Zeitschrift: Bulletin suisse de linguistique appliquée / VALS-ASLA

Herausgeber: Vereinigung für Angewandte Linguistik in der Schweiz = Association suisse de linguistique appliquée

Band: - (2015)

Heft: 102: L'apprentissage de la liaison en français par des locuteurs non natifs : éclairage des corpus oraux = French liaison learning by non-native speakers in the light of oral corpora = Das Erlernen der französischen Liaison durch Nicht-Muttersprachler im Lichte der mündlichen Korpora = L'apprendimento della liaison in francese come lingua straniera alla luce dei corpora orali

Artikel: Corpus oraux, liaison et locuteurs non natifs : de la recherche en phonologie à l'enseignement du français langue étrangère

Autor: Racine, Isabelle / Detey, Sylvain

DOI: https://doi.org/10.5169/seals-978605

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 09.02.2026

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Corpus oraux, liaison et locuteurs non natifs: de la recherche en phonologie à l'enseignement du français langue étrangère

Isabelle RACINE
Université de Genève, Ecole de langue et de civilisation françaises
UNI-Bastions, 5, rue de Candolle, 1211 Genève 4, Suisse
isabelle.racine@unige.ch

Sylvain DETEY
Université Waseda, School of International Liberal Studies
1-6-1 Nishi-Waseda, Shinjuku-ku, Tokyo 169-8050, Japon
detey@waseda.jp

This paper focuses on the acquisition of L2 French liaison. First we present the challenges that liaison involves in L1 and in L2. Then we introduce the methodological framework in which the seven studies included in the present volume have been carried out: the "Interphonologie du français contemporain" project (IPFC), which stands as the non-native part of the PFC project ("Phonologie du français contemporain"), a large sociophonological survey of the pronunciation of French in the French-speaking world. IPFC aims at compiling and analyzing a large database of oral French produced by learners with different L1s on the basis of a single survey protocol. We briefly present the methodology used in the project and we focus on L2 liaison, with a description of the alphanumeric code designed for its treatment. This common methodology has been applied to all of the IPFC data, as is illustrated in the seven papers presented in this volume. This work will enrich our knowledge of L2 liaison and should contribute in the long term to a renewal of the pedagogical material used for teaching L2 French.

Keywords:
liaison, consonant linking, external sandhi, French as Foreign Language, oral French, learner corpora, InterPhonologie du Français Contemporain (IPFC), L2 phonological acquisition.

1. Introduction[1]

Si l'on se place du point de vue d'un apprenant de français langue étrangère (ci-après FLE), on peut aisément imaginer la difficulté qu'il peut rencontrer à rattacher une réalisation comme [ʃɥi-al-mɑ̃][2] à la séquence constituée de trois

[1] Nous tenons à remercier Jacques Durand, Julien Eychenne, Chantal Lyche, Bernard Laks, Alain Kamber, Yuji Kawaguchi, Roberto Paternostro et Françoise Zay, qui ont contribué, par leurs commentaires et suggestions, à faire progresser notre réflexion sur la liaison. Le projet IPFC a bénéficié du soutien, en Suisse, du Fonds national suisse de la recherche scientifique (no 132144, resp. I. Racine), de la Faculté des lettres de l'Université de Genève ainsi que de la Société académique de Genève et, au Japon, de la Société Japonaise pour la Promotion de la Science (Grants-in-Aid no 23320121 et 15H03227, resp. S. Detey) et de l'Université Waseda (Special Research Grant, 2011B-297). Nous remercions également les apprenants qui ont participé à ce projet, ainsi que tous les membres d'IPFC pour leur implication et leur enthousiasme constant.

[2] Les tirets sont utilisés pour indiquer le découpage syllabique.

© 2015 Centre de linguistique appliquée
Université de Neuchâtel

mots "je" + "suis" + "allemand", la première fois qu'il y est confronté. Un certain nombre de phénomènes (chute du schwa[3] dans le pronom "je" et assimilation qui s'ensuit, non-réalisation de la liaison entre "suis" et "allemand") contribuent à éloigner cette séquence de sa forme – que l'on pourrait qualifier de – "canonique", [ʒə-sɥi-(z)almɑ̃] et qui est probablement celle qu'un apprenant aura apprise en classe de FLE, du moins en début d'apprentissage (voir Paternostro 2014: 6).

La liaison, phénomène fréquent en français puisqu'intervenant, selon Boë & Tubach (1992), à peu près une fois tous les 16 mots, constitue donc – au même titre que le schwa – un passage incontournable dans l'enseignement de l'oral en FLE. Une section spécifique est d'ailleurs généralement consacrée à chacun d'entre eux dans les manuels de prononciation (cf. par exemple Abry & Chalaron 1994, 2011; Léon 2003; Charliac & Motron 2006). En ce qui concerne la liaison, ces ouvrages se focalisent souvent sur un seul aspect des difficultés liées à son acquisition/apprentissage[4], celui qui touche le niveau de la *macro-planification* (Racine 2014), c'est-à-dire les problèmes posés par la variation inhérente à la liaison et qui consistent à identifier le contexte – sur la base de facteurs phonologiques, prosodiques, lexicaux, (morpho)syntaxiques et sociolinguistiques (Durand & Lyche à paraître) – afin de déterminer si la liaison doit – ou peut – être réalisée ou non. La deuxième source de difficultés, liée au fait même de réaliser une liaison, et qui se situe donc au niveau de la *micro-planification* (Racine 2014), est généralement traitée de manière beaucoup moins explicite dans les ressources pédagogiques. Elle concerne la nature de la consonne de liaison à réaliser (en lien avec la graphie), le placement de cette consonne dans la structure syllabique (avec ou sans enchaînement) ainsi que les modifications éventuelles de l'environnement immédiat (ouverture/dénasalisation de la voyelle précédente). Nous reviendrons sur ces difficultés plus en détail ultérieurement et les illlustrerons par des exemples. A cela s'ajoute encore – et ce dernier point ne concerne pas que les apprenants mais également les enseignants, d'autant plus s'ils sont natifs – le fait que les mécanismes d'appropriation de la liaison ne sont pas identiques chez les natifs et les non natifs (Wauquier 2009; Harnois-Delpiano, Cavalla & Chevrot 2012; Wauquier & Shoemaker 2013). Si l'apprentissage est généralement implicite et détaché de la graphie en L1, son mode d'appropriation est beaucoup plus explicite en L2 et se fait

---

3 Plus souvent appelé "E caduc", "E muet" ou encore "E instable" dans le domaine du FLE (pour les questions de terminologie, voir Racine 2008).

4 On emploie généralement le terme "acquisition" lorsqu'il s'agit de L1 ou de mode d'appropriation purement implicite, et le terme "apprentissage" lorsqu'il s'agit de langue non première (ci-après L2) ou de mode d'appropriation plus explicite.

généralement à partir des formes écrites, ou du moins avec l'apport de la graphie[5].

On comprend donc aisément que l'enseignement de la liaison constitue un véritable enjeu pour le FLE et que de nombreux travaux se soient intéressés aux productions de liaisons chez les apprenants. Toutefois, comme c'est souvent le cas dans le domaine de l'acquisition/apprentissage phonologique en L2 (Gut 2009), et comme le relève Wauquier (2009), la grande majorité des travaux se sont basés sur des productions d'apprenants anglophones (Mastromonaco 1999; Thomas 2002, 2010; Howard 2005, 2013; Shoemaker 2009; De Moras 2011), ou du moins d'une langue germanique (néerlandais: Matter 1986 et suédois: Stridfeldt 2005), à l'exception du travail de Harnois-Delpiano et al. (2012) portant sur des apprenants coréens. Un élargissement des populations étudiées semble donc bienvenu, notamment en raison du fait que, dans les langues germaniques, la congruence entre frontières syllabiques et lexicales est très forte, et que la resyllabation issue de la liaison, qui masque la frontière lexicale entre deux mots graphiques en français, est donc susceptible d'expliquer les difficultés des apprenants. Ce volume, consacré à la liaison en L2, présente des travaux portant sur sept différentes populations d'apprenants (dans l'ordre du volume: italophones, germanophones, anglophones, hellénophones chypriotes, norvégophones, japonophones et hispanophones).

L'objectif de cet article, qui ouvre le volume, est donc double: il vise à la fois à présenter la problématique ainsi que les enjeux liés à la question de la liaison en L1 (section 2) et en L2 (section 3) et à exposer le cadre dans lequel l'ensemble des travaux qui constituent le présent ouvrage ont été réalisés, à savoir le projet "InterPhonologie du Français Contemporain" (Detey & Kawaguchi 2008; Detey & Racine 2012; Racine, Detey, Zay & Kawaguchi 2012; Detey, Racine, Kawaguchi & Zay à paraître, Racine & Detey à paraître; ci-après IPFC). Nous présenterons brièvement le projet (section 4), puis le traitement de la liaison qui y est proposé (section 5).

2. Le point de référence: la liaison en L1

2.1 Définition

La liaison est, comme le rappellent Durand & Lyche (à paraître), un phénomène de sandhi externe qui implique la présence – ou l'absence – d'une consonne entre deux mots produits conjointement (ex. "les éléphants" [lezelefɑ̃]), le premier mot étant appelé "mot liaisonnant" ou "mot-1" (*vs* mot-2 pour le mot qui suit). Cette consonne n'est réalisée ni en position de coda du mot-1, ni en attaque du mot-2 lorsque les deux unités lexicales sont produites

[5] Voir également Racine (ce volume) pour un développement de la question du mode d'appropriation de la liaison en L1 et en L2.

de manière isolée (ex. "les" [le] et "éléphants" [elefɑ̃]) ou lorsque le mot-2 commence par une consonne (ex. "les souris" [lesuʁi]). En outre, lorsque la liaison est réalisée, la consonne se rattache dans la plupart des cas en position d'attaque du mot-2 (ex. [le-ze-le-fɑ̃]), on parle alors de liaison enchaînée, par opposition à une liaison non enchaînée ou sans enchaînement (ex. [lez-e-le-fɑ̃]). Ainsi, lorsqu'elle est enchaînée, la liaison, qui reflète un état ancien de la langue où toutes les consonnes étaient prononcées (pour un historique, voir Mallet 2008), contribue à renforcer la tendance du français à la syllabation ouverte[6] et sa prédilection pour les syllabes CV (Delattre 1946).

Selon la littérature, les consonnes de liaison possibles sont au nombre de huit ([z, n, t, ʁ, p, v, k, g]). Ce nombre est toutefois soumis à la variation diatopique, les variétés de français du Canada connaissant également des liaisons en [l] (Côté 2012). Comme le montrent Eychenne, Lyche, Durand & Coquillon (2014: 43), [z], [n] et [t] couvrent la plus grande majorité des occurrences de leur corpus, avec respectivement 46.26%, 36.06% et 17.25%, contre moins de 0.5% pour [ʁ] (0.18%) et [p] (0.06%), les trois autres consonnes, [v, k, g], n'apparaissant pas dans leurs données.

On peut souligner encore que la liaison peut aussi avoir un impact sur la voyelle qui précède la consonne de liaison. Celle-ci peut, sous l'effet de la liaison, s'ouvrir (ex. "dernier exemple" sera ainsi produit [dɛʁnjɛʁɛgzɑ̃pl], rendant la forme masculine homophone avec sa contrepartie féminine "dernière") ou se dénasaliser (ex. "bon appétit" sera ainsi produit [bɔnapeti], avec une forme masculine dont la prononciation est identique à celle du féminin "bonne"). Ces modifications (dénasalisation et ouverture) ne sont cependant pas systématiques (ex. dans "son élève" [sɔ̃nelɛv], la voyelle reste nasale, de même que dans "en outre" [ɑ̃nutʁ] ou "un enfant" [ɛ̃nɑ̃fɑ̃]). Elles sont par ailleurs soumises à la variation diatopique, puisque la dénasalisation est notamment plus fréquente, par exemple, dans le français méridional (cf. Coquillon & Durand 2010).

2.2 La liaison, un phénomène variable

La liaison est donc un phénomène complexe. Comme le relèvent Soum Favaro, Coquillon & Chevrot (2014: 1) dans la préface d'un ouvrage récent entièrement consacré à ce phénomène, outre sa complexité, elle se caractérise également par "*[son] hétérogénéité et [sa] variabilité*", qui en font un sujet inépuisable d'étude pour les chercheurs, et ce dans une perspective pluridisciplinaire. Leur ouvrage illustre l'aspect multidimensionnel de la liaison puisqu'il traite des aspects linguistiques, neurolinguistiques, psycho-linguistiques et sociolinguistiques de la liaison.

---

6 Une syllabe ouverte est une syllabe qui se termine par une voyelle (= V) prononcée (ex. V, CV, CCV, etc.), par opposition à une syllabe fermée, qui se termine quant à elle par une consonne (= C) prononcée (ex. VC, VCC, CVC, etc.).

Selon Laks (2015), la liaison est un phénomène "*composite non unitaire*", qui ne peut être envisagé à l'aide d'un mécanisme unique. On peut en effet relever que, si la liaison est variable, elle ne l'est pas dans tous les cas. Certaines liaisons sont en effet réalisées de manière catégorique par tous les locuteurs, notamment entre un déterminant et le nom qui suit (ex. "les oranges" [lezɔrãʒ]), ou dans une expression figée (ex. "en outre" [ãnutʁ]), alors que dans d'autres contextes, elles sont soumises à variation (ex. "c'est évident" [sɛ(t)evidã], "trop aimé" [tʁo(p)eme]). Laks & Calderone (2014) ont ainsi mis en évidence, sur la base d'une analyse quantitative portant sur plus de 16'000 formes, la disparité des contextes morphosyntaxiques dans lesquels la liaison apparaît. Wauquier (2009) distingue quant à elle deux niveaux de variation en ce qui concerne la liaison: une variation conditionnée par le contexte – phonologique, prosodique, lexical et morphosyntaxique –, qui fait que certaines liaisons sont systématiquement réalisées, réalisées de manière variable ou non réalisées. Le deuxième niveau concerne les liaisons variables pour lesquelles on observe une variation libre, qui reflète, selon elle, le choix du locuteur de réaliser ou non la liaison, généralement en fonction de facteurs d'ordre sociolinguistique. Parmi ceux-ci, Eychenne et al. (2014) et Durand & Lyche (à paraître) relèvent notamment le registre, l'âge, la provenance du locuteur, son niveau d'instruction ou encore son degré de connaissance de l'orthographe.

La tripartition généralement utilisée pour catégoriser la liaison en fonction de son degré de réalisation est, selon Laks (2015), très fluctuante et dépend à la fois des auteurs et des époques. D'un point de vue prescriptif et normatif, on parle de liaisons *obligatoires*, *facultatives* et *interdites* (cf. Delattre 1951; Fouché 1959), alors que, dans une perspective descriptive et variationniste, les liaisons se répartissent en *catégoriques*, *variables* et *erratiques* (cf. Encrevé 1988). L'approche de la liaison proposée dans le cadre du projet "Phonologie du français contemporain: usages, variétés et structure" (Durand, Laks & Lyche 2009, 2014; ci-après PFC) se place quant à elle dans une perspective que Laks (2015) qualifie de "*sans a priori*" et utilise de ce fait la dénomination suivante: liaisons *toujours réalisées*, *parfois réalisées* et *jamais réalisées*.

2.3 La liaison et la linguistique de corpus

Si, aujourd'hui, l'apport de la linguistique de corpus à l'étude du français parlé n'est plus à démontrer (Durand 2009; voir également Avanzi, Béguelin & Diémoz à paraître) et que la plupart des études sur la liaison sont basées sur des données authentiques, cela n'a pas toujours été le cas, comme le rappellent Eychenne et al. (2014). De nombreux travaux ont en effet été consacrés au traitement de la liaison en français au cours de la deuxième moitié du 20ème siècle, mais peu d'entre eux ont toutefois pris en compte des données réelles. La plupart des analyses classiques en linguistique ont été

élaborées sur la base de descriptions normatives destinées à faciliter l'apprentissage du phénomène par des apprenants étrangers – les auteurs mentionnent notamment Delattre (1951) et Fouché (1959). De nombreux exemples ont ainsi perduré sans que l'on mette en doute leur authenticité – ils citent notamment "le sot [t]aigle" de Féry (2003). Les années 80 ont amorcé un tournant avec, entre autres, Morin & Kaye (1982) et Morin (1986), qui se sont opposés à une approche uniquement théorique de la liaison. En parallèle, les premiers travaux sur corpus sont apparus et, comme le mentionnent Eychenne et al. (2014), ont commencé à remettre en question un certain nombre d'idées reçues à propos de la liaison. Le premier que ces auteurs mentionnent est celui d'Ågren (1973), qui a présenté, dans sa thèse de doctorat, une analyse de la liaison basée sur un corpus radiophonique. Son travail a notamment permis de montrer le lien entre le registre et la réalisation de la liaison, le nombre de liaisons facultatives réalisées étant inversement proportionnel au degré de familiarité. Son travail a également souligné l'importance de la fréquence lexicale, facteur qui est également mis en évidence, quelques années plus tard, par De Jong (1994), qui se base quant à lui sur un corpus composé d'entretiens semi-formels récoltés auprès de 45 locuteurs à Orléans. Son étude révèle également une différence entre hommes et femmes – les femmes produisant davantage de liaisons que les hommes –, en fonction de l'âge – les jeunes locuteurs réalisant moins de liaisons que les locuteurs âgés – et enfin selon la classe sociale – les locuteurs socialement favorisés présentant un taux de liaison plus élevé que ceux des milieux moins favorisés. Eychenne et al. (2014) mentionnent encore le travail d'Encrevé (1988), qui, sur la base d'un corpus rassemblant les discours enregistrés de 21 hommes politiques français qui ont marqué les années 1978-1981, a montré que la liaison n'est pas systématiquement enchaînée (ex. "il est arrivé" produit [i-lɛt-a-ʁi-ve]), la consonne de liaison se plaçant en coda de la dernière syllabe du mot liaisonnant.

2.4 La liaison dans le projet PFC

Les recherches menées dans le cadre du projet PFC (Durand et al. 2009, 2014, www.projet-pfc.net) viennent compléter ce tableau. Ce programme collaboratif, visant à constituer une importante base de données permettant de rendre compte de la diversité des usages oraux du français dans l'ensemble de la francophonie, regroupe plus de soixante chercheurs de différents pays et fait aujourd'hui figure de corpus de référence pour le français parlé. Sur la base d'un protocole commun, les données de plus de 700 locuteurs issus de l'ensemble de la francophonie ont été collectées – dont un peu plus de 40% ont été transcrites et analysées. Le corpus PFC comprend actuellement 37 points d'enquête dans l'espace francophone entièrement consultables en ligne, soit 396 locuteurs. Pour chaque point d'enquête, les enquêteurs sélectionnent dix à douze locuteurs selon la

méthode des réseaux denses (Milroy 1980), répartis de manière équilibrée en termes de sexe et couvrant généralement trois tranches d'âge. Une certaine diversité socio-économique est visée, même si cet aspect est plus difficile à assurer avec des groupes de locuteurs aussi restreints. La méthodologie de recueil de données, dans la droite ligne des travaux classiques de Labov, inclut, pour chaque locuteur, quatre tâches: la lecture d'une liste de mots, la lecture d'un texte, un entretien semi-dirigé ainsi qu'un entretien libre entre deux – parfois trois – locuteurs[7]. Tous les enregistrements ainsi obtenus sont alignés sur le signal et transcrits orthographiquement sous *Praat* (Boersma & Weenink 2014). Le schwa et la liaison constituent deux cibles privilégiées dans le projet PFC. Pour toutes les enquêtes, ces deux phénomènes sont traités de manière systématique par le biais d'un codage alphanumérique, effectué sous *Praat*, et inséré dans des tires séparées (une tire par phénomène). Divers outils génériques ou propres au projet permettent ensuite d'explorer les données et de les croiser avec des métadonnées telles que l'âge de la personne ou sa profession, par exemple. A ce jour, 202'089 sites ont ainsi été codés pour le schwa et 53'561 en ce qui concerne la liaison. Comme Laks (2015) le souligne, le codage a été conçu dans une perspective qui n'intègre aucun a priori, l'objectif étant de fournir une description des usages actuels de la liaison dans la francophonie. Tous les sites traditionnellement considérés comme des sites de liaison potentielle sont codés au moyen d'un code qui comprend trois champs. Le premier indique le nombre de syllabes du mot liaisonnant; le deuxième fournit des indications quant à la réalisation de la liaison (absence de liaison, présence de liaison enchaînée ou non-enchaînée, liaison incertaine, liaison réalisée avec une consonne épenthétique); le troisième champ s'intéresse à la nature de la consonne de liaison si elle est réalisée et permet également, dans le cas de liaison avec voyelle nasale, d'intégrer des informations concernant la dénasalisation éventuelle de la voyelle nasale (ex. "bon appétit" [bɔnapeti] *vs* "mon ami" [mɔ̃nami]). Enfin, le symbole "h" permet d'indiquer que la transition entre les mots ne se fait pas de manière fluide (p. ex. présence d'un coup de glotte). Ainsi, la séquence "il est à Genève depuis un an", produite [i-lɛ-ta-ʒə-nɛv-də-pɥi-ɛ̃-nɑ̃], sera codée: "il est11t à Genève depuis20 un11nVN an"[8]. Ces données peuvent ensuite être analysées par le biais de la plateforme *Dolmen* (Eychenne & Paternostro à paraître), qui permet d'effectuer des requêtes ciblées et fournit des statistiques descriptives sur les données codées.

Cette procédure de traitement de la liaison a permis d'effectuer des analyses détaillées (voir, par exemple, Bordal & Lyche 2008; Durand & Lyche 2008;

---

7 Pour plus de détails concernant la méthodologie et l'état actuel du projet, voir Racine, Durand & Andreassen (à paraître) ainsi que le volume édité par Detey, Durand, Laks & Lyche (à paraître).

8 "est11t": monosyllabe, liaison réalisée en [t]; "depuis20": polysyllabe, liaison non réalisée; "un11nVN": monosyllabe, liaison réalisée en [n], sans dénasalisation de la voyelle nasale du mot "un".

Mallet 2008; Durand, Calderone, Laks & Tchobanov 2011; Côté 2012; Côté 2013; Laks, Celata & Calderone 2014; Boutin 2014; Boutin & Lyche 2014; Eychenne et al. 2014; Barreca & Christodoulides 2015). Ces travaux ont notamment contribué à mettre en lumière certaines grandes tendances. En ce qui concerne par exemple la classification des liaisons, les données PFC confortent et affinent les résultats de De Jong (1994). Durand & Lyche (2008) réduisent ainsi, pour le français hexagonal, à quatre contextes – qui se situent tous à l'intérieur d'une unité prosodique – les liaisons catégoriques dans les usages: déterminant + substantif (ex. "u**n** enfant"), proclitique + verbe ("il**s** avaient", "il**s** y allaient")[9], verbe + enclitique (ex. "di**t**-il", "mange**s**-en") et expressions figées (ex. "de plu**s** en plus"). Les données PFC montrent également que la liaison variable est très peu réalisée en conversation spontanée et que les liaisons sont presque toutes systématiquement produites avec enchaînement. Du point de vue des facteurs externes, elles révèlent que la liaison semble essentiellement influencée par l'âge et l'origine géographique des locuteurs ainsi que par le registre (pour une présentation plus détaillée de ces résultats, voir Eychenne et al. 2014).

2.5 L'enseignement de la liaison en FLE[10]

Ce panorama contraste avec les descriptions que l'on trouve dans la plupart des manuels de prononciation du français. En effet, dans le domaine du FLE, c'est encore souvent la tripartition prescriptive et normative qui sert de fondement, comme l'illustre la terminologie utilisée. La plupart des manuels parlent en effet de liaisons *obligatoires*, *facultatives* et *interdites*. La liaison y est présentée sous forme de tableaux ou de listes basées sur des généralisations d'ordre morphosyntaxique (p. ex. "*la liaison est obligatoire entre le déterminant et le nom qui le suit*"), suivies d'exemples. Les autres facteurs de variation – d'ordre sociolinguistique, lexical ou prosodique – ne sont généralement pas mentionnés, mis à part la distinction, pour les liaisons facultatives, entre registre courant/standard et soutenu. On peut souligner que le nombre de règles énoncées est plutôt élevé (selon les manuels, entre six et neuf pour chacune des catégories "obligatoire" et "interdite"), ce qui contraste, pour la catégorie "obligatoire", avec les quatre contextes mentionnés par Durand & Lyche (2008). Par ailleurs, outre les importantes divergences qui peuvent être observées d'un manuel à l'autre, on peut relever que les explications sont parfois opaques[11].

---

9 A noter toutefois que les variétés canadiennes font exception puisque la liaison n'est pas réalisée après "ils" (Côté 2012).

10 Les réflexions présentées dans cette section ainsi que dans la partie 3 sont en partie reprises de Racine (2014).

11 On trouve par exemple, pour expliquer les contextes dans lesquels la liaison est obligatoire, l'explication suivante: "*En style standard, la liaison est obligatoire à l'intérieur des groupes rythmiques lorsque la cohésion lexicale ou syntaxique entre les mots est maximale*" (Charliac & Motron 2006: 30).

Certes, comme le relève Thomas (1998), la complexité de la liaison rend difficile sa didactisation par des règles simples. Force est de constater cependant que certaines règles énoncées dans les manuels ne correspondent pas à la réalité des usages actuels, tels que décrits notamment par les travaux basés sur des corpus oraux récents. Ainsi, par exemple, la liaison est présentée comme systématiquement réalisée (liaison "obligatoire") après les prépositions monosyllabiques "en", "dans", "chez", "sans" et "sous" (Abry & Chalaron 1994: 110). Or, Eychenne et al. (2014: 44) montrent, en se basant sur le corpus PFC, que la liaison n'est pas réalisée de manière uniforme après toutes les prépositions monosyllabiques. Si elle l'est en effet toujours après "en" (taux de réalisation: 99.35%), c'est moins le cas après "sans" (93.13%), et encore moins après "chez" (75.68%). Un examen plus détaillé des liaisons après "chez" révèle que la liaison est systématiquement réalisée lorsque la préposition est suivie d'un clitique, alors qu'elle n'est présente que dans la moitié des occurrences avant un syntagme nominal, ce qui, selon les auteurs, illustre le rôle primordial joué par la prosodie.

Ainsi, comme le soulignent Eychenne et al. (2014), les travaux basés sur des corpus oraux ont clairement montré que la liaison ne peut être expliquée en se basant uniquement sur des principes d'association mécanique mais que d'autres facteurs, tels que par exemple la fréquence lexicale du mot liaisonnant et la prosodie doivent être pris en compte. On peut ajouter à cela que le décalage entre descriptions et usages réels peut également s'expliquer par le fait que l'on ne s'exprime plus aujourd'hui de la même manière qu'à l'époque où ces descriptions ont été élaborées. L'importance du facteur âge a en effet été mis en évidence par De Jong (1994): les locuteurs âgés de son corpus réalisent un taux de liaison significativement supérieur par rapport aux jeunes. Il observe un changement entre la tranche des 18-29 ans et celle des 30-49 ans, ce qui situe ce changement majeur au milieu des années soixante[12], les descriptions traditionnelles mentionnées – Delattre (1951) et Fouché (1959) – y étant donc antérieures.

Ainsi, le constat effectué au niveau linguistique par Eychenne et al. (2014: 56), à savoir qu'il n'est plus possible aujourd'hui de se satisfaire de données "*fabriquées*" et que l'on se doit d'intégrer les résultats des études sur corpus dans les modélisations phonologiques, semble s'imposer également en ce qui concerne l'enseignement du FLE: les descriptions des manuels de phonétique doivent être renouvelées afin de mieux correspondre aux usages réels en vigueur.

---

[12] Ce changement coïnciderait, selon Laks (2013), avec le mouvement de mai 68.

3. La liaison: les enjeux pour le FLE

3.1 Les difficultés au niveau de la micro- et de la macro-planification

Comme nous l'avons mentionné précédemment, la liaison constitue une source de difficulté pour les apprenants, tant au niveau de la *micro-planification* que de celui de la *macro-planification* (Racine 2014). Si la plupart des ressources pédagogiques consacrées à la liaison se focalisent sur les difficultés liées à la *macro-planification*, qui consiste à identifier le contexte afin de déterminer si la liaison doit – ou peut – être réalisée ou non, nous avons vu dans la section précédente qu'un renouvellement du matériel pédagogique est nécessaire, afin que l'enseignement corresponde davantage aux usages en vigueur et non à une vision purement théorique et normative de la liaison. Ce mouvement semble avoir été amorcé récemment puisque certaines descriptions ont commencé à intégrer le travail conséquent effectué sur la liaison dans le domaine de la linguistique de corpus (voir par exemple Lauret 2007; Abry & Chalaron 2011; Pustka 2011).

Les difficultés au niveau de la *micro-planification*, liées au moment même où la liaison est réalisée, sont quant à elle traitées de manière beaucoup moins explicite dans les ressources pédagogiques. La réalisation de la liaison va pourtant à l'encontre de certains principes que l'apprenant a dû – ou est en train – d'intégrer, comme le souligne Howard (2013). Il a ainsi dû apprendre que la consonne finale graphique d'un grand nombre de mots ne se prononce pas, ce qui constitue une difficulté majeure pour les apprenants dont la L1 dispose d'une très bonne correspondance entre la phonie et la graphie. Or, dans le cas de la liaison, cette consonne finale habituellement non prononcée peut tout de même l'être sous certaines conditions. De plus, outre le fait que la consonne finale se resyllabe avec la voyelle initiale du mot suivant, sa réalisation ne correspond pas toujours à la graphie (ex. "grand", avec une consonne finale non prononcée qui, en liaison, se réalise en [t] et non [d], "gran**d** éléphant", [gʁɑ̃-te-le-fɑ̃]). La liaison en [n] renforce ce phénomène. Si l'acquisition des voyelles nasales du français constitue déjà en soi une difficulté majeure pour la plupart des apprenants, la dimension graphique leur complique encore la tâche. En effet, en plus du fait que les voyelles nasales possèdent des graphies multiples, l'apprenant doit acquérir des graphèmes complexes dans lesquels la consonne nasale ne se prononce pas (ex. "moye**n**" [mwajɛ̃]). Or, lorsqu'il s'agit d'une liaison en [n], non seulement la consonne nasale graphique se prononce mais la voyelle qui précède peut aussi perdre sa nasalité (ex. "Moyen Âge" [mwajɛnaʒ]), sans que cela soit systématique (ex. "bien entendu" [bjɛ̃nɑ̃tɑ̃dy]). Si la plupart des manuels de prononciation présentent les différentes consonnes de liaison, et leurs différentes graphies, ainsi que le phénomène de resyllabation qui permet d'enchaîner la liaison, ces éléments le sont souvent de manière sommaire, et avec un nombre d'exercices restreint. Le fait que la liaison va à l'encontre de

certains principes (non prononciation des consonnes finales et de la consonne nasale dans les graphèmes correspondant aux voyelles nasales) n'est généralement pas thématisé et en tout cas pas exercé de manière contrastive. On trouve par ailleurs très peu d'explications ou d'exercices concernant les modifications de la voyelle précédente, notamment la dénasalisation. On peut enfin relever que la liaison est généralement traitée de manière totalement déconnectée à la fois des aspects prosodiques et de ceux qui touchent le niveau segmental. Ainsi, concernant l'interaction avec les aspects segmentaux, les difficultés spécifiques liées à la liaison en [n] se répartissent généralement sur deux sections, celle portant sur la liaison et celle consacrée aux voyelles nasales, sans toutefois présenter de manière explicite les liens entre les deux niveaux.

3.2 Les travaux précédents

Les travaux qui se sont penchés sur les réalisations de la liaison dans des productions d'apprenants de FLE (Mastromonaco 1999; Thomas 2002, 2010; Howard 2005, 2013; De Moras 2011; Harnois-Delpiano et al. 2012) ont permis de confirmer que les difficultés rencontrées ne se situent pas uniquement au niveau de la *macro-planification*, mais que les aspects liés à la *micro-planification* sont également problématiques.

Malgré la grande hétérogénéité qui caractérisent ces travaux, en termes de niveau (apprenants allant de débutants à avancés), de contexte d'apprentissage (avec et sans séjour prolongé dans un milieu francophone) et de tâche (production de séquences isolées, description d'images, texte lu, entretien guidé et parole spontanée), les résultats sont relativement homogènes. Ils révèlent que les liaisons "obligatoires" ne semblent pas poser de problèmes aux apprenants avancés. L'étude de Howard (2013), qui comporte une dimension longitudinale, montre en outre une progression dans le taux de réalisation de certaines catégories, notamment entre déterminant et substantif et après les pronoms sujets. Howard (2013) observe toutefois que certaines liaisons appartenant à la catégorie "obligatoire" si l'on s'en tient aux descriptions traditionnelles telles que celle de Delattre (1951), sont en revanche moins systématiquement réalisées par les apprenants. C'est le cas notamment de la liaison entre un adjectif préposé et le substantif qui le suit. Cette observation nous ramène au décalage observé par Eychenne et al. (2014) entre la perspective prescriptive/normative des descriptions traditionnelles et les usages réels. On peut rappeler ici que la liaison entre un adjectif préposé et le substantif n'apparaît pas comme l'un des quatre contextes dans lesquels la liaison est systématiquement réalisée dans les données de Durand & Lyche (2008). La comparaison avec les natifs doit par conséquent être considérée avec précaution car les études portant sur les apprenants n'ont souvent pas constitué de corpus de données natives

comparables[13] et s'appuient sur la littérature dans le domaine qu'elles utilisent comme modèle à suivre. Or, comme De Moras (2011) le souligne et comme nous l'avons vu précédemment, les différentes catégories ne sont pas définies de manière uniforme dans la littérature.

La question de la variété de référence devient encore plus prégnante en ce qui concerne les liaisons réalisées de manière variable. Les travaux dans le domaine s'accordent généralement sur le fait que les apprenants réalisent moins de liaisons variables que les locuteurs natifs (Mastromonaco 1999; Thomas 2002; Howard 2005, 2013; De Moras 2011). Il faut toutefois se demander comment est effectuée la comparaison avec les natifs. En effet, comme le rappellent Eychenne et al. (2014), les liaisons variables sont extrêmement sensibles à la variation diaphasique, ce qui se traduit, dans les données PFC, par un taux de réalisation de la liaison variable moins élevé en conversation par rapport à la lecture du texte. Pour la liaison après la forme "(c')est", ces auteurs observent par exemple un taux de réalisation de 31.68% en conversation, contre 76.92% pour le texte lu. Les travaux portant sur les apprenants montrant une forte hétérogénéité au niveau des tâches, il n'est donc pas surprenant d'observer des taux de réalisation très variables d'une étude à l'autre. Il est également intéressant de constater que Thomas (2002) compare les réalisations de ses apprenants anglophones aux chiffres obtenus par Ågren (1973), calculés sur la base de 134 extraits radiophoniques enregistrés entre 1960 et 1961. Pour la liaison après "(c')est", le taux de réalisation s'élève chez Ågren (1973) à 97%, ce qui est beaucoup plus élevé que le taux obtenu en conversation (31.68%), dans les données PFC, par Eychenne et al. (2014). Ces éléments incitent par conséquent à la prudence en ce qui concerne les analyses des contextes de réalisation de la liaison produits par les apprenants, les tâches dont sont issues les productions examinées ainsi que la base servant de référence à la comparaison devant être soigneusement prises en compte.

En ce qui concerne les difficultés liées à la *micro-planification*, les études précédentes ont permis de mettre en évidence deux phénomènes principaux, à savoir des difficultés au niveau de la nature de la consonne de liaison, en lien avec la graphie, ainsi qu'au niveau de l'enchaînement accompagnant la liaison. Ainsi, d'une part, Mastromonaco (1999), Thomas (2002) et Harnois-Delpiano et al. (2012) observent en effet des erreurs de type [gʁɑ̃dami] pour "grand ami" ou [gʁɑ̃nɔm] pour "grand homme", que ces auteurs qualifient de prononciation "orthographique" et qui ne sont pas observées chez les enfants natifs (Wauquier 2009). Mastromonaco (1999) et De Moras (2011) mentionnent en outre la difficulté particulière engendrée par les liaisons en [n],

13 Ce n'est pas le cas de toutes les études. De Moras (2011) a par exemple constitué un groupe témoin de natifs qui sert de référence pour la comparaison avec les apprenants anglophones de son étude.

en lien avec la question des voyelles nasales. D'autre part, Mastromonaco (1999), Thomas (2002) et De Moras (2011) observent un taux important de liaisons réalisées sans enchaînement par les apprenants, alors que ce type de réalisation n'apparaît pas chez les enfants natifs (Wauquier 2009) et est quasiment absent des données de conversation du corpus PFC (Eychenne et al. 2014). Deux explications sont avancées pour expliquer ces réalisations sans enchaînement: cela peut refléter une hésitation devant un mot difficile (Thomas 2002), ou bien l'influence de la tendance à la syllabation fermée de l'anglais, la L1 des apprenants des études de Thomas (2002) et De Moras (2011). Ces résultats montrent que les difficultés au niveau de la *micro-planification* ne sont pas triviales et qu'il semble nécessaire de les traiter plus explicitement au niveau de l'enseignement du FLE. De Moras (2011) observe en effet une amélioration notoire au niveau de l'enchaînement des liaisons obligatoires après un enseignement explicite consacré à la liaison.

Comme le relève Wauquier (2009), un élargissement des populations d'apprenants examinées est souhaitable afin de mieux cerner les difficultés que la liaison pose aux apprenants. L'apprentissage de la liaison se faisant généralement à partir des formes écrites, ou du moins avec l'apport de la graphie, la prise en compte plus systématique d'apprenants dont le système graphique repose sur d'autres correspondances (p. ex. japonophones, sinophones, arabophones, etc.), ou sur un alphabet différent (p. ex. hellénophones, russophones, etc.) semble pertinente[14]. En outre, la comparabilité des données, non seulement avec les productions natives mais également entre les différentes populations d'apprenants doit être améliorée, de même qu'en termes de tâche(s) effectuée(s). Le projet IPFC semble offrir un cadre adéquat pour poursuivre ces différents objectifs.

4. Le projet IPFC

Dans certaines régions francophones telles que la Louisiane ou sur le continent africain par exemple, les locuteurs des enquêtes PFC ne sont pas uniquement francophones mais bilingues, voire plurilingues. Un volet non natif du projet constituait par conséquent une extension logique de ce programme. C'est en 2008 que le projet "Interphonologie du français contemporain" (IPFC) a été lancé (Detey & Kawaguchi 2008; Detey & Racine 2012; Racine et al. 2012; Detey et al. à paraître; Racine & Detey à paraître). Son objectif est de décrire la prononciation du français langue étrangère par des apprenants de différentes L1, sur la base d'un large corpus collecté avec un protocole commun. À terme, ces données seront utiles à des fins théoriques, notamment pour tester les prédictions des modèles d'acquisition phonologique

[14] A noter que Harnois-Delpiano et al. (2012) ont initié ce travail en examinant les réalisations de liaisons chez les apprenants coréens, dont le système graphique repose sur un alphabet différent (cf. Han & Eychenne à paraître).

en L2, aussi bien qu'applicatives, dans le domaine du traitement automatique du langage.

IPFC a permis de combler un manque puisque, bien que la linguistique de corpus ait également pris son essor dans le domaine de la L2 et que des corpus d'apprenants collectées à des fins d'études phonétiques et phonologiques existent pour différentes langues (pour l'allemand, voir Gut 2005; Zimmerer, Trouvain & Bonneau 2015; pour l'anglais, voir Gut 2005; Visceglia, Tseng, Kondo, Meng & Sagisaka 2009; Andreassen, Herry-Bénit & Kamiyama 2015; pour l'espagnol, voir Carranza sous presse; pour le néerlandais, voir Neri, Cucchiarini & Strik 2006 et pour le polonais, voir Cylwik, Wagner & Demenko 2009), on ne trouvait rien de comparable pour le français avant 2008. Le lancement du projet IPFC a été suivi par celui de deux autres corpus, COREIL (Delais-Roussarie & Yoo 2010; Delais-Roussarie, Santiago & Yoo 2015) et PhoDiFle (Landron et al. 2012).

Le projet a débuté par la constitution d'un corpus d'apprenants japonophones (IPFC-japonais, S. Detey & Y. Kawaguchi), rapidement suivi par un corpus d'apprenants hispanophones (IPFC-espagnol, I. Racine), qui sont les deux corpus les plus avancés en termes de méthodologie et d'analyses effectuées sur les données. D'autres équipes se sont jointes au projet, et le recueil de données d'apprenants de quinze L1 différentes (allemand, anglais, arabe, coréen, danois, espagnol, grec, italien, japonais, néerlandais, norvégien, portugais, suédois, russe et turc) est actuellement en cours, avec parfois une prise en compte de la variation diatopique en L1 puisque plusieurs points d'enquêtes différents peuvent être collectés pour une même L1 (ex. pour l'allemand, en Allemagne, en Autriche et en Suisse)[15].

La comparaison avec les natifs étant, on l'aura compris (cf. section 3.2), un élément crucial pour l'analyse des réalisations des apprenants, le protocole IPFC a été élaboré au plus proche de celui de PFC. Il comporte les six tâches suivantes:

- La répétition d'une liste spécifique de mots, constituée de 34 mots comportant des difficultés avérées pour l'ensemble des apprenants de FLE (p. ex. voyelles nasales, voyelles arrondies, etc.) et de 25-35 mots comportant des difficultés spécifiques à la population en question (ex. /b/-/v/ pour les hispanophones, /r/-/l/ et groupes consonantiques pour les japonophones, dévoisement pour les germanophones, etc.);
- La lecture de la liste de mots PFC, composée de 94 mots;

[15] Pour plus de détails concernant les équipes et les points d'enquête, voir le site du projet (http://cblle.tufs.ac.jp/ipfc/).

- La lecture de la liste spécifique, ce qui permet de disposer de cibles identiques, produites à l'aide de deux tâches différentes (répétition et lecture);
- La lecture du texte PFC, intitulé "Le Premier Ministre ira-t-il à Beaulieu?";
- Un entretien guidé avec un natif, généralement un enseignant, qui pose des questions fermées et ouvertes, élaborées en fonction du niveau de l'étudiant (A1-B1 et B2-C2 du CECRL[16]);
- Une conversation libre entre deux apprenants.

Le protocole fournit environ une heure de données par apprenant et un questionnaire sociolinguistique visant à établir le profil de l'apprenant le complète.

Sur le modèle PFC, les données sont alignées sur le signal sous *Praat* et sont transcrites orthographiquement selon des conventions qui ont été adaptées aux spécificités de la L2 (voir Racine, Zay, Detey & Kawaguchi 2011). En effet, si la transcription orthographique constitue déjà une forme d'interprétation puisque, comme le relève Delais-Roussarie (2009) pour la parole native, elle est le "*résultat d'une analyse, ou plutôt d'une abstraction, des données réelles*", cette question se pose de manière encore plus accrue pour des productions d'apprenants (MacWhinney 2015). On comprend dès lors aisément pourquoi une transcription complète du corpus à l'aide d'un alphabet phonétique, par exemple API (Alphabet Phonétique International) ou Sampa (*Speech Assessment Methods Phonetic Alphabet*) n'est pas envisageable pour le projet IPFC. Comme le relèvent Durand & Lyche (2003), une transcription phonémique consisterait à mettre "*la charrue avant les boeufs*", en supposant que le système que l'on cherche à établir par le biais de l'enquête a déjà été découvert. Si, en revanche, on privilégie une transcription de type allophonique, on peut alors se demander "*quel degré de finesse phonétique*" devrait être adopté, "*bon nombre de réalisations ne correspondant pas à des choix binaires mais à des valeurs sur des échelles continues*" (Durand & Lyche 2003: 230). Comme cela a été fait dans PFC pour la liaison et le schwa, l'approche par le biais d'un codage alphanumérique a donc été privilégiée dans le cadre du projet IPFC et a également été appliquée sur le plan segmental (cf. Detey 2012, 2014). Une procédure de codage a ainsi été développée pour les voyelles nasales (Detey, Racine & Kawaguchi 2014; Detey, Racine, Eychenne & Kawaguchi 2014), les voyelles orales (Detey et al. à paraître), les consonnes et les groupes consonantiques (Detey & Racine 2014), ainsi que pour le schwa (Racine, Detey & Andreassen

[16] *Cadre Européen Commun de Référence pour les langues* (Conseil de l'Europe 2001).

en préparation) et la liaison. Ce dernier sera présenté de manière détaillée dans la section suivante.

Ce codage permet d'inclure à la fois des champs descriptifs, tels que par exemple le segment/élément cible ou le contexte phonologique gauche et droit, qui vont permettre d'effectuer des requêtes ciblées. Il comprend également des éléments pour lesquels l'évaluation perceptive d'un élément doit être effectuée (p. ex. nasalité de la voyelle, adéquation de son timbre, présence/absence d'un élément consonantique après la voyelle nasale, etc.). Cette analyse privilégie donc une approche perceptive de la parole non native, sans toutefois exclure les analyses acoustiques, qui peuvent être menées dans un deuxième temps sur les formes jugées déviantes, de manière à établir les corrélats acoustiques qui ont conduit à cette analyse perceptive. La validité de cette approche par codage a été évaluée, pour les voyelles nasales (Detey et al. 2014) et arrondies (Racine 2012) en comparant les résultats de tests de perception classiques, avec des locuteurs natifs non experts, à ceux obtenus sur la base du codage. Une procédure de double, voire triple, codage en aveugle, avec comparaison entre les codeurs, a également été mise en place. Des requêtes peuvent ensuite être effectuées sur les données codées par le biais du concordancier *Dolmen*, pour lequel des interfaces spécifiques pour IPFC ont été développées par J. Eychenne (cf. Eychenne & Paternostro à paraître).

5. Le traitement de la liaison dans IPFC

Cette approche par codage a également été appliquée pour la liaison dans le projet IPFC, sur le modèle de la méthodologie adoptée dans PFC. Dans la ligne définie par Laks (2015) d'appliquer un codage visant à décrire sans "*a priori*" les réalisations des apprenants, nous avons pris le parti de considérer comme un site potentiel de liaison tout mot se terminant par une consonne graphique habituellement non prononcée situé devant un mot commençant par une voyelle ou un h muet. Si l'on se place en amont du système que l'on souhaite découvrir, il aurait en effet été biaisé de considérer que l'apprenant est capable de différencier un contexte tel que "un chat e**t** un chien", la liaison après "et" n'étant jamais réalisée, de contextes comme "devan**t** une porte", "tro**p** intelligent", "che**z** elle" ou "de**s** enfants", où elle peut ou doit l'être, à des degrés variables, et de se limiter à ne coder que les cas où la liaison peut être réalisée. Cela explique que, dans le projet PFC, 35 sites potentiels de liaison ont été codés dans le texte, contre 37 dans le projet IPFC (voir Annexe)[17].

[17] Deux sites potentiels de liaison ont été ajoutés: entre "région" et "en" ainsi qu'entre "opposition" et "aurait" par rapport à ceux du projet PFC. Comme nous l'avons vu précédemment, le code PFC n'intègre aucune information concernant les catégories de liaison. Toutefois, Durand, Laks & Lyche (2002), dans la présentation du codage liaison, ont comparé les sites du texte PFC avec la classification de Delattre (1951). Le texte PFC contient 35 sites de liaison potentiels,

Inspiré par le code PFC, le codage mis en place dans IPFC a dû être adapté de manière à prendre en compte certaines spécificités de la L2. Le code PFC a donc dû être adapté à certaines spécificités de la L2. Le format du codage a également été revu, de manière à conserver une certaine cohérence avec les autres codes développés dans IPFC. Le code IPFC pour la liaison comprend donc sept champs, séparés par un tiret bas (*underscore*). Les quatre premiers champs portent sur des éléments descriptifs, alors que les trois derniers ciblent l'évaluation perceptive de la liaison. Il se présente de la manière suivante:

- Champ 1: nature de la consonne de liaison cible;
- Champ 2: catégorie syntaxique du mot liaisonnant;
- Champ 3: catégorie syntaxique du mot qui suit;
- Champ 4: nombre de syllabes du mot liaisonnant et nature – orale ou nasale – de la voyelle du mot liaisonnant;
- Champ 5: réalisation de la liaison et si oui, avec ou sans enchaînement;
- Champ 6: nature et caractéristiques de la consonne de liaison (correspondance avec la cible ou non, etc.);
- Champ 7: présence d'une pause, d'une hésitation ou d'un coup de glotte.

Le champ 1 comprend deux éléments et permet de coder la nature de la consonne de liaison cible ([z] = 10 et 11, [n] = 20, [t] = 30 et 31, [ʁ] = 40, [p] = 50 et [g] = 60), en différenciant, pour [z] et [t], entre les formes du singulier (p. ex. "gros insecte", "est allé"), codées 10 pour [z] et 30 pour [t], et celles du pluriel (p. ex. "des insectes", "sont allés"), codées 11 pour [z] et 31 pour [t]. Le codage prévoit également les cas de liaisons épenthétiques, où il n'y a aucune consonne cible (ex. "si évident" produit [sitevidɑ̃]), qui sont codés quant à eux 70.

Pour les champs 2 et 3, l'étiquetage a été adapté sur la base des catégories morphosyntaxiques de *TreeTager* (Schmid 1994, 1995)[18] et qui sont les suivantes: ABR = abréviation; ADJ = adjectif; ADV = adverbe; AUX = verbe auxiliaire (être, avoir, pouvoir, vouloir, devoir); CON = conjonction; DET = déterminant (y compris adjectif démonstratif et possessif); EXF = élément appartement à une expression figée (ex. de temps en temps, de plus en plus, etc.); INF = verbe à l'infinitif; INT = interjection; NAM = nom propre; NEG = négation; NOM = nom; NUM = numéral; PDE = préposition + article (ex. au,

---

dont 17 de liaisons obligatoires, 14 de liaisons facultatives et 4 de liaisons interdites. Pour IPFC, deux sites potentiels de liaison interdite s'ajoutent à cet inventaire.

18 Voir également le site http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/.

du, aux, des); PPA = participe passé; PPR = participe présent; PRO = pronom; PRP = préposition; VER = autre verbe (sans distinction de temps ou de mode); XXX = indéfinissable (p. ex. faux départs, hésitations, etc.).

Le champ 4, qui comporte trois éléments, permet d'indiquer le nombre de syllabes du mot liaisonnant, à savoir 100 pour un monosyllabe et 200 pour un polysyllabe. Pour les cibles en [n], la nature de la consonne précédente est également indiquée. Ainsi, si la voyelle précédente est dénasalisée, le code sera 1VO ou 2VO, selon s'il s'agit d'un mono- ou d'un polysyllabe (ex. "en outre" réalisé [anutʁ]). S'il n'y a pas de dénasalisation, le code sera alors 1VN ou 2VN (ex. "en outre" produit [ãnutʁ]).

Le champ 5, qui est constitué de deux chiffres, code la réalisation de la liaison de la manière suivante: 00 indique que la liaison n'est pas réalisée, 11 que la liaison est réalisée avec enchaînement, 12, qu'elle l'est sans enchaînement et 13, une liaison réalisée avec incertitude quant à l'enchaînement.

Le champ 6, qui comprend un élément, fournit des indications quant à la nature de la consonne de liaison. Ainsi, 0 est utilisé lorsque la liaison n'est pas réalisée, 1, lorsque la consonne est conforme à la cible et 2, lorsqu'elle est presque conforme à la cible, à savoir produite avec une différence phonétique (ex. "plus important" réalisée [plysɛ̃pɔʁtɑ̃]). Lorsque la liaison est réalisée avec une consonne orthographique non finale mais présente dans le mot, le code 3 est utilisé (ex. "grand émoi" produit [gʁɑ̃nemwa]). Une liaison réalisée avec une consonne épenthétique non présente dans le mot liaisonnant est codée quant à elle en 4 (ex. "grand émoi" produit [gʁɑ̃gemwa]).

Le champ 7, composé d'un seul chiffre, permet d'indiquer la présence d'une pause, d'une hésitation ou d'un coup de glotte, sans toutefois différencier ces éléments entre eux. Les liaisons réalisées sans pause, hésitation ou coup de glotte entre le mot-1 et le mot-2 sont codées 0 pour ce champ, alors que celles produites avec pause, hésitation ou coup de glotte sont codées en 1.

Le codage de l'ensemble des champs se fait sous *Praat*, dans une tire dédiée. Ainsi, par exemple, la séquence "il est à Genève depuis un an", produite [i-lɛ-ta-ʒə-nɛv-də-pɥi-ɛ̃-nɑ̃] sera codée "il est31_AUX_PREP_100_11_1_0 à Genève depuis10_PREP_DET_200_00_0_0 un20_DET_NOM_1VN_11_1_0 an"[19]. Les 37 sites potentiels de liaison du texte PFC ainsi que le codage des quatre premiers champs (champs descriptifs) sont présentés dans l'Annexe (cf. note 17 pour la différence dans le nombre de sites potentiels de liaison du texte dans les projets PFC et IPFC).

Le codage est généralement effectué par un codeur natif et est au minimum vérifié par un deuxième évaluateur, pour autant qu'une procédure de double-

---

[19] Rappelons que cette séquence, avec le code PFC, était codée "il est11t à Genève depuis20 un11nVN an" (cf. section 2.4).

codage en aveugle ne soit pas appliquée. Le concordancier *Dolmen* permet de comparer aisément les codages effectués par les deux codeurs et de calculer leur degré de corrélation (ICC: "*intraclass correlation coefficient*") pour l'ensemble du codage ou pour chaque champ séparément. *Dolmen* interagissant avec *Praat*, les erreurs concernant les champs descriptifs peuvent ainsi facilement être corrigées, directement dans les fichiers textgrids. En revanche, en ce qui concerne les divergences dans les champs concernant l'évaluation perceptive, qui ne peuvent évidemment pas être considérées comme des erreurs, une procédure doit être mise en place (p. ex. recours à un troisième évaluateur qui tranche). Enfin, le décodage par le biais du concordancier *Dolmen* permet d'obtenir facilement des statistiques descriptives, telles que le nombre de liaisons en [n] réalisées, le nombre de liaisons entre un déterminant et un substantif, le nombre de liaisons produites avec enchaînement, etc.

6. Conclusion

Le code liaison IPFC a été diffusé aux différentes équipes du projet, qui l'ont appliqué à leurs données. L'intérêt de disposer d'une méthodologie commune (protocole de recueil de données, conventions de transcription, code et outil de décodage) est de rendre possible des comparaisons à grande échelle, qui peuvent prendre en compte différents paramètres tels que, par exemple, la population (en fonction de la L1, du niveau des apprenants ou du contexte d'apprentissage) ou la tâche effectuée (lecture *vs* conversation). Nous avons également vu que la plupart des études en L2 se basent sur la littérature pour établir des comparaisons avec les natifs, littérature qui, en ce qui concerne la liaison, peut diverger fortement d'une étude à l'autre, comme l'a relevé De Moras (2011). La proximité avec le projet PFC facilite donc aussi la comparaison avec les natifs. Ainsi, par exemple, des apprenants qui étudient le français à Genève ou à Québec peuvent directement être comparés avec des locuteurs genevois ou québécois, ce qui permet de prendre en compte les spécificités locales quand elles existent. De même, l'usage du même texte dans les deux projets rend les données directement comparables.

Sept contributions examinant les réalisations de la liaison par des apprenants de L1 variées (dans l'ordre du volume: italophones, germanophones, anglophones, hellénophones chypriotes, norvégophones, japonophones et hispanophones) sont présentées dans la suite de ce volume. Toutes les études ont été effectuées dans le cadre du projet IPFC et s'appuient donc sur la méthodologie exposée dans le présent article.

Nous espérons que ces études contribueront à approfondir les connaissances au niveau de l'apprentissage de la liaison en L2, en élargissant les populations d'apprenants à d'autres L1 que l'anglais. Cette diversification des travaux combinée à une meilleure prise en compte des usages réels en vigueur chez

les natifs, rendue possible par le travail conséquent issu de la linguistique de corpus, devraient permettre, à terme, un renouvellement adéquat des ressources pédagogiques pour l'enseignement de la liaison.

BIBLIOGRAPHIE

Abry, D. & Chalaron, M. (1994). *350 Exercices de phonétique*. Paris: Hachette.

Abry, D. & Chalaron, M. (2011). *Les 500 Exercices de phonétique*. Paris: Hachette.

Ågren, J. (1973). *Enquête sur quelques liaisons facultatives dans le français de conversation radiophonique*. Uppsala: Acta Universitatis Uspaliensis.

Andreassen, H. N., Herry-Bénit, N. & Kamiyama, T. (2015). The ICE-IPAC project: testing the protocol on Norwegian and French learners of English. *Proceedings of the workshop: "Phonetic Learner Corpora"*, ICPHS, Glasgow, 9-11.

Avanzi, M., Béguelin, M. J. & Diémoz, F. (à paraître). Corpus de français parlés et français parlés des corpus, *Cahiers Corpus*.

Barreca, G. & Christodoulides, G. (2015). Une analyse de la distribution de la liaison dans le corpus PFC. *Journées d'été PFC: "PFC: phonologie, corpus, méthode"*, Université de Vienne, 17-18 juillet 2015.

Boë, L.-J. & Tubach, J.-P. (1992). *De A à Zut: dictionnaire phonétique du français parlé*. Grenoble: Ellug.

Boersma, P. & Weenink, D. (2014). *Praat: doing phonetics by computer*. http://www.praat.org.

Bordal, G. & Lyche, C. (2008). La liaison en terre africaine. *Journées PFC 2008: variation, interfaces, cognition*, Paris, FMSH, 11-13 décembre 2008.

Boutin, B. A. (2014). Liaisons en français et terrains africains. In J. Durand, G. Kristoffersen & B. Laks et avec la collaboration de J. Peuvergne (éds.), *La phonologie du français: normes, périphéries, modélisation. Mélanges pour Chantal Lyche* (pp. 153-171). Presses universitaires de Paris Ouest.

Boutin, B. A. & Lyche, C. (2014). Ce que nous apprennent des locuteurs francophones non-lecteurs sur la liaison. In C. Soum-Favaro, A.-L. Coquillon & J.-P. Chevrot (éds.), *La liaison: approches contemporaines* (pp. 283-310). Berne: Peter Lang.

Carranza, M. (sous presse). Transcription and annotation of spontaneous non-native spoken corpora. In E. Martín-Monje, I. Elorza & B. García Riaza (éds.), *Technological Advances in Specialized Linguistic Domains: Learning on the Move*. London: Routledge.

Charliac, L. & Motron, A.-C. (2006). *Phonétique progressive du français* (niveau avancé). Paris: CLE International.

Coquillon, A.-L. & Durand, J. (2010). Le français méridional: éléments de synthèse. In S. Detey, J. Durand, B. Laks & C. Lyche (éds.), *Les variétés du français parlé dans l'espace francophone: ressources pour l'enseignement* (pp. 185-197). Paris: Ophrys.

Côté, M.-H. (2012). Laurentian French (Quebec): extra vowels, missing schwas and surprising liaison consonnants. In R. Gess, C. Lyche & T. Meisenburg (éds.), *Phonological variation in French: Illustrations from three continents* (pp. 235-274). Amsterdam: John Benjamins.

Côté, M.-H. (2013) Understanding cohesion in French liaison. *Language Sciences, 39*, 156-166.

Cylwik, N., Wagner, A. & Demenko, G. (2009). The EURONOUNCE corpus of non-native Polish for ASR-based Pronunciation Tutoring System. *Proceedings of SlaTE 2009 – 2009 ISCA Workshop on Speech and Language Technology in Education*. Birmingham, UK.

Delais-Roussarie, E. (2009). *Conventions CHAT de Transcription des données*. Document interne, BDD Interlangue, janvier 2009.

Delais-Roussarie, E., Santiago, F. & Yoo, H.-Y. (2015). The extended COREIL corpus: first outcomes and methodological issues. *Proceedings of the workshop: "Phonetic Learner Corpora"*, ICPHS, Glasgow, 57-59.

Delais-Roussarie, E. & Yoo, H.-Y. (2010). The COREIL corpus: a learner corpus designed for studying phrasal phonology and intonation. In K. Deziubalska-Kolaczyk, M. Wrembel. & M. Kul (éds.), *Proceedings of New Sounds2010* (pp. 100-105), Poznan, Poland.

Delattre, P. (1946). Pour imiter un disque de français parlé, *The French Review, 20 (1),* 43-48.

Delattre, P. (1951). *Principes de phonétique française à l'usage des étudiants anglo-américains.* Middlebury College.

De Jong, D. (1994). La sociophonologie de la liaison orléanaise. In C. Lyche (éd.), *French Generative Phonology: Retrospective and Perspectives* (pp. 95-130). AFLS/ESRI.

De Moras, N. (2011). *Acquisition de la liaison et de l'enchaînement en français L2: le rôle de la fréquence.* Ph.D. Dissertation, University of Western Ontario.

Detey, S. (2012). Coding an L2 phonological corpus: from perceptual assessment to non-native speech models – an illustration with French nasal vowels. In Y. Tono, Y. Kawaguchi & M. Minegishi (éds.), *Developmental and Crosslinguistic Perspectives in Learner Corpus Research* (pp. 229-250). Amsterdam and Philadelphia: John Benjamins.

Detey, S. (2014). Vers une évaluation par codage perceptif sur corpus de la production des liquides françaises /R/ et /l/ des apprenants japonais en singleton et en groupe consonantique. *Flambeau, 40*, 1-17.

Detey, S., Durand, J., Laks, B. & Lyche, C. (à paraître). *Varieties of Spoken French.* Oxford: Oxford University Press.

Detey, S. & Kawaguchi, Y. (2008). Interphonologie du Français Contemporain (IPFC): récolte automatisée des données et apprenants japonais. *Journées PFC: Phonologie du français contemporain: variation, interfaces, cognition,* Paris, 11-13 décembre 2008.

Detey, S. & Racine, I. (2012). Les apprenants de français face aux normes de prononciation: quelle(s) entrée(s) pour quelle(s) sortie(s)?, *Revue française de linguistique appliquée*, *17 (1)*, 81-96.

Detey, S. & Racine, I. (2014). Coder les consonnes dans IPFC: segments et syllabe. *Rencontres FLORAL 2014: Corpus oraux et enseignement de la prononciation en FLE & Interphonologie et corpus oraux,* Paris, 8-9 décembre 2014

Detey, S., Racine, I., Eychenne, J. & Kawaguchi, Y. (2014). Corpus-based L2 phonological data and semi-automatic perceptual analysis: the case of nasal vowels produced by beginner Japanese learners of French. *Proceedings of Interspeech2014.* Singapore: 539-544.

Detey, S., Racine, I. & Kawaguchi, Y. (2014). Des modèles prescriptifs à la variabilité des performances non-natives: les voyelles nasales des apprenants japonais et espagnols dans le projet IPFC. In J. Durand, G. Kristoffersen & B. Laks et avec la collaboration de J. Peuvergne (éds.), *La phonologie du français: normes, périphéries, modélisation. Mélanges pour Chantal Lyche* (pp. 197-226). Paris: Presses Universitaires de Paris Ouest.

Detey, S., Racine, I., Kawaguchi, Y. & Zay, F. (à paraître). Variation among non-native speakers: the Interphonology of Contemporary French. In S. Detey, J. Durand, B. Laks & C. Lyche (éds.), *Varieties of Spoken French.* Oxford: Oxford University Press.

Durand, J. (2009). On the scope of linguistics: data, intuitions, corpora. In Y. Kawaguchi, M. Minegishi & J. Durand (éds.), *Corpus Analysis and Variation in Linguistics* (pp. 25-52). Amsterdam and Philadelphia: John Benjamins.

Durand, J., Calderone, B., Laks, B. & Tchobanov, A. (2011). Que savons-nous de la liaison aujourd'hui?, *Langue française, 169*, 103-135.

Durand, J., Laks, B. & Lyche, C. (2002). Directions d'analyse. In J. Durand, B. Laks & C. Lyche (éds.), *Bulletin PFC1* (pp. 35-70). http://www.projet-pfc.net/bulletins-et-colloques/cat_view/918-bulletins-pfc/919-.html.

Durand, J., Laks, B. & Lyche, C. (2009). Le projet PFC (phonologie du français contemporain): Une source de données primaires structurées. In J. Durand, B. Laks & C. Lyche (éds.), *Phonologie, variation et accents du français* (pp. 19-62). Paris: Hermès.

Durand, J., Laks, B. & Lyche, C. (2014). French phonology from a corpus perspective: the PFC programme. In J. Durand, U. Gut & G. Kristoffersen (éds.), *The Oxford Handbook of Corpus Phonology* (pp. 486-497). Oxford: Oxford University Press.

Durand, J. & Lyche, C. (2003). Le projet Phonologie du français contemporain (PFC) et sa méthodologie. In E. Delais-Roussarie & J. Durand (éds.), *Corpus et variation en phonologie du français: méthodes et analyse* (pp. 213-276). Toulouse: Presses universitaires du Mirail.

Durand, J. & Lyche C. (2008). French liaison in the light of corpus data. *Journal of French and Language Studies*, *18 (1)*, 33-66.

Durand, J. & Lyche, C. (à paraître). Approaching variation in PFC: the liaison level. In S. Detey, J. Durand, B. Laks & C. Lyche (éds.), *Varieties of Spoken French.* Oxford: Oxford University Press.

Encrevé, P. (1988). *La liaison avec et sans enchaînement.* Paris: Seuil.

Eychenne, J., Lyche, C., Durand, J. & Coquillon, A.-L. (2014). Quelles données pour la liaison en français: la question des corpus. In C. Soum-Favaro, A.-L. Coquillon & J.-P. Chevrot (éds.), *La liaison: approches contemporaines* (pp. 33-60). Berne: Peter Lang.

Eychenne, J. & Paternostro, R. (à paraître). Analyzing transcribed speech with Dolmen. In S. Detey, J. Durand, B. Laks & C. Lyche (éds.), *Varieties of Spoken French*. Oxford: Oxford University Press.

Féry, J. (2003). *Liaison and syllable structure in French*. Manuscrit.

Fouché, P. (1959). *Traité de prononciation française.* Paris: Klincksieck. 2e édition.

Gut, U. (2005). Corpus-based pronunciation training. *Proceedings of Phonetics, Teaching and Learning Conference*, London.

Gut, U. (2009). *Non-native Speech: a Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German*. Wien: Peter Lang.

Han, M.H. & Eychenne, J. (à paraître). Les coréanophones. In S. Detey, I. Racine, Y. Kawaguchi & J. Eychenne (éds.), *La prononciation du français dans le monde: du natif à l'apprenant.* Paris: CLE international.

Harnois-Delpiano, M., Cavalla, C., & Chevrot, J.-P. (2012). L'acquisition de la liaison en L2: étude longitudinale chez des apprenants coréens de FLE et comparaison avec enfants francophones natifs. In F. Neveu, V. M. Toké, P. Blumenthal, T. Klingler, P. Ligas, S. Prévost & S. Teston-Bonnard (éds.), *Actes du 3ème Congrès Mondial de Linguistique Française* (pp. 1575-1589). Paris: ILF.

Howard, M. (2005). L'acquisition de la liaison en français langue seconde - Une analyse quantitative d'apprenants avancés en milieu guidé et en milieu naturel. *CORELA,* numéro spécial: Colloque AFLS, <http://corela.edel.univ-poitiers.fr/index.php?id=1127>.

Howard, M. (2013). La liaison en français langue seconde: une étude longitudinale préliminaire. *Language, Interaction, and Acquisition*, *4/2*, 190-231.

Laks, B. (2013). Diachronie de la liaison dans la parole publique. *Journées PFC 2013: "Phonologie du français contemporain: Regards croisés sur les corpus oraux"*, Paris, 5-7 décembre 2013.

Laks, B. (2015). Histoire de la liaison. Cours donné dans le cadre de l'école d'été PFC: "*PFC: phonologie, corpus, méthodes*", Université de Vienne, 13-18 juillet 2015.

Laks, B. & Calderone, B. (2014). La liaison en français contemporain: approches lexicales et exemplaristes. In C. Soum-Favaro, A.-L. Coquillon & J.-P. Chevrot (éds.), *La liaison: approches contemporaines* (pp. 79-109). Bern: Peter Lang.

Laks, B., Celata, C & Calderone, B. (2014). French liaison and the lexical repository. In C. Celata & S. Calamai (éds.), *Advances in sociophonetics* (pp. 30-52). Amsterdam: John Benjamins.

Landron, S., Paillereau, N., Nawafleh, A. & al. (2011). Le corpus PhoDiFLE: un corpus commun de français langue étrangère pour une étude phonétique des productions de locuteurs de langues maternelles plurielles. *CJC Praxiling2011*, Montpellier.

Lauret, B. (2007). *Enseigner la prononciation du français, questions et outils*. Paris: Hachette.

Léon, M. (2003). *Exercices systématiques de prononciation française*. Paris: Didier.

MacWhinney, B. (2015). *The CHILDES project: Tools for Analyzing Talk*, édition électronique consultée le 3 août 2015: http://childes.psy.cmu.edu/manuals/CHAT.pdf).

Mallet, G. (2008). *La liaison en français: description et analyses dans le corpus PFC*. Thèse de Doctorat, Université Paris Ouest Nanterre la Défense.

Mastromonaco, S. M. (1999). *Liaison in French as a Second Language*. PhD Dissertation, University of Toronto.

Matter, J. F. (1986). *A la recherche des frontières perdues: Etude sur la perception de la parole en français.* PhD Dissertation, University of Utrecht, Amsterdam: De Werelt.

Morin, Y.-C. (1986). On the morphologization of word-final consonant deletion in French. In H. Andersen (éd.), *Sandhi Phenomena in the Languages of Europe* (pp. 167-210). Berlin: Mouton de Gruyter.

Morin, Y.-C. & Kaye, J. (1982). The syntactic bases for French liaison. *Journal of Linguistics*, *18*, 291-330.

Milroy, J. (1980). *Language and social networks*. Oxford: Blackwell.

Neri, A., Cucchiarini, C. & Strik, H. (2006). Selecting segmental errors in non-native Dutch for optimal pronunciation training, *IRAL - International Review of Applied Linguistics in Language Teaching*, *44*, 357-404

Paternostro, R. (2014). L'éveil à la variation phonétique en Français Langue Etrangère: enjeux et outils, *LIDIL*, *50*, 105-124.

Pustka, E. (2011). *Einführung in die Phonetik und Phonologie des Französischen*. Berlin: Erich Schmidt Verlag.

Racine, I. (2008). *Les effets de l'effacement du schwa sur la production et la perception de la parole en français*. Thèse de doctorat, Université de Genève.

Racine, I. (2012). Spanish learners' productions of French close rounded vowels: a corpus-based perceptual study. In Y. Tono, Y. Kawaguchi & M. Minegishi (éds.), *Developmental and Crosslinguistic Perspectives in Learner Corpus Research* (pp. 205-228). Amsterdam/Philadelphia: John Benjamins.

Racine, I. (2014). Une approche par corpus de la liaison chez les apprenants hispanophones de français langue étrangère: quelles conséquences pour l'enseignement du FLE?, *Flambeau*, *40*, 18-37.

Racine, I. & Detey, S. (à paraître). La liaison dans un corpus d'apprenants: le projet "Interphonologie du français contemporain" (IPFC). In M. Avanzi, M.-J., Béguelin & F. Diémoz (éds.), *Corpus de français parlés et français parlés des corpus*, Cahiers Corpus.

Racine, I., Detey, S. & Andreassen, H. (en préparation). *French schwa alternation for non-native speakers: A representational or a processing problem.*

Racine, I., Detey, S., Zay, F. & Kawaguchi, Y. (2012). Des atouts d'un corpus multitâches pour l'étude de la phonologie en L2: l'exemple du projet "Interphonologie du français contemporain" (IPFC). In A. Kamber & C. Skupiens (éds.), *Recherches récentes en FLE* (pp. 1-19). Berne: Peter Lang.

Racine, I., Durand, J. & Andreassen, H. (à paraître). PFC, codages et représentations: la question du schwa. In M. Avanzi, M.-J., Béguelin & F. Diémoz (éds.), *Corpus de français parlés et français parlés des corpus*, Cahiers Corpus.

Racine, I., Zay, F., Detey, S. & Kawaguchi, Y. (2011). De la transcription de corpus à l'analyse interphonologique: enjeux méthodologiques en FLE. In G. Col & S.N. Osu (éds.), *Transcrire, écrire, Formaliser (1)* (pp 13-30). Rennes: PUR.

Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.

Schmid, H. (1995). Improvements in part-of-speech tagging with an application to German. *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland.

Shoemaker, E. (2009). *Acoustic Cues to Speech Segmentation in Spoken French: Native and Non-native Strategies.* PhD Dissertation, University of Texas.

Soum-Favaro, C., Coquillon, A.-L. & Chevrot, J.-P. (éds.) (2014). *La liaison: approches contemporaines*. Bern: Peter Lang

Stridfeldt, M. (2005). *La perception du français oral par des apprenants suédois.* Thèse de Doctorat, Université de Umeå.

Thomas, A. (1998). La liaison et son enseignement: des modèles orthoépiques à la réalité linguistique, *Canadian Modern Language Review, 54/4*, 543-552.

Thomas, A. (2002). La variation phonétique en français langue seconde au niveau universitaire avancé, *AILE, 17,*101-121.

Thomas, A. (2010). La complexité en FLE2 universitaire avancé. In U. Paprocka-Piotrowska, C. Martinot & S. Gerolimich (éds.), *Actes du colloque La complexité en langue et son acquisition* (pp. 149-152). Paris: Université Descartes.

Visceglia, T., Tseng, C.-Y., Kondo, M., Meng, H. & Sagisaka, Y. (2009). Phonetic aspects of content design in AESOP (Asian English Speech cOrpus Project). *Proceedings of Oriental-COCOSDA*, Urumuqi, Chine.

Wauquier, S. (2009). Acquisition de la liaison en L1 et L2: stratégies phonologiques ou lexicales?, *Phonétique, bilinguisme et acquisition, AILE*, 2, 93-130.

Wauquier, S. & Shoemaker, E. (2013). Convergence and divergence in the acquisition of French liaison by native and non-native speakers: a review of existing data and avenues for future research, *Language, Interaction, and Acquisition*, *4/2*, 161–189.

Zimmerer, F., Trouvain, J. & Bonneau, A. (2015). One corpus, one research question, three methods: German vowels produced by French speakers. *Proceedings of the workshop: "Phonetic Learner Corpora*", ICPHS, Glasgow, 25-27.

Annexe

Texte PFC-IPFC avec indication des 37 sites de liaisons potentielles et pré-codage à l'aide du code liaison IPFC (pour les quatre champs descriptifs)

Le Premier Ministre ira-t-il à Beaulieu?

Le village de Beaulieu est30_AUX_PRP_100_XX_X_X en grand30_ADJ_NOM_100_XX_X_X émoi. Le Premier Ministre a en20_EXF_EXF_1XX_XX_X_X effet décidé de faire étape dans cette commune au cours de sa tournée de la région20_NOM_PRP_2XX_XX_X_X en fin d'année. Jusqu'ici les seuls titres de gloire de Beaulieu étaient son vin blanc sec, ses chemises11_NOM_PRP_200_XX_X_X en soie, un champion local de course à pied (Louis Garret), quatrième aux jeux11_EXF_EXF_100_XX_X_X olympiques de Berlin20_NAM_PRP_2XX_XX_X_X en 1936, et plus récemment, son20_DET_NOM_1XX_XX_X_X usine de pâtes11_NOM_ADJ_100_XX_X_X italiennes. Qu'est-ce qui a donc valu à Beaulieu ce grand30_ADJ_NOM_100_XX_X_X honneur? Le hasard, tout bêtement, car le Premier Ministre, lassé des circuits11_NOM_ADJ_200_XX_X_X habituels qui tournaient toujours10_ADV_PRP_200_XX_X_X autour des mêmes villes, veut découvrir ce qu'il appelle la campagne profonde.

Le maire de Beaulieu - Marc Blanc – est30_AUX_PRP_100_XX_X_X en revanche très10_ADV_ADJ_100_XX_X_X inquiet. La cote du Premier Ministre ne cesse de baisser depuis les11_DET_NOM_100_XX_X_X élections. Comment30_ADV_PRP_200_XX_X_X, en plus, éviter les manifestations qui ont31_AUX_PPA_100_XX_X_X eu tendance à se multiplier lors des visites11_NOM_ADJ_200_XX_X_X officielles? La côte escarpée du Mont Saint-Pierre qui mène au village connaît des barrages chaque fois que les11_DET_NOM_100_XX_X_X opposants de tous les bords manifestent leur colère. D'un20_DET_ADJ_1XX_XX_X_X autre côté, à chaque voyage du Premier Ministre, le gouvernement prend contact avec la préfecture la plus proche et s'assure que tout30_PRO_AUX_100_XX_X_X est fait pour le protéger. Or, un gros détachement de police, comme on20_PRO_PRO_1XX_XX_X_X en20_PRO_AUX_1XX_XX_X_X a vu à Jonquière, et des vérifications d'identité risquent de provoquer40_INF_DET_200_XX_X_X une explosion. Un jeune membre de l'opposition20_NOM_AUX_2XX_XX_X_X aurait déclaré: "Dans le coin20_NOM_PRO_1XX_XX_X_X, on20_PRO_AUX_1XX_XX_X_X est jaloux de notre liberté. S'il faut montrer patte blanche pour circuler, nous ne répondons pas de la réaction des gens du pays. Nous11_PRO_AUX_100_XX_X_X avons le soutien du village entier." De plus, quelques11_DET_NOM_XXX_XX_X_X articles parus dans La Dépêche du Centre, L'Express, Ouest Liberté et Le Nouvel Observateur indiqueraient que des11_DET_NOM_100_XX_X_X activistes des communes voisines préparent31_VER_DET_200_XX_X_X une journée chaude au Premier Ministre. Quelques fanatiques11_NOM_AUX_200_XX_X_X auraient même entamé un jeûne prolongé dans l'église de Saint-Martinville.

Le sympathique maire de Beaulieu ne sait plus10_ADV_PRP_100_XX_X_X à quel saint se vouer. Il a le sentiment de se trouver dans10_PRP_DET_100_XX_X_X une impasse stupide. Il s'est30_AUX_PRP_100_XX_X_X, en désespoir de cause, décidé à écrire au Premier Ministre pour vérifier si son village était vraiment30_ADV_DET_200_XX_X_X une étape nécessaire dans la tournée prévue. Beaulieu préfère être inconnue et tranquille plutôt que de se trouver40_INF_PDE_200_XX_X_X au centre d'une bataille politique dont, par la télévision, seraient témoins des millions d'électeurs.