta	じらもった	ziramonai	じらもない	ziramoranai	じらもらない
95	ORU	4	Acc.	Nég.	'e'anoru	えあのる	'e'anota	えあのた	'e'anotta	えあのった	'e'anonai	えあのない	'e'anoranai	えあのらない
96	ORU	4	Nég.	Acc.	sewakoru	せわこる	sewakota	せわこた	sewakotta	せわこった	sewakonai	せわこない	sewakoranai	せわこらない
97	ORU	4	Nég.	Acc.	sidotoru	しどとる	sidotota	しどとた	sidototta	しどとった	sidotonai	しどとない	sidotoranai	しどとらない
98	ORU	4	Nég.	Acc.	hanutoru	はぬとる	hanutota	はぬとた	hanutotta	はぬとった	hanutonai	はぬとない	hanutoranai	はぬとらない
99	ORU	4	Nég.	Acc.	'anisoru	あにそる	'anisota	あにそた	'anisotta	あにそった	'anisonai	あにそない	'anisoranai	あにそらない
100	ORU	4	Nég.	Acc.	hozururu	ほずろる	hozurota	ほずろた	hozurotta	ほずろった	hozuronai	ほずろない	hozuroranai	ほずろらない
101	URU	3	Acc.	Nég.	niruru	にるる	niruta	にるた	nirutta	にるった	nirunai	にるない	niruranai	にるらない
102	URU	3	Acc.	Nég.	'aturu	あつる	'atuta	あつた	'atutta	あつった	'atunai	あつない	'aturanai	あつらない

103	URU	3	Acc.	Nég.	bosuru	ぼする	bosuta	ぼすた	bosutta	ぼすった	bosunai	ぼすない	bosuranai	ぼすらない
104	URU	3	Acc.	Nég.	samuru	さむる	samuta	さむた	samutta	さむった	samunai	さむない	samuranai	さむらない
105	URU	3	Acc.	Nég.	yoguru	よぐる	yoguta	よぐた	yogutta	よぐった	yogunai	よぐない	yoguranai	よぐらない
106	URU	3	Nég.	Acc.	'u'uru	ううる	'u'uta	ううた	'u'utta	ううった	'u'unai	ううない	'u'uranai	ううらない
107	URU	3	Nég.	Acc.	sokuru	そくる	sokuta	そくた	sokutta	そくった	sokunai	そくない	sokuranai	そくらない
108	URU	3	Nég.	Acc.	na'uru	なうる	na'uta	なうた	na'utta	なうった	na'unai	なうない	na'uranai	なうらない
109	URU	3	Nég.	Acc.	momuru	もむる	momuta	もむた	momutta	もむった	momunai	もむない	momuranai	もむらない
110	URU	3	Nég.	Acc.	'iguru	いぐる	'iguta	いぐた	'igutta	いぐった	'igunai	いぐない	'iguranai	いぐらない
111	URU	4	Acc.	Nég.	sosumuru	そすむる	sosumuta	そすむた	sosumutta	そすむった	sosumunai	そすむない	sosumuranai	そすむらない
112	URU	4	Acc.	Nég.	satozuru	さとずる	satozuta	さとずた	satozutta	さとずった	satozunai	さとずない	satozuranai	さとずらない
113	URU	4	Acc.	Nég.	wazokuru	わぞくる	wazokuta	わぞくた	wazokutta	わぞくった	wazokunai	わぞくない	wazokuranai	わぞくらない
114	URU	4	Acc.	Nég.	'ebeyuru	えべゆる	'ebeyuta	えべゆた	'ebeyutta	えべゆった	'ebeyunai	えべゆない	'ebeyuranai	えべゆらない
115	URU	4	Acc.	Nég.	tokumuru	とくむる	tokumuta	とくむた	tokumutta	とくむった	tokumunai	とくむない	tokumuranai	とくむらない
116	URU	4	Nég.	Acc.	horusuru	ほるする	horusuta	ほるすた	horusutta	ほるすった	horusunai	ほるすない	horusuranai	ほるすらない
117	URU	4	Nég.	Acc.	tusemuru	つせむる	tusemuta	つせむた	tusemutta	つせむった	tusemunai	つせむない	tusemuranai	つせむらない
118	URU	4	Nég.	Acc.	mogusuru	もぐする	mogusuta	もぐすた	mogusutta	もぐすった	mogusunai	もぐすない	mogusuranai	もぐすらない
119	URU	4	Nég.	Acc.	muNzuru	むんずる	muNzuta	むんずた	muNzutta	むんずった	muNzunai	むんずない	muNzuranai	むんずらない
120	URU	4	Nég.	Acc.	kenakuru	けなくる	kenakuta	けなくた	kenakutta	けなくった	kenakunai	けなくない	kenakuranai	けなくらない

Annexe 10 : liste des dialogues utilisés pour les items des expériences sur les flexions verbales.

Description des colonnes :

- N° : numéro arbitrairement donné aux dialogues pour les identifier.
- Flex. : la flexion verbale que requiert ce dialogue pour le pseudo-verbe dans la réponse du deuxième personnage (Acc. = accomplie ; Nég. = négation).
- Lang. : la langue dans laquelle le dialogue est retranscrit sur cette ligne (Jap. = japonais ; Trad. = traduction française). Dans les transcriptions japonaises, la place du pseudo-verbe est remplacée par un blanc « ____ ». Dans les traductions françaises, le radical du pseudo-verb es est remplacé par X mais la conjugaison reste apparente, par souci de fluidité de lecture.
- Réplique du premier personnage : la première phrase du dialogue dans laquelle le pseudo-verbe apparaît à la forme inaccomplie. Même si la traduction ne laisse pas toujours transparaître qu’une forme inaccomplie est attendue, c’est bien le cas en japonais.
- Réponse du deuxième personnage : la deuxième partie du dialogue dans laquelle le pseudo-verbe devrait apparaître à la forme indiquée dans la colonne Flex. Même si la traduction ne laisse pas toujours transparaître que la flexion indiquée dans la colonne Flex. est attendue, c’est bien le cas en japonais.

N°	Flex.	Lang.	Réplique premier personnage	Réponse deuxième personnage
1	Acc.	Jap.	萌 : 翔太、やばい! 社長は仕事の時に ____ 人を絶対に許さないって言ってたよ!	翔太 : そうなんだよ。一度 ____ だけなのに、クビになるかもしれない。
		Trad.	Moe : Shōta ! C’est grave ! Le patron a dit qu’il ne tolérerait pas que l’on X pendant les heures de travail.	Shōta : C’est vrai ! Même si j’ai Xé qu’une fois, je risque de me faire virer.
2	Acc.	Jap.	萌 : 翔太と一緒に ____ のが大好き。	翔太 : そうだね、前回一緒に ____ 時はとても楽しかったな。

		Trad.	Moe : Shōta, j'adore Xer avec toi.	Shōta : Ouais. C'était vraiment marrant la dernière fois qu'on a Xé ensemble !
3	Acc.	Jap.	萌 : 翔太が次に____予定はいつなの？	翔太 : 先月もう____から、再来週かな。
		Trad.	Moe : Shōta, quand est-ce que tu as prévu de Xer de nouveau ?	Shōta : Comme j'ai déjà Xé ce mois-ci, peut-être dans deux semaines.
4	Acc.	Jap.	萌 : 翔太は今日5時に____はずじゃないの？	翔太 : そうだけど、今朝暇だったから、午前中に____よ。
		Trad.	Moe : Shōta, tu n'étais pas censé Xer à 17h aujourd'hui ?	Shōta : Si, mais comme j'avais du temps libre, j'ai Xé ce matin.
5	Acc.	Jap.	萌 : 彼が本当に____ことができるというのは嘘だと思う。	翔太 : 嘘じゃないよ。昨日、彼は僕の前で____よ。
		Trad.	Moe : Je pense qu'il nous ment sur le fait qu'il sache Xer.	Shōta : C'est pas des mensonges ! Hier, je l'ai vu Xé juste sous mes yeux.
6	Acc.	Jap.	萌 : 翔太は____の好き？	翔太 : ____ことがないから、わからないな。
		Trad.	Moe : Shōta, est-ce que tu aimes Xer ?	Shōta : Je n'ai jamais Xé, donc je ne sais pas.
7	Acc.	Jap.	萌 : 小林さんは今晚も____の？	翔太 : いや、今朝は____みたいだけど、今晚は何もしないらしい。
		Trad.	Moe : M.Kobayashi va Xer ce soir aussi ?	Shōta : Non, il a Xé ce matin, mais il paraît que ce soir il ne fera rien.
8	Acc.	Jap.	萌 : この場合、ちょっと待つかすぐに____か、どっちがいいかな？	翔太 : すぐ____ほうがいいと思う。
		Trad.	Moe : Dans ce cas, est-ce que je devrais attendre un peu ou bien Xer tout de suite ?	Shōta : Je pense que c'est mieux de Xer (acc.).
9	Acc.	Jap.	萌 : 彼は一分間で何回____ことができるの？	翔太 : 僕が彼を録音した時、一分間で八回____よ。
		Trad.	Moe : Il peut Xer combien de fois par minute ?	Shōta : Quand je l'ai enregistré, il a Xé 8 fois en une minute.
10	Acc.	Jap.	萌 : 健康のためには、毎日____ほうがいいよ。	翔太 : みんなはそう言うけど、毎日____のに風邪をひいちゃった。
		Trad.	Moe : Pour ta santé, il faut Xer tous les jours.	Shōta : Tout le monde dit ça, mais même si j'ai Xé tous les jours, j'ai quand même attrapé froid.

11	Acc.	Jap.	萌：彼女は今日何時に____の？	翔太：彼女は今日はもう____よ。
		Trad.	Moe : Elle X à quelle heure aujourd'hui ?	Shōta : Elle a déjà Xé aujourd'hui !
12	Acc.	Jap.	萌：あの人はもうすぐ____はずじゃないの？	翔太：でも、____ばかりだから、しばらくないと思う。
		Trad.	Moe : Cette personne ne devrait-elle pas Xer très prochainement ?	Shōta : Eh bien, elle vient tout juste de Xer (acc.), donc je ne crois pas qu'elle recommence de sitôt.
13	Acc.	Jap.	萌：彼女はうまく____と聞いたよ。	翔太：でも、昨日一緒にいた時、彼女はなんとなく____けど、うまくなかった。
		Trad.	Moe : J'ai entendu dire qu'elle savait bien Xer !	Shōta : Eh bien, on était ensemble hier, elle a plus ou moins Xé, mais ce n'était pas très bon.
14	Acc.	Jap.	萌：鈴木さんはよく____と思う？	翔太：今日だけで、鈴木さんは5回も____ね。
		Trad.	Moe : Tu crois que Mme Suzuki X souvent ?	Shōta : Rien qu'aujourd'hui, Mme Suzuki a Xé 5 fois !
15	Acc.	Jap.	萌：彼は初めて____らしいわ。	翔太：いや、それは違うと思う。先週、僕と一緒に____よ。
		Trad.	Moe : Il paraît que c'était la première fois qu'il Xait.	Shōta : Ce n'est pas vrai. Il a Xé avec moi la semaine dernière !
16	Acc.	Jap.	萌：____のは本当に無駄なの？	翔太：何回も____のに、何もが変わらなっただし、無駄そう。
		Trad.	Moe : Est-ce vraiment inutile de Xer ?	Shōta : Même après avoir Xé je ne sais combien de fois, ça n'a rien changé, donc ça a l'air inutile.
17	Acc.	Jap.	萌：____のは素晴らしいけど、毎日が多すぎるね。	翔太：そうだね。僕は何回も____からもう飽きたよ。
		Trad.	Moe : C'est génial de Xer, mais si c'est tous les jours, c'est trop.	Shōta : Ouais. Après avoir Xé une centaine de fois, tu dois en avoir marre.
18	Acc.	Jap.	萌：こういう時は、翔太は____方がいいんじゃない？	翔太：それは知ってるけどさあ、前に____ら社長が怒っちゃったんだよね。
		Trad.	Moe : Quand ça arrive, n'es-tu pas censé Xer ?	Shōta : Je sais bien, mais même si j'ai Xé, le patron était en colère.
19	Acc.	Jap.	萌：今から____ことにしたよ。	翔太：マジか。萌はもう____と思ってた。

		Trad.	Moe : À partir de maintenant, j'ai décidé de Xer !	Shōta : Oh vraiment ? Je croyais que tu Xais déjà tout le temps.
20	Acc.	Jap.	萌 : 明日____のはどうかな？	翔太 : 三日前____から、ちょっと早すぎると思う。
		Trad.	Moe : Ça te dirait de Xer demain ?	Shōta : Comme on a déjà Xé il y a trois jours, je pense que c'est un peu trop tôt.
21	Acc.	Jap.	萌 : 翔太は一日で何回____の？	翔太 : ええと、月曜日から一日四回ぐらい____んだ。
		Trad.	Moe : Shōta, combien de fois par jour est-ce que tu X ?	Shōta : Hmm. Depuis lundi, j'ai Xé 4 fois par jour.
22	Acc.	Jap.	萌 : 彼は最近よく____と思う。	翔太 : そうだね。彼は先月から何回か____ね。
		Trad.	Moe : Je crois qu'il X beaucoup ces derniers temps	Shōta : En effet. Depuis le mois dernier, il a Xé je ne sais combien de fois.
23	Acc.	Jap.	萌 : 翔太は普段誰と____の？	翔太 : 最近は山田さんと____よ。
		Trad.	Moe : Shōta, avec qui est-ce que tu X ?	Shōta : Ces derniers temps, je Xais avec M.Yamada.
24	Acc.	Jap.	萌 : 私はいつ____のがいいかな？	翔太 : 萌は今日の夜____ほうがいいと思う。
		Trad.	Moe : Quand est-ce que je pourrais Xer ?	Shōta : Je pense que tu pourrais Xer ce soir.
25	Acc.	Jap.	萌 : 今回も____つもりなの？	翔太 : いや、しないよ。前回____時迷惑かけちゃったからね。
		Trad.	Moe : Tu penses Xer cette fois encore ?	Shōta : Non ! La dernière fois que j'ai Xé, j'ai eu des problèmes.
26	Acc.	Jap.	萌 : 翔太は普段どこで____の？	翔太 : 色々だけど、先週末は僕の部屋で____よ。
		Trad.	Moe : Shōta, où est-ce que tu X d'habitude ?	Shōta : Ça dépend, mais le weekend dernier, j'ai Xé dans ma chambre.
27	Acc.	Jap.	萌 : 今朝は____予定があったの？	翔太 : うん。今朝は8時から9時まで____よ。
		Trad.	Moe : Tu avais prévu de Xer ce matin ?	Shōta : Oui. Ce matin j'ai Xé entre 8h et 9h.
28	Acc.	Jap.	萌 : 彼らは明日も一緒に____と思っていた。	翔太 : いや、今日たくさん____から、もういって言ってたよ。

		Trad.	Moe : Je pensais qu'ils allaient Xer ensemble demain aussi.	Shōta : Non. Ils ont dit que comme ils avaient beaucoup Xé aujourd'hui, ça suffisait.
29	Acc.	Jap.	萌 : 誰が____のが一番上手だと思う？	翔太 : 佐藤さん以外、____ことある人がいないから、佐藤さんが一番だね。
		Trad.	Moe : Qui est-ce qui X le mieux, à ton avis ?	Shōta : À part M.Satou, personne n'a jamais Xé, donc c'est lui le meilleur.
30	Acc.	Jap.	萌 : 彼はどうして____のが好きじゃないのかな？	翔太 : 前回____時にうまくいかなかったからじゃないかな。
		Trad.	Moe : Pourquoi est-ce qu'il n'aime pas Xer ?	Shōta : La dernière fois qu'il a Xé, peut-être que ça ne s'est pas bien passé.
31	Acc.	Jap.	翔太 : 萌は____ことはできるの？	萌 : 一度も____ことがないから、できないと思う。
		Trad.	Shōta : Moe, tu sais Xer ?	Moe : Je n'ai jamais Xé une seule fois, donc je pense que je ne sais pas.
32	Acc.	Jap.	翔太 : この記事曰く____のが体にいいみたい。	萌 : それなら、できるだけよく____ほうがいいよ。
		Trad.	Shōta : D'après cet article, Xer serait bon pour le corps.	Moe : Si c'est le cas, alors on devrait Xer autant que possible.
33	Acc.	Jap.	翔太 : 萌は____のを止めたの？	萌 : うん、先月は毎日____から、もう飽きたんだよね。
		Trad.	Shōta : Moe, est-ce que tu as arrêté de Xer ?	Moe : Oui. J'ai Xé tous les jours depuis le mois dernier, donc j'en ai eu marre.
34	Acc.	Jap.	翔太 : ここで____のは大丈夫かな。	萌 : 大丈夫じゃないかな。私は何回も____けど、誰も文句を言わなかったよ。
		Trad.	Shōta : Est-ce qu'on peut Xer ici ?	Moe : Je crois bien que oui. Même si j'ai Xé plein de fois ici, personne ne s'est jamais plaint.
35	Acc.	Jap.	翔太 : 田中さんは____のが好きじゃないみたいだよ。	萌 : そうだね。昨日三十分____けど、彼はそれだけでも長すぎるって言ってたよ。
		Trad.	Shōta : On dirait que M.Tanaka n'aime pas Xer.	Moe : On dirait bien. Même s'il n'a Xé que 30 minutes hier, il a dit que c'était déjà trop pour lui.
36	Acc.	Jap.	翔太 : 今度の休みは____つもりなの？	萌 : うん、前回の休みの時に____のがとても面白かったから今回もやってみたいんだよね。

		Trad.	Shōta : Tu as prévu de Xer pendant les vacances ?	Moe : Oui. C'était super bien de Xer les dernières vacances. Donc je veux refaire pareil cette fois.
37	Acc.	Jap.	翔太 : 僕たちは明日絶対____よ。	萌 : そうだね。前回____のが先月だったから、明日やるしかないね。
		Trad.	Shōta : On doit absolument Xer demain.	Moe : Oui. La dernière fois qu'on a Xé c'était le mois dernier, alors il n'y a pas le choix.
38	Acc.	Jap.	翔太 : 山田さんは来週も____はずだね。	萌 : ちがうよ。山田さんは先週____んだ。来週は鈴木さんだね。
		Trad.	Shōta : M.Tanaka est censé Xer la semaine prochaine, n'est-ce pas ?	Moe : Non. M. Tanaka a Xé la semaine dernière. La semaine prochaine c'est M.Suzuki.
39	Acc.	Jap.	翔太 : この部屋では____のが禁止されてるよ。	萌 : あ！ そうなの！ でも、私は隣の部屋で____から大丈夫。
		Trad.	Shōta : Il est interdit de Xer dans cette pièce.	Moe : Ah ! Vraiment ? Mais c'est dans la pièce d'à côté que j'ai Xé, donc pas de problème.
40	Acc.	Jap.	翔太 : ちゃんと____ためには、練習が必要だ。	萌 : そうだね。でも私は子供の頃よく____から、今も結構できるはず。
		Trad.	Shōta : Pour Xer correctement, il faut s'entraîner.	Moe : C'est vrai. Mais comme j'ai beaucoup Xé quand j'étais petit, même aujourd'hui je m'en sors pas mal.
41	Acc.	Jap.	翔太 : 萌は今日____予定はあるの？	萌 : いや、昨日____からもうしないよ。
		Trad.	Shōta : Moe, tu as prévu de Xer aujourd'hui ?	Moe : Non. J'ai déjà Xé hier. Ça suffit.
42	Acc.	Jap.	翔太 : 彼は____と、何か起こるの？	萌 : まー。彼は前回一度____けど、何も起こらなかったなあ。
		Trad.	Shōta : Que se passe-t-il lorsqu'il X ?	Moe : Eh bien, la dernière fois, il a Xé une fois, et il ne s'est rien passé...
43	Acc.	Jap.	翔太 : 彼女はあの時から、毎日____の？	萌 : いや、あの時から、三回____けど、結局やめちゃったよ。
		Trad.	Shōta : Elle X tous les jours depuis la dernière fois ?	Moe : Non, depuis la dernière fois, elle a Xé trois fois, mais après elle a arrêté.
44	Acc.	Jap.	翔太 : 僕は午後 2 時から 4 時まで____必要があるんだよね。	萌 : 大変だね。翔太は今朝も一生懸命____のに。

		Trad.	Shōta : Je dois Xer de 14h à 16h aujourd'hui.	Moe : C'est dur ! Alors que tu as déjà Xé de toutes tes forces ce matin...
45	Acc.	Jap.	翔太 : これを____のには何が必要なの？	萌 : 前回____時は特別な道具を使ったよ。
		Trad.	Shōta : Tu as besoin de quoi pour Xer ça ?	Moe : La dernière fois que j'ai Xé ça, j'ai utilisé un outil particulier.
46	Acc.	Jap.	翔太 : 萌は最近よく____の？	萌 : うん。昨日だけでも十二回も____よ。
		Trad.	Shōta : Moe, est-ce tu X souvent ces derniers temps ?	Moe : Oui. Rien que hier, j'ai Xé 12 fois.
47	Acc.	Jap.	翔太 : 萌は今日山田さんと____の？	萌 : いや、山田さんとは昨日____よ。今日は鈴木さんとだね。
		Trad.	Shōta : Moe, est-ce que tu X avec M.Yamada aujourd'hui ?	Moe : Non. J'ai Xé avec M.Yamada hier. Aujourd'hui c'est le tour de M.Suzuki.
48	Acc.	Jap.	翔太 : 彼女は三時間____のができると思う？	萌 : 絶対できるよ。昨日、彼女は五時間も____からね。
		Trad.	Shōta : Tu crois qu'elle peut Xer pendant 3 heures ?	Moe : Sans l'ombre d'un doute ! Hier, elle a Xé pendant 5 heures.
49	Acc.	Jap.	翔太 : よく____人は知ってる？	萌 : ええと。佐藤さんは____ことがあると思うけど、一回だけだと思う。
		Trad.	Shōta : Tu connais quelqu'un qui X souvent ?	Moe : Hmm. M.Satou a déjà Xé, mais je crois que c'était une seule fois.
50	Acc.	Jap.	翔太 : 萌の弟は____のが上手なの？	萌 : うん。十五歳になってからたくさん____から、結構上手になったよ。
		Trad.	Shōta : Moe, est-ce ton petit frère est bon pour Xer ?	Moe : Ouais. Depuis qu'il a 15 ans, il a beaucoup Xé. Il est devenu plutôt bon.
51	Acc.	Jap.	翔太 : 今朝の____練習はどうだった？	萌 : よかったよ。____後でお兄さんと遊んだし、楽しい朝だったね。
		Trad.	Shōta : C'était comment l'entraînement à Xer ce matin ?	Moe : Ça s'est bien passé. En plus, après avoir Xé, j'ai joué avec mon grand frère, c'était vraiment une bonne matinée.
52	Acc.	Jap.	翔太 : この状況では____しかない。	萌 : それはそうだけど、____後はどうすればいい？
		Trad.	Shōta : Dans cette situation, tu dois Xer.	Moe : C'est vrai. Mais après avoir Xé, qu'est-ce que je devrais faire ?

53	Acc.	Jap.	翔太：萌は今日何時に____の？	萌：今____ばかりだよ。
		Trad.	Shōta : Moe, quand est-ce que tu X aujourd'hui ?	Moe : Je viens juste de Xer.
54	Acc.	Jap.	翔太：これを____と、何かが起きるかも。	萌：そう思ったけど、前にこれを____のにも効き目がなかったんだよね。
		Trad.	Shōta : Si tu X ça, peut-être qu'il va se passer quelque chose.	Moe : Je pensais aussi, mais quand j'ai Xé ça tout à l'heure, ça n'a eu aucun effet.
55	Acc.	Jap.	翔太：このテストを受ける時は____ことは禁止されているよ。	萌：そうだね。去年____人がすぐに0点になったって聞いたなあ。
		Trad.	Shōta : Il est interdit de Xer quand on passe cet examen.	Moe : Effectivement. J'ai entendu dire que les personnes qui ont Xé l'an dernier ont reçu un 0 immédiatement.
56	Acc.	Jap.	翔太：彼女は犬と____のが好きなの？	萌：前は猫とも____ことがあるみたいだけど、今は犬としかしないね。
		Trad.	Shōta : est-ce qu'elle aime Xer avec les chiens ?	Moe : Elle a déjà Xé avec des chats auparavant, mais maintenant elle ne le fait plus qu'avec les chiens.
57	Acc.	Jap.	翔太：疲れているから、私は____ともう何もできないや。	萌：そうだね、私も昨日____後すぐに寝ちゃったよ。
		Trad.	Shōta : Comme c'est fatiguant, après avoir Xé, je ne peux plus rien faire.	Moe : C'est bien vrai. Hier, après avoir Xé, je suis directement allé au lit.
58	Acc.	Jap.	翔太：萌は今日午後8時に____って聞いたよ。	萌：いや、今日は午前8時にもう____よ。夜は暇だよ。
		Trad.	Shōta : Moe, j'ai entendu dire que tu Xais à 20h aujourd'hui !	Moe : Mais non. J'ai déjà Xé à 8h, ce soir je suis libre.
59	Acc.	Jap.	翔太：今年は絶対に毎日____と決めたよ。	萌：私は去年毎日____けどそんなに難しくないよ。翔太ならきっとできると思う。
		Trad.	Shōta : Cette année, j'ai décidé de Xer absolument tous les jours.	Moe : J'ai Xé tous les jours l'an dernier. Ce n'est pas si difficile. Je suis sûr que tu peux le faire.
60	Acc.	Jap.	翔太：明日は誰が____の？	萌：ええと。私は二日前____から、明日は鈴木さんの番だね。
		Trad.	Shōta : Qui est-ce qui X demain ?	Moe : Hmm. J'ai Xé il y a deux jours, donc demain, c'est le tour de M.Suzuki.

61	Nég.	Jap.	萌：翔太は毎日____の？	翔太：そんなことない、週末しか____よ。
		Trad.	Moe : Shōta, est-ce que tu X tous les jours ?	Shōta : Non. Je ne X pas en dehors des weekend.
62	Nég.	Jap.	萌：鈴木さんは毎週金曜日____よね？	翔太：いや、鈴木さんは金曜日は____よ。火曜日だと思う。
		Trad.	Moe : M.Suzuki X tous les vendredi, c'est ça ?	Shōta : Non. M.Suzuki ne X pas le vendredi. Je crois que c'est le mardi.
63	Nég.	Jap.	萌：みんな毎日____よね？	翔太：いや、あまり____人もいるんじゃないかな。
		Trad.	Moe : Tout le monde X tous les jours n'est-ce pas ?	Shōta : Malheureusement, certains jours, beaucoup de personnes ne X pas.
64	Nég.	Jap.	萌：彼が____かどうか知ってる？	翔太：彼は____らしいよ。
		Trad.	Moe : Est-ce tu sais s'il X ou pas ?	Shōta : On m'a dit qu'il ne Xait pas
65	Nég.	Jap.	萌：彼女がそんなによく____なんて信じられないね。	翔太：確かに嘘かもなあ。彼女はそんなによく____と思う。
		Trad.	Moe : Je n'arrive pas à croire qu'elle X si souvent !	Shōta : Effectivement. C'est peut-être un mensonge. Je pense qu'il ne X pas aussi souvent.
66	Nég.	Jap.	萌：この仕事には、よく____人が必要だね？	翔太：うん、あまり____人は絶対にだめだね。
		Trad.	Moe : Pour ce travail, on a besoin de personnes qui X souvent, pas vrai ?	Shōta : Oui. Les personnes qui ne X pas tellement ne sont d'aucune utilité.
67	Nég.	Jap.	萌：翔太の友達で____人はいないの？	翔太：僕が知る限りは、誰も____なあ。
		Trad.	Moe : Shōta, est-ce que tu n'aurais pas des amis qui X ?	Shōta : D'après ce que je sais, personne ne X.
68	Nég.	Jap.	萌：渡辺さんの妻は____のかな？	翔太：もちろんだね。渡辺さんは____人が嫌いなんだ。
		Trad.	Moe : Tu penses que la femme de M.Watanabe X ?	Shōta : Bien sûr. M.Watanabe déteste les gens qui ne X pas.
69	Nég.	Jap.	萌：翔太は明日____の？	翔太：ううん。明日は____かな。

		Trad.	Moe : Shōta, est-ce que tu vas Xer demain ?	Shōta : Non, je ne pense pas Xer demain.
70	Nég.	Jap.	萌 : 翔太の部屋で何か____の？	翔太 : いや、僕の部屋では何も____よ。
		Trad.	Moe : Shōta, est-ce qu'il y a quelque chose qui X dans ta chambre ?	Shōta : Non, il n'y a rien qui X dans ma chambre !
71	Nég.	Jap.	萌 : 私が____といつも悪いことが起こっちゃうんだよね。	翔太 : もしかしたら、もう____ほうがいいのかもね。
		Trad.	Moe : À chaque fois que je X, quelque chose de mal se produit.	Shōta : C'est peut-être préférable de ne plus Xer.
72	Nég.	Jap.	萌 : これを____ためにはどうすればいいのかな？	翔太 : いや、もうやめた方がいいんじゃないかな。____に越したことはない。
		Trad.	Moe : Pour Xer ça, qu'est-ce que je dois faire ?	Shōta : Tu ferais mieux d'arrêter. C'est mieux de ne pas le Xer.
73	Nég.	Jap.	萌 : 翔太は来週も____つもりなの？	翔太 : もういやだ。僕は二度と____よ。
		Trad.	Moe : Shōta, tu as prévu de Xer la semaine prochaine aussi ?	Shōta : Non, je déteste ça. Jamais plus je ne Xerai.
74	Nég.	Jap.	萌 : 翔太は週末によく____って聞いたんだけど。	翔太 : 週末にはあまり____なあ。大好きだけど、平日だけだよ。
		Trad.	Moe : J'ai entendu dire que tu Xais souvent le weekend.	Shōta : Je ne X pas souvent le weekend. Même si j'adore ça, j'essaye de limiter à la semaine.
75	Nég.	Jap.	萌 : 翔太のお母さんって本当に____の？	翔太 : ううん、母は全然____よ。むしろ父だよ。
		Trad.	Moe : Shōta, est-ce ta mère X vraiment ?	Shōta : Non, ma mère ne X pas du tout. Mais mon père si !
76	Nég.	Jap.	萌 : 明日はいつ____の？	翔太 : 明日は____んだ。明後日の10時だよ。
		Trad.	Moe : Quand X tu demain ?	Shōta : Je ne X pas demain. C'est après-demain à 10h.
77	Nég.	Jap.	萌 : 翔太はだれと____のがいいと思う？	翔太 : 田中さんとは____ほうがいいと思うけど、彼以外なら誰でもいいよ。
		Trad.	Moe : Shōta, avec qui penses-tu qu'il soit préférable de Xer ?	Shōta : Je pense que c'est mieux de Xer avec M.Tanaka. Mais sinon, n'importe qui fera l'affaire.

78	Nég.	Jap.	萌：翔太は毎日____の？	翔太：いや、全然____日もあるよ。
		Trad.	Moe : Shōta, est-ce que tu X tous les jours ?	Shōta : Non, il y a des jours où je ne X pas du tout.
79	Nég.	Jap.	萌：今からできるだけよく____と決めたよ。	翔太：いいね。僕はあまり____人は好きじゃないな。
		Trad.	Moe : À partir de maintenant, j'ai décidé de Xer autant que possible.	Shōta : Tant mieux. Je n'aime pas les gens qui ne X pas trop.
80	Nég.	Jap.	萌：彼はどうして一人で____の？	翔太：彼の友達は____から、仕方がないよ。
		Trad.	Moe : Pourquoi est-ce qu'il X tout seul ?	Shōta : Vu que ses amis ne X pas, il n'a pas le choix.
81	Nég.	Jap.	萌：この俳優はすべての映画でよく____と気づいたんだけど。	翔太：確かに。でも実生活ではあまり____らしいよ。
		Trad.	Moe : J'ai remarqué que cet acteur Xait bien dans tous les films.	Shōta : C'est vrai, mais il paraît qu'il ne X pas souvent dans la vraie vie.
82	Nég.	Jap.	萌：このゲームでは、____のがいい作戦だね。	翔太：そうだね。ビギナーの人はあまり____から負けやすいんだよね。
		Trad.	Moe : Dans ce jeu, c'est une bonne stratégie de Xer.	Shōta : En effet. Beaucoup de débutants perdent parce qu'ils ne X pas.
83	Nég.	Jap.	萌：あの会社員は____はずだよね？	翔太：そうだね。____会社員はすぐに首になるみたいだよ。
		Trad.	Moe : Ces employés sont censés Xer, n'est-ce pas ?	Shōta : Oui. Il semblerait que les employés que ne X pas se font renvoyer immédiatement.
84	Nég.	Jap.	萌：____ものは左の箱に置こう。	翔太：うん。じゃあ、____ものは右の箱なんだね？
		Trad.	Moe : Mettons les objets qui X dans la boîte de gauche.	Shōta : D'accord. Alors les objets qui ne X pas vont dans la boîte de droite, c'est ça ?
85	Nég.	Jap.	萌：翔太のクラスでは誰が____の？	翔太：ええと、うちのクラスだと僕しか____と思う。
		Trad.	Moe : Shōta, Qui est-ce qui X dans ta classe ?	Shōta : Hmm. Je crois que personne ne X à part moi.
86	Nég.	Jap.	萌：彼は____のが上手だって言ってたよ。	翔太：うそー！？彼は全然____よ。

		Trad.	Moe : Il a dit qu'il savait bien Xer.	Shōta : C'est faux ! Il ne X pas du tout !
87	Nég.	Jap.	萌 : ____許可が下りなかったら、どうすればいいかな？	翔太 : その場合は、 ____方がいいね。
		Trad.	Moe : Que va-t-on faire si on n'a pas l'autorisation de Xer ?	Shōta : Dans ce cas, il vaudra mieux ne pas Xer.
88	Nég.	Jap.	萌 : 鈴木さんは毎週よく ____みたいだよ。	翔太 : らしいね。でも週末ではあまり ____みたいだよ。
		Trad.	Moe : On dirait que M.Suzuki X beaucoup chaque semaine.	Shōta : On dirait bien. Mais il paraît qu'il ne X pas trop le weekend.
89	Nég.	Jap.	萌 : 田中先生は ____学生が好きみたいだよ。	翔太 : ああ、確かに。 ____学生にはとても厳しいしね。
		Trad.	Moe : On dirait que M.Tanaka apprécie les étudiant qui X.	Shōta : Oh, c'est sûr. Il est très sévère avec les étudiants qui ne X pas.
90	Nég.	Jap.	萌 : この本は ____子供に向いていると思う。	翔太 : 確かに。でも ____子供にも読みやすいんじゃないかな。
		Trad.	Moe : Je pensais que ce livre était fait pour les enfants qui X.	Shōta : C'est vrai. Mais il est aussi facile à lire pour les enfants qui ne X pas.
91	Nég.	Jap.	翔太 : 萌はなんで ____の？	萌 : いやあ、だって ____人はよく病気になるって聞いたよね。
		Trad.	Shōta : Moe, pourquoi est-ce que tu X ?	Moe : C'est parce que j'ai entendu dire que les personnes qui ne X pas sont souvent malades.
92	Nég.	Jap.	翔太 : これは ____と問題になりそうじゃないかな？	萌 : そうなんだけど、さすがに絶対 ____はずだと思う。
		Trad.	Shōta : Ne serait-ce pas un problème si ceci venait à Xer ?	Moe : En effet. Mais c'est prévu pour n'absolument jamais Xer.
93	Nég.	Jap.	翔太 : 僕は ____のはあまり好きじゃないなあ。	萌 : じゃ、できるだけ ____ようにしようか。
		Trad.	Shōta : Je n'aime pas vraiment Xer.	Moe : Bon, alors faisons de notre mieux pour ne pas Xer.
94	Nég.	Jap.	翔太 : 山田さんは一日に何回も ____けど、一向に上達しないんだよね。	萌 : 正しく ____からじゃない？
		Trad.	Shōta : Peu importe combien de fois M.Yamada X chaque jour, il ne progresse pas du tout.	Moe : N'est-ce pas parce qu'il ne X pas correctement ?

95	Nég.	Jap.	翔太：この仕事では、ちょっとだけしか____のができない人は本当にだめなの？	萌：ちゃんと____人はこの仕事では全然役に立たないから、だめだと思う。
		Trad.	Shōta : Pour ce travail, les personnes qui ne peuvent que Xer un petit peu ne sont vraiment pas acceptées ?	Moe : Les personnes qui ne X pas correctement ne sont d'aucune utilité, donc non.
96	Nég.	Jap.	翔太：萌はどんなところで____の？	萌：図書館では____けど、そこ以外ならどこでもいいかな。
		Trad.	Shōta : Moe, dans quel genre d'endroit X tu ?	Moe : Je ne X pas dans les bibliothèques, mais à part ça, je peux le faire n'importe où.
97	Nég.	Jap.	翔太：これを____と、なんか変な音が出るよ。	萌：だから____ほうがいいって言ったじゃん。
		Trad.	Shōta : Quand je X ceci, ça fait un son bizarre.	Moe : C'est pour ça que je t'ai dit de ne pas le Xer.
98	Nég.	Jap.	翔太：萌は本当に毎朝____の？	萌：うん。もう習慣になってるから、____朝はあまりないね。
		Trad.	Shōta : Moe, tu X vraiment tous les jours ?	Moe : Oui. Depuis que c'est devenu une habitude, c'est rare que je ne X pas le matin.
99	Nég.	Jap.	翔太：今すぐ____のは本当にだめなの？	萌：準備が整わない限りは____ほうがいいかな。
		Trad.	Shōta : N'est-ce vraiment pas possible de X tout de suite ?	Moe : Il vaut mieux ne pas Xer tant que les préparatifs ne sont pas achevés.
100	Nég.	Jap.	翔太：時々____人はもっと優しいと思う。	萌：そうだね。ぜんぜん____人は厳しそうだね。
		Trad.	Shōta : Je pense que les gens qui X de temps en temps sont sympas.	Moe : C'est vrai ! Les gens qui ne X pas du tout ont l'air stricte.
101	Nég.	Jap.	翔太：____ペットが最高だ。	萌：そう？私は____ペットも好きだよ。
		Trad.	Shōta : Les animaux de compagnie qui X sont les meilleurs.	Moe : Tu crois ? J'aime bien les animaux de compagnie qui ne X pas aussi !
102	Nég.	Jap.	翔太：朝か夜か、____にはどっちが一番いい時間だと思う？	萌：私は全然____からよく分からないな。
		Trad.	Shōta : Matin ou soir, lequel est le meilleur moment pour Xer, à ton avis ?	Moe : Vu que je ne X jamais, je m'en fiche.

103	Nég.	Jap.	翔太：他の多くの動物も____のことができると思ったんだけど。	萌：ううん。人間しか____んだよね。
		Trad.	Shōta : Je pensais qu'un tas d'autres animaux pouvaient Xer.	Moe : Non. Seuls les humains X.
104	Nég.	Jap.	翔太：萌は本当に月曜日に____の？	萌：ううん、冗談。月曜日は____よ。
		Trad.	Shōta : Moe, tu vas vraiment Xer lundi ?	Moe : Non, c'était une blague. Je ne vais pas Xer lundi !
105	Nég.	Jap.	翔太：その国では皆は何歳の人でも____って聞いたんだけど。	萌：そんなことないと思う。その国でも大人しか____よ。
		Trad.	Shōta : J'ai entendu dire que tout le monde, de tous les ages, Xait dans ce pays.	Moe : Ce n'est pas vrai. Même dans ce pays, seuls les adultes X.
106	Nég.	Jap.	翔太：彼はいつも____みたいだね。	萌：うん。彼が____時はほとんどないね。
		Trad.	Shōta : On dirait qu'il X tout le temps.	Moe : Ouais. Les moments où il ne X pas sont vraiment rares.
107	Nég.	Jap.	翔太：萌は来週も____つもりなの？	萌：ううん。これからはもう____ほうがいいと思う。
		Trad.	Shōta : Moe, tu prévois de Xer la semaine prochaine aussi ?	Moe : Non. À partir de maintenant, je pense qu'il vaut mieux ne plus Xer.
108	Nég.	Jap.	翔太：次回、彼女は本当に____はずじゃないかな。	萌：いや、何が何でも彼女は____と思う。
		Trad.	Shōta : La prochaine fois, il faudrait vraiment qu'elle X.	Moe : Quoi qu'on fasse, je pense qu'elle ne Xera pas.
109	Nég.	Jap.	翔太：萌は本当に____のをやめるの？	萌：うん。もう____って約束したよ。
		Trad.	Shōta : Moe, tu vas vraiment arrêter de Xer ?	Moe : Oui. J'ai juré de ne plus Xer.
110	Nég.	Jap.	翔太：あのグループでは____人がいると思う？	萌：ううん。誰も____と思う。
		Trad.	Shōta : Tu crois qu'il y a des personnes qui X dans ce groupe ?	Moe : Non. Je pense que personne ne X.
111	Nég.	Jap.	翔太：これを____のは本当に必要なの？	萌：うん。これを____限りは進めないんだよね。

		Trad.	Shōta : C'est vraiment indispensable de Xer ceci ?	Moe : Oui. On ne peut pas progresser tant qu'on n'a pas Xé ceci.
112	Nég.	Jap.	翔太 : 萌は明日____予定があったんだね。	萌 : そうだったんだけど、予定が変わったから、結局明日は____かな。
		Trad.	Shōta : Moe, tu avais prévu de Xer demain, n'est-ce pas ?	Moe : Oui, mais mes plans ont changé. Je ne Xerai pas demain finalement.
113	Nég.	Jap.	翔太 : 萌は____歌手しか好きじゃないんだと思ってた。	萌 : いや、____歌手も好きだよ。
		Trad.	Shōta : Je croyais que tu n'aimais que les chanteurs qui X.	Moe : Par vraiment. J'aime aussi les chanteurs qui ne X pas.
114	Nég.	Jap.	翔太 : この森には____動物しかいないの？	萌 : ううん。____動物もいるよ。
		Trad.	Shōta : Dans cette forêt, il y a seulement des animaux qui X ?	Moe : Non. Il y a aussi des animaux qui ne X pas !
115	Nég.	Jap.	翔太 : 萌は来週毎日____から、結構忙しいね。	萌 : そうだね。でも次の週は全然____から暇だよ。
		Trad.	Shōta : Moe, puisque tu X tous les jours de la semaine prochaine, tu seras plutôt occupée.	Moe : Oui. Mais comme je ne Xerai pas beaucoup la semaine suivante, j'aurai du temps libre.
116	Nég.	Jap.	翔太 : いつも____からとても疲れているんだね。	萌 : そうだね。体に大変だから、たまには____ほうがいいね。
		Trad.	Shōta : N'est-ce pas fatiguant de Xer tout le temps ?	Moe : En effet. Comme c'est dur pour le corps, il est préférable de pas Xer de temps en temps.
117	Nég.	Jap.	翔太 : 萌の友達をよく____の。	萌 : 残念だけど、私の友達はあまり____んだよね。
		Trad.	Shōta : Est-ce que ton ami X beaucoup ?	Moe : Malheureusement, mon ami ne X pas beaucoup.
118	Nég.	Jap.	翔太 : 明日は誰と____の。	萌 : 明日は一人だから、誰とも____よ。
		Trad.	Shōta : Avec qui est-ce que tu X demain ?	Moe : Comme je serai seule, je ne Xerai avec personne.
119	Nég.	Jap.	翔太 : この調査曰く、嘘をつく時多くの人は気づかないうちに____らしい。	萌 : 確かに。嘘をつかない人は全然____よね。
		Trad.	Shōta : D'après cette étude, beaucoup de personnes X inconsciemment lorsqu'ils mentent.	Moe : Effectivement. Les gens qui disent la vérité ne X pas du tout !

120	Nég.	Jap.	翔太 : 萌の兄弟も ____ のは上手なの。	萌 : ええと、弟はあまり ____ から、上手だとは言えないんだけど、姉は本当にすごいよ。
		Trad.	Shōta : Moe, est-ce que tes frères et soeurs sont aussi bon à Xer ?	Moe : Eh bien, mon petit frère ne X pas souvent, donc je ne peux pas dire qu'il soit bon. Mais ma grande sœur est vraiment incroyable !