

Variabilité de la liaison facultative en français standard

Mathilde Hutin¹, Yaru Wu¹, Nori Kondo², Samantha Ruvoletto³, Ioana Vasilescu¹, Lori Lamel¹, Martine Adda-Decker^{1,4}

(1) Université Paris-Saclay, CNRS, UPR 3251 LIMSI, Bât. 507, rue du Belvédère, 91405 Orsay, France

(2) Nagoya University of Foreign Studies, 57, Takenoyama, Iwasakicho, 470-0197, Nisshin, Japon

(3) Université de Lorraine, CNRS / UL, UMR 7118 ATILF, 23 Boulevard Albert 1er, 54000 Nancy, France

(4) Université Paris 3 Sorbonne Nouvelle, CNRS, UMR 7018, LPP, 19 rue des Bernardins, 75005 Paris, France

La liaison est un phénomène de sandhi externe très fréquent en français. Les éléments impliqués dans le processus de liaison sont deux mots (Mot1 et Mot2) et une consonne de liaison (CL). Cette consonne latente peut apparaître entre deux mots si Mot2 commence par une voyelle, comme dans *les + amis* = [le.za.mi], mais pas dans *les + copains* = [le.ko.pɛ̃].

La cohésion syntaxique entre Mot1 et Mot2 est posée comme déterminante pour la réalisation de la liaison : sans cohésion syntaxique, pas de liaison. Les études des catégories syntaxiques possibles des Mot1 et Mot2 ont permis une première classification faite par Delattre (1947) qui distingue trois catégories de liaison : « obligatoire », « facultative » et « interdite ».

Dans ce projet, nous nous concentrons sur le phénomène de la liaison facultative. Elle se définit par le fait que le locuteur peut prononcer la séquence de deux mots soit avec liaison, soit sans liaison. Ni la production de la liaison ni son omission ne sont agrammaticales. Notre objectif est de confronter les connaissances actuelles sur la liaison facultative à de grands corpus alignés automatiquement et de développer une recherche plus théorique sur la représentation de la CL et de la variation.

En effet, plusieurs points restent décrits de façon incomplète dans la littérature. Pour commencer, les analyses du corpus PFC (Durand et al. 2011) ont amené à remettre en question les appellations comme « obligatoire » ou « facultative ». Dans des contextes jusqu'à présent reconnus comme étant des contextes de liaison obligatoire, notamment le contexte « Mot invariable monosyllabique + X », Mallet (2008) ou encore Adda-Decker et al. (2012) montrent que la liaison n'est que *presque* toujours réalisée ; et Durand & Lyche (2008) remarquent que, dans la catégorie définie par Delattre comme « obligatoire », il y a de la *variation* dans la réalisation dépendant de phénomènes morphologiques et catégoriels. Quant aux facteurs stylistiques, Delattre (1947) note une forte différence dans la fréquence des liaisons variables selon les registres utilisés par les locuteurs. Entre le registre familier et le registre plus soutenu, la fréquence de production des liaisons variables augmente. Les trois modalités prévues dans le protocole PFC permettent également d'avoir des données sur des « styles » différents, plus ou moins surveillés, et Eychenne et al. (2014) montrent qu'un pourcentage plus élevé de liaison est produit en lecture, puis en conversation guidée, puis en conversation spontanée. Concernant les facteurs individuels qui influencent la réalisation de la liaison variable, des études (Malécot 1975, Ashby 1981) montrent que les locuteurs plus âgés réalisent plus de liaisons variables, surtout en conversation spontanée. En ce qui concerne le sexe des locuteurs, les résultats sont controversés : les analyses de Malécot (1975) ou Booij et De Jong (1987) montrent un usage plus fréquent de la liaison chez les femmes alors qu'Ashby (1981) trouve une réalisation plus fréquente chez les hommes, là où les recherches plus récentes montrent qu'il n'y a pas de différence statistiquement significative entre les deux sexes (Durand et al. 2011).

Dans ce projet, nous souhaitons affiner l'état des lieux sur la liaison variable en français standard à l'aide de trois grands corpus manuellement transcrits orthographiquement (sans présupposition sur la réalisation de la liaison). Le corpus ESTER (Galliano et al. 2005) comprend originellement 80h de discours journalistique que nous avons filtrées pour enlever les données en français non-métropolitain (RFI et RTM) et ne retenir que les quelque 40h en français standard.

Le corpus ETAPE (Gravier et al. 2012) contient 13h30 de radio et 29h de télévision, notamment des débats et des conversations. Enfin, NCCFr (Torreira et al. 2010) contient 31h d'interaction de visu, spontanée, entre amis. Nos trois grands corpus représentent donc chacun un style de parole différent : ESTER est caractéristique du discours préparé, soigné, et potentiellement lu à haute voix ; ETAPE de monologues semi-préparés et de conversations à deux ou plusieurs interlocuteurs ; et NCCFr de discours spontané entre amis. Ils ont l'avantage, par rapport à PFC, de représenter de la parole plus naturelle que celle qu'on peut récolter lors d'enquêtes linguistiques, englobant un échantillon intéressant de situations variées de communication reflétant la parole « vivante », entraînant une richesse lexicale, syntaxique, thématique... qui permet de produire des résultats contrastifs, et donc plus complets et plus détaillés.

Comme ces données ne sauraient être segmentées à la main, elles ont été traitées suivant la méthode décrite dans Hallé & Adda-Decker (2011). Un système de reconnaissance automatique de la parole du français (Gauvain 2002) a été utilisé pour forcer un alignement pour lequel des variantes avec et sans consonnes de liaison étaient possibles. Par exemple, la construction *trop irrité* pouvait être alignée avec les transcriptions [tʁoʁiʁite] ou [tʁopiʁite] selon que la machine estimait qu'une consonne était réalisée ou non à la fin de Mot1. Il sera alors possible, sur des dizaines de milliers d'occurrences estimées, de faire un état des lieux approfondi de la liaison variable en français standard et des facteurs de sa réalisation, qu'ils soient d'ordre linguistique (phonétiques, phonologiques, phonotactiques, morpho-syntaxiques...) ou d'ordre sociolinguistique (le style de parole, évidemment, mais aussi et surtout le sexe des locuteurs, dont l'effet n'a pas été bien démontré dans la littérature). La comparaison de deux corpus journalistiques – ETAPE (années 2000) et les données d'Ågren (1973) – nous permettra également d'établir si et à quel point la liaison variable a subi un changement diachronique.

Nous présenterons à la conférence les résultats préliminaires de cette étude de grande ampleur. Une telle étude aura, à terme, un double impact : compléter l'état de nos connaissances sur la liaison variable en français grâce à des données dont la quantité et la diversité sont inégalées, et initier une recherche plus théorique sur les phénomènes de sandhi et la représentation de la variation.

Bibliographie

- Adda-Decker, M., Fougeron, C., Gendrot, C., Delais-Roussarie, E. & Lamel, L. 2012. La liaison dans la parole spontanée familière : une étude sur grand corpus. *Revue française de linguistique appliquée*, vol. xvii(1), 113-128
- Ågren, J. 1973. *Étude sur quelques liaisons facultatives dans le français de conversation radiophonique : fréquence et facteurs*. Uppsala : Acta Universitatis Upsaliensis.
- Ashby, W. J. 1981. French liaison as a sociolinguistic phenomenon. In W. W. Cressy & D. J. Napoli (ed.). *LSRL 9*. Washington DC: Georgetown University Press, 46-57
- Booij, G. & De Jong, D. 1987. The domain of liaison: theories and data. *Linguistics*, 25, 1005-1025.
- Delattre, P. 1947. La liaison en français, tendances et classification. *The French Review*, 21 (2), 148-157.
- Durand, J., Laks, B., Calderone, B. & Tchobanov, A. 2011. Que savons-nous de la liaison aujourd'hui ? *Langue Française*, 169, 1, 103-126.
- Durand, J. & Lyche, C. 2008. French liaison in the light of corpus data. *Journal of French Language Studies*, 18 (1), 33-66.
- Durand, J., Laks, B. & Lyche, C. 2002. La phonologie du français contemporain : usages, variétés et structure. In C. Pusch & W. Raible (éds), *Romanistische Korpuslinguistik- Korpora und gesprochene Sprache/Romance Corpus Linguistics – Corpora and Spoken Language*. Tübingen : Gunter Narr Verlag, 93-106.
- Eychenne, J., Lyche, C., Durand, J. & Coquillon, A. 2014. Quelles données pour la liaison aujourd'hui : la question des corpus, dans C. Soum, A. Coquillon & J.-P. Chevrot (eds) *La liaison : approches contemporaines*, Bern: Peter Lang, 33-60.
- Galliano, S., Geoffrois, E., Mostefa, D., Choukri, K., Bonastre, J.-F. & Gravier, J. 2005. ESTER Phase II Evaluation Campaign for the Rich Transcription of French Broadcast News. *Interspeech 2005*, 2453-2456
- Gauvain, J.-L., Lamel, L. & Adda, G. 2002. The LIMSI broadcast news transcription system. *Speech communication* 37 (1-2), 89-108

- Gravier, G., Adda, G., Paulson, N., Carré, M., Giraudel, A. & Galibert, O. 2012. The ETAPE corpus for the evaluation of speech-based TV content processing in the French language. *LREC Eighth international conference on Language Resources and Evaluation*.
- Hallé, P. & Adda-Decker, M. 2011. Voice assimilation in French obstruents: A gradient or a categorical process? *Tones and features: A festschrift for Nick Clements*, De Gruyter, 149–175.
- Malécot, A. 1975. French liaison as a function of grammatical, phonetic and paralinguistic variables. *Phonetica*, 32, 161-179.
- Mallet, G. 2008. *La liaison en français : descriptions et analyses dans le corpus PFC*. Thèse de doctorat, Université Paris Ouest-Nanterre-La Défense.
- Torreira F., Adda-Decker M. & Ernestus M. 2010. The Nijmegen Corpus of Casual French. *Speech Communication*, Elsevier: North-Holland, 2010, 52 (3)