


ARTICLE

La liaison dans un corpus de hits francophones (1956–2017)

Gasparde Coutanson^{a,*}  and Flora Badin^b

^aUniversité Paris Nanterre and ^bUniversité d'Orléans

*Email: gasparde.coutanson@parisnanterre.fr

(Received 11 February 2020; revised 13 January 2021; accepted 20 January 2021)

RÉSUMÉ

Dans cette contribution, nous analysons un corpus de chansons ayant accédé à la première place de classements musicaux réalisés pour la France entre 1956 et 2017. Cette base de données nous permet de nous questionner sur le type d'oralité que produit l'industrie du disque francophone. Si les chansons présentent des caractéristiques conceptionnelles relevant de la distance (Koch et Oesterreicher, 2001), quelques traits linguistiques relevés sont pourtant illustratifs d'une proximité conceptionnelle. Nous nous intéressons plus particulièrement aux liaisons dans les chansons, à leur possible réalisation dans ces productions préparées en comparaison avec des corpus spontanés ou de locuteurs professionnels, aux liaisons variables (que nous considérons comme une marque de distance) et à l'évolution diachronique de leur réalisation. Pour mener à bien cette analyse, nous avons utilisé un outil facilitant l'étude sur corpus des liaisons, que nous présenterons. C'est surtout la réalisation de certaines liaisons variables, comme après *suis*, *étais* ou *jamais*, plus spécifiquement réalisées dans notre corpus, qui concourt au caractère distant de ces productions langagières. Au fil du temps, la réalisation des liaisons variables a diminué dans les chansons, par exemple après *est* ou préposition, les rapprochant de productions relevant de la proximité.

INTRODUCTION

Les chansons sont des productions linguistiques qui se positionnent de façon toute particulière entre oralité et scripturalité. Söll (1974) a tout d'abord proposé de distinguer l'*aspect médial* (le canal utilisé : *phonique* vs *graphique*) de l'*aspect conceptionnel* (l'allure linguistique d'une production : *oral* vs *écrit*). Koch et Oesterreicher (2001) ont par la suite défendu l'idée que les diverses productions linguistiques peuvent être replacées sur un continuum conceptionnel entre *immédiateté* (ou *proximité*) et *distance*. Si les chansons sont phoniques, elles présentent principalement des caractéristiques communicationnelles relevant de la distance : détachement actionnel et situationnel, communication publique, interlocuteur inconnu, communication préparée, monologue (selon les paramètres de Koch et Oesterreicher, 2001: 586). Au niveau linguistique, notre corpus se

caractérise également par une absence de disfluences (particules d'hésitation, amorces, reformulations, réparations) mais présente pourtant d'autres caractéristiques relevant de la proximité, que ce soit au niveau phonétique : réductions (par exemple de [ʒəsɥi] à [ʃɥi] ou [ʃy]), simplification de groupes consonantiques (dont des non-réalisations de la consonne finale des pronoms sujets *il* et *elle* devant initiale consonantique) ou encore non-réalisation de voyelles (comme le [e] de *déjà* prononcé [dʒa]) ; au niveau morphosyntaxique : utilisation du suffixe *-zer*,¹ mots tronqués (comme *Lambo* pour *Lamborghini*), non-réalisation du *ne* dans la négation, du *il* à valeur impersonnelle dans *il y a* ou *il faut*... Certains faits linguistiques venant d'être exposés comme relevant de la variation entre oral et écrit, ou variation diamésique, peuvent également être propres à d'autres dimensions, celles-ci se combinant entre elles. Pour la dimension socioculturelle (variation diastratique), les sociolectes des catégories sociales défavorisées tendraient vers la proximité alors que les sociolectes des catégories sociales favorisées tendraient vers la distance. Pour la dimension géographique (variation diatopique), les diverses variétés régionales tendraient également vers la proximité tandis que pour la dimension situationnelle (variation diaphasique), les registres les moins soutenus tendraient vers la proximité et les plus soutenus vers la distance. L'utilisation de prononciations et termes régionaux ou argotiques tirerait ainsi les productions du côté de la proximité conceptionnelle. Les règles de versification classiques propres au français ne sont également pas toujours respectées, notamment au niveau de la réalisation des schwas (habituellement tous prononcés, excepté en fin de vers, devant ou après voyelle phonique) et des liaisons (réalisation d'un maximum de liaisons variables). Ces consonnes à la frontière entre deux mots (mot 1, M1 et mot 2, M2) pouvant être produites devant initiale vocalique mais ne l'étant pas devant initiale consonantique ou lorsque les mots sont prononcés en isolation, constituent ici notre objet d'étude. Si le phénomène est phonique, la réalisation de nombreuses liaisons variables est caractéristique de la distance en cela que ce paramètre prestigieux dénote une certaine maîtrise du code graphique (voir néanmoins Hornsby, 2019 pour l'apport de données sur la liaison à l'hypothèse diglossique du français). Comme leur nom l'indique, les liaisons variables sont variablement réalisées, contrairement aux liaisons invariables qui sont généralement toujours réalisées par les locuteurs francophones natifs.

Dans une perspective fondée sur l'usage, la grammaire émerge de l'expérience linguistique d'un locuteur en production comme en réception et influence les productions de ce locuteur ; la grammaire émane donc de l'usage tout comme elle détermine l'usage (Bybee, 2006). L'importance du rôle d'auditeur était déjà notée avant l'émergence de la linguistique cognitive par les sociolinguistes : 'la langue d'un sujet, contrairement au jugement commun, ce n'est pas la langue qu'il parle, c'est la langue qu'il entend' (Encrevé, 1977: 6). Cela nous a amenées à nous intéresser à un corpus de chansons. Cet attrait pour le groupe social des chanteurs de variété découle surtout du caractère public de leurs productions,

¹-zer serait originaire de Grigny (Bidan et Cuccagna, 2016), en banlieue parisienne. Il a été popularisé par le rappeur Booba (ex : *la hainezer que je trainezer est insoignable*, OKLM, Booba, 2014 [2014]).

et plus qu'une simple influence entre pairs au sein d'un groupe d'appartenance, c'est ici la possible influence des chanteurs sur les auditeurs francophones qui nous a semblée pertinente. La musique est en effet aujourd'hui un objet culturel du quotidien qui passe souvent par une pratique d'écoute, privée ou collective (Maisonneuve, 2009: 13). La pratique d'écoute privée est aujourd'hui massive puisque selon un sondage IPSOS (Le Gorju, 2014), 99% des Français disent écouter de la musique, environ 17h par semaine et indiquent que leur genre préféré est la variété française. Les Français sont en outre confrontés à la musique dans l'espace public, dans les commerces et au travail (71% des commerçants diffusent de la musique dans leur établissement, touchant à la fois les clients et les employés ; Le Gorju, 2014). Notons qu'il ne semble pas exister à l'heure actuelle d'étude portant sur la liaison dans les chansons, excepté celle de Nardy et al. (2014) portant sur les enfantines en synchronie, ou 'genres oraux utilisés par les enfants au cours de leurs jeux chantés : chansons, formulettes, comptines, devinettes, charades, etc.' (Chauvin et Colletta, 2003: 40). Comme l'avaient proposé ces auteurs : 'une séquence des enfantines incluant un contexte de liaison peut fournir le matériau constitutif d'un schéma à la condition que cette séquence soit aussi suffisamment fréquente dans les échanges quotidiens' (Nardy et al., 2014: 258).

Notre première partie présentera un état de l'art en matière d'étude diachronique de la liaison. Nous présenterons dans une seconde partie notre corpus et dans une troisième la méthodologie adoptée pour transcrire et annoter nos données ainsi que l'outil que nous avons développé pour l'étude de la liaison. Dans une quatrième partie, nous observerons les pratiques de nos chanteurs en matière de liaison. L'étude de leur évolution diachronique nous permettra enfin de modérer nos observations.

Nous souhaitons comparer la réalisation des liaisons dans notre corpus avec celle des liaisons de corpus oraux spontanés, car les chansons étant des productions préparées, la façon de liasonner a pu être pensée en amont, ce qui pourrait impacter l'authenticité de ces productions. Nous voulons en outre observer dans quelle mesure les liaisons variables, indicatrices de distance, sont produites : les chanteurs souhaitent-ils se démarquer de productions orales spontanées ou au contraire s'en rapprocher par souci d'authenticité ?

LIAISON ET DIACHRONIE

L'enregistrement sonore constitue une avancée technologique relativement récente, de la sorte, il existe encore peu de corpus oraux diachroniques pour le français. Les données orales les plus anciennes ont été enregistrées selon deux perspectives, premièrement, conserver les productions de grands hommes et femmes d'une époque (personnalités politiques, écrivains, artistes, inventeurs, chanteurs), deuxièmement, préserver la mémoire de pratiques linguistiques (dialectologiques) et sociales (ethnographiques, musicologiques) menacées de disparition (Descamps, 2005: 71). Pour accéder à des données sonores anciennes pour le français, il faut donc soit se pencher sur les productions de professionnels de la parole (pour la plupart), ce sera notre cas, soit s'intéresser à des productions non standard ou provenant d'une autre discipline. En ce qui concerne l'étude diachronique de la

liaison, seules les personnalités politiques ont été étudiées (Encrevé, 1983 ; Laks, 2009 ; Laks et Peuvergne, 2017), bien que des journalistes aient aussi fait l'objet d'études en synchronie. Les chanteurs, aux côtés des professeurs, acteurs ou encore personnels administratifs pourraient également avoir un impact sur les productions des locuteurs. Des études ont aussi porté sur la liaison en diachronie chez les locuteurs des villes de Tours (Ashby, 2003) et d'Orléans (Dugua et Baude, 2017), mais nous n'avons pas de grand corpus oral pour le français antérieur aux années 1960. Les études diachroniques portant sur l'oral restent de la sorte micro-diachroniques et ne nous permettent d'étudier que des tendances récentes. Sur le long terme, on ne peut nier une évolution diachronique, puisque ce sont d'anciennes consonnes finales qui ont évolué vers les liaisons actuelles. Déjà en latin mérovingien, on observe des cas de sandhis externes consonantiques prenant la forme d'alternances conditionnées phonologiquement (Russo, 2014). Mais la tendance n'a pas toujours été à la diminution, puisque l'influence de l'écrit (Durand et al., 2011: 111) et de l'analogie (Morin, 1982: 28) ont pu favoriser la prononciation de liaisons auparavant non réalisées. En l'absence de données empiriques, l'idée reçue selon laquelle la liaison serait actuellement en déclin a souvent été reprise (Léon, 1992: 246 ; Chigarevskaïa, 1973: 162 ; Thomas, 1998: 546). Les corpus oraux micro-diachroniques existants convergent en revanche dans la direction opposée : depuis le siècle dernier, le taux global de liaisons aurait peu évolué, que ce soit chez les personnalités politiques (Laks, 2009: 243 ; Laks et Peuvergne, 2017: 62) ou les locuteurs orléanais (Dugua et Baude, 2017: 51). Ashby (2003) n'apporte aucune indication à ce sujet.

Malgré cela, ces études antérieures ont montré que cette stabilité apparente cachait certaines évolutions au niveau de la catégorie des liaisons variables tout comme des évolutions individuelles différentes. Notons que dans l'étude portant sur les enfantines, Nardy et al. (2014) avaient observé une réalisation élevée de liaisons variables et une variation interindividuelle quasi inexistante entre les différentes versions d'une même enfantine. Le changement linguistique peut être appréhendé de deux façons. Une première concerne les études en temps réel, dans lesquelles on s'intéresse au changement dans la communauté linguistique en divers points temporels (Labov, 1994: 73). Ashby (2003), en comparant deux corpus de français tourangeau enregistrés à vingt ans d'écart, note une diminution du taux de liaison variable entre 1976 (586/1617, 36%) et 1995 (290/1034, 28%). Laks (2009: 245) met quant à lui en évidence pour le corpus HPOL1 (personnalités politiques françaises, 1908–1998) une fluctuation du taux de liaisons variables à la hausse comme à la baisse : la décroissance régulière s'est en effet inversée dans les années 1970, au même moment où le taux de liaisons non enchaînées² a atteint son pic. Cette fluctuation est aussi visible dans les trajectoires individuelles des présidents de la République de la période étudiée (Laks, 2009: 249–251). Enfin, si la liaison variable après des M1 polysyllabiques ou singuliers aurait décliné, après des M1 monosyllabiques ou pluriels elle se

²Les liaisons non enchaînées sont des liaisons prononcées en coda de la dernière syllabe du M1 (habituellement les liaisons sont enchaînées, c'est-à-dire qu'elles sont prononcées à l'initiale de la première syllabe du M2) ; elles ont été observées pour la première fois par Encrevé (1983).

serait renforcée (Laks, 2009: 256–257). L’entreprise la plus difficile reste encore de parvenir à obtenir un corpus oral diachronique de mêmes locuteurs, pari relevé pour sept locuteurs orléanais enregistrés à quarante ans d’intervalle entre les années 1970 et les années 2010 dans le cadre d’un sous-corpus des ESLO (*Enquêtes sociolinguistiques à Orléans*, voir LLL, 2017 et Eshkol-Taravella et al., 2011). Globalement, la variation interindividuelle au niveau des taux de réalisation des liaisons variables est moindre dans le second volet du corpus comparé au premier, ce qui laisse dire aux auteurs qu’il pourrait y avoir une plus grande convergence des usages dans les pratiques récentes (Dugua et Baude, 2017: 51). A quarante ans d’intervalle, et avec les trajectoires sociales des locuteurs à disposition, on peut également observer des évolutions individuelles des taux de liaisons variables réalisées (Dugua et Baude, 2017: 52). Les études en temps apparent constituent une deuxième façon d’appréhender des données orales diachroniques. Pour celles-ci, on s’intéresse au changement linguistique entre les différents groupes d’âge de la communauté linguistique (Labov, 1994: 28). Laks et Peuvergne (2017: 65–66) ont notamment montré que pour le corpus HPOL2 (personnalités politiques françaises, 1999–2015), la réalisation des liaisons variables était moindre au fil des générations, changement qui serait mené par les femmes. Ceci est contraire aux résultats de l’étude d’Ashby (2003) en temps réel où le changement semble provenir des hommes³. Laks et Peuvergne (2017: 69–70) indiquent également que la réalisation des liaisons variables de certaines constructions spécifiques peut aussi fluctuer en temps apparent (par exemple, la liaison après *pas* est moins réalisée au fil des générations).

Enfin, une autre dimension que la dimension diachronique peut entrer en jeu dans le changement linguistique. Laks (2009: 245), en s’intéressant aux liaisons variables dans quatre corpus : HPOL1, PFC (*Phonologie du français contemporain*, corpus variationniste recueilli dans l’espace francophone,⁴ 1996–2015 ; voir MoDyCo et RUG, 2017 et Durand et al., 2009), VIL (adolescents issus de classes populaires, Villejuif, 1975) et BRU (ouvriers parisiens, 1912–1913), note par exemple « le rôle dominant des facteurs diastratiques et diaphasiques sur les facteurs proprement diachroniques » : diastratiques, dans le sens où les professionnels de la parole publique que sont les personnalités politiques (HPOL1) liasonneraient plus que les ouvriers (BRU), les adolescents issus de classes populaires (VIL) ou plus généralement les locuteurs d’une interaction spontanée (PFC) et diaphasiques, dans le sens où ils liasonneraient pourtant autant que des locuteurs en lecture (PFC), en opposition aux locuteurs d’interactions plus spontanées (VIL, BRU, conversations PFC). On pourrait encore voir là de la variation diamésique, avec une influence de l’écrit sur les productions orales, car pour reprendre un des arguments avancés par Koch et Oesterreicher (2001: 586), la parole politique est souvent préparée.

³Il semblerait notamment que cela soit le fait des hommes des catégories sociales supérieures, bien qu’il y ait un problème dans les légendes des tableaux du document, les résultats étant présentés deux fois pour les « locuteurs de la classe “A” (supérieure) ».

⁴Le protocole d’enquête de PFC comprend deux tâches de lecture (d’une même liste de mots et d’un même texte) et l’enregistrement de deux conversations (une dite guidée et une dite libre).

CORPUS DE TOPS 1

Nous avons constitué un corpus à partir de chansons ayant accédé à la première place de classements musicaux réalisés pour la France entre 1956 et 2017. Le choix de ces chansons nous a semblé pertinent en tant qu'input linguistique pour les auditeurs français et représentatif du champ musical d'une période donnée. Le classement du Top 50 a débuté en 1984, il est ainsi aisé d'avoir accès aux titres classés numéro un depuis cette date.⁵ Il a tout d'abord comptabilisé seulement les singles physiques avant de prendre également en compte, dès 2016, avec donc un certain délai, les téléchargements et le streaming (limité aux écoutes en streaming payantes à partir de 2018 ; cf SNEP, 2019). Avant 1984, il faut se contenter d'estimations et les titres sélectionnés pour cette période dans notre corpus ne le sont qu'à titre indicatif. Notons que ces classements ne comptabilisent ni la diffusion dans les médias, ni le prêt entre particuliers ou en bibliothèque, ni la copie, ni le téléchargement illégal ou encore la revente, classements qui peuvent aussi refléter des enjeux économiques plus qu'un réel engouement de la part des consommateurs. Nos trois critères de sélection étaient les suivants : que ce soit une chanson (sont ainsi exclus les sketches humoristiques), ayant été classée top 1 et qu'elle soit francophone ou partiellement francophone. Ce dernier choix est motivé par l'absence de liaisons dans les autres langues ayant accédé à cette position (telles l'anglais, le latin, le créole guadeloupéen ou encore le douala sans compter les titres uniquement instrumentaux). Une chanson peut contenir de trois vers en français pour la chanson *Bad Romance* de Lady Gaga (2010 [2009]) à tous ses vers en français. Étant donné que notre intérêt porte principalement sur ce qui est proposé aux auditeurs, nous n'avons exclu ni les chansons mixtes, ni celles de chanteurs n'étant pas francophones natifs.

Le corpus ainsi constitué a une durée totale de 21h 27min (temps musicaux et alternance codique compris), comptabilise 368⁶ chansons, 107 550 mots,⁷ 5914 contextes de liaison possibles pour 3252 liaisons réalisées.⁸ Nous considérons comme site potentiel de liaison une frontière externe de mots entre tout mot ayant une consonne finale graphique non prononcée et potentiellement de liaison (M1) et un mot à initiale vocalique à l'oral (M2),⁹ quel que soit le contexte prosodique, syntaxique de l'énoncé, le statut lexical ou encore morphosyntaxique des mots 1 et 2. Même si cela provoquera pour bon nombre de corpus un bruit important, toute comparaison est alors permise, même avec les corpus les plus atypiques : corpus acquisitionnels, de locuteurs non natifs, d'écrits oralisés ou encore corpus de chansons traditionnelles. Nous ne pouvons que recommander d'engager une réflexion sur nos pratiques et de rendre publics à la fois nos corpus, nos jeux de données étudiés et de présenter dans nos résultats les effectifs et non pas seulement des pourcentages afin de voir émerger des méta-analyses cohérentes.

⁵Les données sont accessibles en ligne jusqu'à présent sur le site <http://www.chartsinfrance.net>.

⁶Pour cet article, deux chansons ont été exclues en raison de problèmes techniques (ici, n=366).

⁷Le nombre de mots dans le corpus a été calculé avec la commande FREQ du logiciel CLAN qui permet d'éviter de prendre en compte la ponctuation et surtout les codes propres au format CHAT.

⁸Voir la note de bas de page 16 pour les cas exclus de cette analyse.

⁹Voir notes de bas de page 13 et 14.

Notre corpus peut être vu comme socialement situé car de par leur image publique nous pouvons facilement obtenir des informations d'ordre sociolinguistique sur les chanteurs et sur leur place dans le champ musical. Même s'ils ne constituent pas une population sociolinguistique équilibrée, notre corpus étant exhaustif, elle est néanmoins socialement représentative. Nous n'avons alors pas eu à constituer un échantillon représentatif de la communauté des chanteurs francophones, ni du champ musical français (on peut noter dans notre corpus des chansons relevant de plusieurs genres musicaux : chansons d'amour, chansons de foot, rap...), étant donné que nous avons à notre disposition la population totale visée, soit l'ensemble des titres numéro un, sur une période temporelle précise.

Quelques protagonistes (Johnny Hallyday > Mylène Farmer > Dalida = Michel Sardou > Claude François¹⁰) cumulent les titres numéro un (18% des titres dépendent de ces cinq chanteurs ; 72/402¹¹) alors qu'un très grand nombre de chanteurs n'atteindra qu'une fois la première place du hit-parade (43% ; 173/402). Si tous nos sujets exercent le métier de chanteur, on trouve plus d'hommes (67% ; 270/403) que de femmes (33% ; 133/403), et la plupart des chanteurs étaient plutôt jeunes (78%, soit 276/354 ont dans la vingtaine ou la trentaine), avec une fourchette allant de 4 (Jordy avec *Dur dur d'être bébé !*, 1992 [1992]) à 65 ans (Johnny Hallyday avec *Ça n'finira jamais*, 2008 [2008]). Tous ne sont aussi pas nés en France et certains ne sont pas de langue maternelle française (comme Shakira, de langue maternelle espagnole, chantant *Je l'aime à mourir*, 2011 [2011]). Un bon nombre de chanteurs nés en France proviennent pour leur part de la région parisienne. Il est à noter que le lieu de naissance n'est pas toujours un critère pertinent, les chanteurs étant un groupe socio-culturel ayant une grande mobilité géographique, mais cette donnée est plus facilement disponible que par exemple le lieu de résidence lors de l'accession au top 1.

Le nombre de chansons par année varie d'une seule (2012) à treize chansons (2007), en fonction des autres titres en tête du hit-parade. Le nombre de contextes de liaison possibles variant de ce fait selon les années, nous avons arbitrairement décidé de regrouper les chansons par décennie pour notre étude micro-diachronique. La Figure 1 montre que la plupart des chansons ne font qu'une irruption brève en tête des charts (par exemple, uniquement une semaine pour 20% des chansons, soit 73/368), alors que seul un petit nombre réussira à s'y maintenir longtemps. Des chansons prégénantes pourraient davantage toucher leurs contemporains, et de la sorte certains contextes de liaison pourraient être mis en avant dans l'ambiance musicale d'une époque. On citera encore le cas problématique des chansons qui accèdent au top 1 bien après leur sortie et pour lesquelles la façon de lier des chanteurs est prise en compte avec l'année d'accession au top 1 alors qu'elle est le produit d'une époque (par exemple, *Je te promets* de Johnny Hallyday, 2017 [1986], numéro un à la mort du chanteur

¹⁰Pour le Top 50 seul (dès 1984), le classement est le suivant : Mylène Farmer > Johnny Hallyday > Céline Dion > Garou = Michaël Youn = Stromae.

¹¹L'effectif total varie selon ce qui est recherché car dans une même chanson, il peut y avoir plusieurs chanteurs. En outre, certaines valeurs peuvent être manquantes.

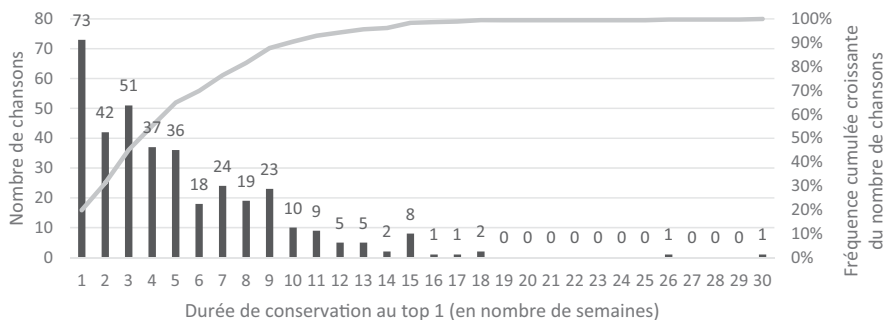


Figure 1. Nombre de chansons selon la durée de conservation au top 1.

en 2017), ou le cas des reprises ou référence à des chansons existant antérieurement où, par fidélité avec la chanson d'origine, les chanteurs peuvent adopter une façon de liasonner qui n'est plus en phase avec leur époque (par exemple, *Ah... si tu pouvais fermer ta gueule...* de Patrick Sébastien, 2008 [2008] fait référence à la *Marseillaise* dans le vers *allons [z] enfants ça sera notre hymne à nous*).

TRANSCRIPTION ET ANNOTATION DU CORPUS

Les chansons retenues ont été transcrites et alignées sur le son (vers par vers) à l'aide du logiciel CLAN (*Computerized Language Analysis*, MacWhinney, 2000), utilisé à l'origine pour traiter des données acquisitionnelles. Afin de faciliter le repérage automatique des liaisons, seuls les vers en français ont été transcrits et en cas d'alternance codique au sein du même vers, les parties non francophones n'ont pas été conservées. Les transcriptions se fondent sur des paroles de chansons proposées par des internautes (notamment sur des sites tels que *Lyricstranslate*¹²) adaptées au format CHAT (conventions de transcription et d'annotation de CLAN).

Plusieurs étapes ont été nécessaires en partant de ces transcriptions pour arriver vers des transcriptions sous le logiciel ELAN (*EUDICO Linguistic Annotator*, cf Max Planck Institute for Psycholinguistics, 2018 et Brugman et al., 2004) avec pré-annotation automatique des liaisons. Le format XML des transcriptions facilite l'extraction des informations qui nous intéressent ainsi que des segments sonores associés. Comme on le voit dans la Figure 2, le logiciel propose une structuration plus maniable, avec un alignement temporel (segmentation au mot) et structurel (vers, liaison, partie du discours, lemme, métadonnées). Tout d'abord, les métadonnées ont été récupérées depuis les fichiers sources CLAN. Les transcriptions ont été converties au format .trs (logiciel Transcriber) grâce à l'outil TeiCorpo (cf MoDyCo, 2016 et Liégeois et al., 2015). Ce format nous a permis d'utiliser le logiciel Jtrans pour aligner notre transcription sur le signal sonore (voir Cerisara et al., 2009) selon un découpage au niveau de séparateurs comme l'apostrophe, le tiret et l'espace sans distinction de frontière interne de

¹²Le site *Paroles.net* a par exemple lui aussi été utilisé. Ces sites sont accessibles aux adresses suivantes : <https://lyricstranslate.com/>, <https://www.paroles.net/>.

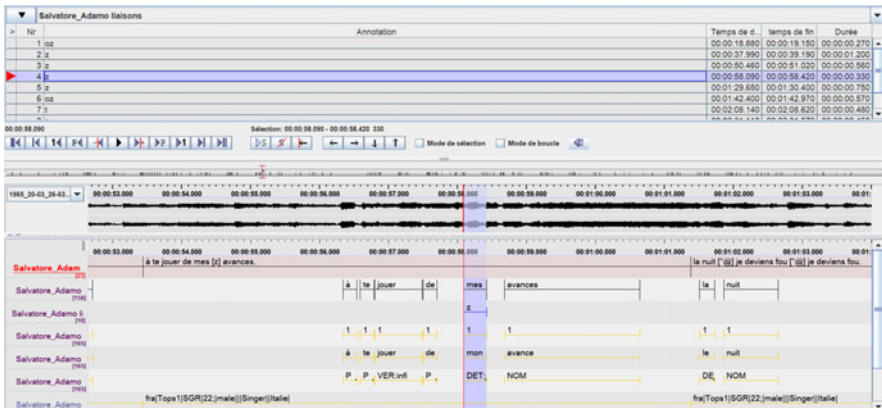


Figure 2. Structuration des fichiers ELAN.

mot. Ainsi nous obtenons aujourd'hui mais aussi l'arbre. Ce paramètre propre au logiciel n'est pas modifiable mais nous permet d'avoir un alignement optimal sur le son et un découpage en petites unités qui nous permettent de déterminer par la suite des contextes de liaison. À partir du résultat obtenu, nous avons pu créer une transcription au format .eaf (logiciel ELAN) en ajoutant une ligne d'annotation de la liaison pour chaque locuteur et en proposant des liaisons potentielles pour chaque contexte de liaison possible. Ces derniers sont repérés en fonction de critères définis pour chaque corpus : derniers¹³ et premiers¹⁴ caractères des mots, liste d'exclusion de certains M1 à consonne finale toujours prononcée (tels *Dallas*, *short*, verbes à terminaison en *-ir*), voire de M2, pas de liaison possible entre deux énoncés (ici entre deux vers). On peut choisir de délimiter les énoncés selon le découpage des fichiers sources (ce fut notre cas avec les fichiers CLAN) ou bien de considérer une pause de 800 millisecondes comme un changement de tour de parole (critère utile pour d'autres corpus oraux).¹⁵ De plus nous avons annoté notre corpus en parties du discours (ou PoS pour *parts of speech*) et en lemmes alignés sur les mots via le logiciel Treetagger (cf Schmid, 1999) et en utilisant à nouveau l'outil TeiCorpo.

Depuis ELAN ont ensuite été vérifiés manuellement à la fois les contextes de liaison préannotés (suppression du bruit) et les liaisons, préannotées par défaut

¹³⟨s⟩, ⟨x⟩ et ⟨z⟩ pour des liaisons en [z], ⟨t⟩ et ⟨d⟩ pour des liaisons en [t], ⟨n⟩ et ⟨m⟩ pour des liaisons en [n], ⟨r⟩ pour des liaisons en [ʁ], ⟨p⟩ pour des liaisons en [p] et ⟨g⟩ pour des liaisons en [k]. ⟨f⟩ n'a pas été pris en compte car dans les sandhis externes du type *neuf heures* prononcé [nœvœʁ], le [v] ne correspond pas à notre définition de la liaison puisque le mot *neuf* est prononcé [nœf] en isolation et devant un mot à initiale consonnantique en français contemporain.

¹⁴Voyelles et ⟨h⟩ majuscules et minuscules. Des occurrences telles que *et plu(s) mon coeur sous [z] x@l* (*Oui mais... non*, Mylène Farmer, 2010 [2010]) doivent être préannotées dans CLAN.

¹⁵Souvent lorsque le M1 est séparé du M2 par une pause jugée trop longue, ce contexte n'est pas considéré comme un site potentiel de liaison. Pour notre corpus, nous n'avons pas estimé qu'une pause à l'intérieur d'un vers bloque la liaison car par exemple, pour la chanson *Pour un infidèle* de Cœur de pirate et Julien Doré (2010 [2010]) nous avons trouvé une pause de 0,47s et une inspiration avant la liaison du vers *les plus douces de tes [z] histoires*.

comme non réalisées, ont été réannotées à l'aide d'un menu déroulant¹⁶ sur une base perceptive. L'annotation des chants à l'unisson reste problématique, il a été nécessaire de trancher en fonction de ce qui était globalement perçu (les cas les moins clairs ont été annotés comme incertains). Enfin, ELAN permet des recherches sur corpus en ayant la possibilité de revenir facilement aux contextes de réalisation et de faire de premières statistiques, ou d'exporter les données dans un tableur.

En résumé, l'outil ici présenté, développé pour l'étude de la liaison, permet de préannoter automatiquement les contextes de liaison possibles à partir de transcriptions existantes. Leur vérification manuelle se fait en allant de contexte en contexte qui sont directement réécoutables. Cet outil nous semble prometteur pour faciliter des études futures, afin de traiter plus rapidement des données toujours plus importantes. Il reste à perfectionner, notamment en réfléchissant à une possible automatization de l'obtention d'informations complémentaires absentes du codage actuel (nature de la consonne de liaison et réalisation de cette consonne) mais prises en compte dans d'autres systèmes d'annotation de la liaison (notamment PFC, cf Durand et al., 2009) : nombre de syllabes du M1, enchaînement, présence d'une pause, nasalité de la voyelle précédant les liaisons en [n].

LA LIAISON DANS UN CORPUS CHANTÉ

Globalement, les liaisons sont plus réalisées dans ce corpus de chansons (54,99% ; 3252/5914) que dans des interactions entre connaissances (PFC 44,11% ; 18760/42528),¹⁷ ce qui serait caractéristique de productions distantes. Ceci n'est pas sans rappeler les taux plus élevés de réalisation observés en lecture (PFC : 62,16% ; 7723/12425).¹⁸ L'ordre de fréquence des consonnes de liaison du corpus de tops 1 (en excluant la non-liaison) est le même que celui des conversations du corpus PFC (Tableau 1) : /z, n, t ʁ, p/, mais on observe, tout comme en lecture dans le corpus PFC, que la part des liaisons en /n/ est moins importante et celle des liaisons en /t/ plus importante.

Cette analyse peut être affinée en observant plus en détail ce qui se produit au niveau de certaines liaisons variables. Avant de faire quelques observations, nous ferons deux remarques. Tout d'abord, les chiffres sont en réalité très

¹⁶Les liaisons non réalisées ont été annotées sur le schéma ox (oz, on, ot, or, op, ok), les liaisons réalisées sur le schéma x (z, n, t, r, p, k), les incertitudes sur le schéma x* (z*, n*, t*, r*, p*, k*) et les sandhis externes non attendus en fonction de la consonne finale graphique du M1 sur le schéma *x (*z, *n, *t, *r, *p, *k), par exemple pour le pataqués *pour qu'il sourie [ʔt] en s'endormant* (La Mamma, Aznavour, 1964 [1963]). Les deux derniers cas ne sont pas pris en compte dans l'analyse, le dernier nécessite d'être préannoté dès les fichiers.cha. Cette façon d'annoter découle des réflexions d'un groupe de travail animé d'une volonté d'annotation commune des liaisons (Dugua et al., 2017).

¹⁷Les données issues du corpus PFC ont été extraites en décembre 2020. N'ont ici été prises en compte que les données issues des conversations libres et guidées et ont été exclues les données pour le [k] et le [v] ainsi que les liaisons annotées comme incertaines et épenthétiques.

¹⁸Ici n'ont été prises en compte que les données issues de la tâche de lecture du texte, avec les mêmes exclusions. Notons que les contextes de liaison et les liaisons produites sont fortement dépendantes du texte sélectionné.

Tableau 1. Fréquence des consonnes de liaison dans PFC et le corpus de tops 1

		/z/	/n/	/t/	/ʁ/	/p/	/k/
PFC conversations	%	46,29%	39,14%	14,34%	0,14%	0,09%	0%
	n	8524/18416	7208/18416	2641/18416	26/18416	17/18416	0/18416
PFC lecture	%	44,58%	30,09%	25,13%	0,21%	0%	0%
	n	3399/7625	2294/7625	1916/7625	16/7625	0/7625	0/7625
Tops 1	%	45,39%	30,63%	22,82%	1,08%	0,09%	0%
	n	1476/3252	996/3252	742/3252	35/3252	3/3252	0/3252

difficilement comparables étant donné que les études précédentes, à part certaines études sur les locuteurs non natifs du français, ne tiennent pas compte des contextes morphosyntaxiques et lexicaux considérés comme impossibles. Dans les mêmes conditions, nos taux de réalisation pourraient encore être supérieurs aux taux ici présentés, puisque nous prenons en compte tous les contextes, même ceux où des liaisons n'ont jamais été notées. De plus, nous avons, pour le Tableau 2, présenté des résultats différents pour le même corpus, PFC, selon la masse de données disponibles lors des deux analyses. Nous pouvons voir que même pour un corpus aussi massif que PFC, les effectifs semblent être encore trop petits pour obtenir des résultats fiables.

Après le verbe *être* conjugué à la première personne du présent ou de l'imparfait de l'indicatif (Tableau 2), on peut notamment noter un taux de réalisation des liaisons élevé chez les chanteurs, comparé aux corpus déjà existants, surtout par rapport aux corpus variationnistes.

Après des adverbes polysyllabiques (Tableau 3), on notera que les chanteurs liaisonnent souvent après *jamais* (à 64%), ce que Laks (2009: 261) avait déjà constaté dans la parole politique, mais contrairement aux personnalités politiques, jamais après *toujours* (les effectifs sont trop restreints pour proposer une quelconque conclusion quant aux autres adverbes polysyllabiques). C'est surtout en comparant nos données à des productions plus spontanées que ce type de liaisons variables apparaît comme un moyen de se démarquer de productions relevant de la proximité.

DIACHRONIE DE LA LIAISON DANS LES CHANSONS

Contrairement aux études diachroniques antérieures sur la liaison, pour notre corpus de titres ayant accédé à la première place de classements musicaux, la diminution du taux de liaisons réalisées est visible, liaisons variables comme invariables confondues. Entre la première décennie et la dernière, on passe en effet de 68% de liaisons réalisées à 44% (annotation 'x', Figure 3 : 1956–1966 : 68% (644/950), 1967–1976 : 64% (615/959), 1977–1986 : 60% (492/824), 1987–1996 : 54% (375/699), 1997–2006 : 47% (613/1313), 2007–2017 : 44% (513/1169)). La chute du taux de liaison semble s'effectuer principalement au

Tableau 2. Liaisons après le verbe être conjugué

	Ågren 1973 (radio)		De Jong 1994 (Orléans)		Mallet 2008 (PFC)		Pustka 2017 (livres audio)		PFC 2020		Tops 1	
	%	n	%	n	%	n	%	n	%	n	%	n
<i>suis</i>	47%	139	29%	209	13%	430	45%	33	14%	628	52%	25
<i>es</i>	88%	8	0%	1	0%	53	41%	22	5%	97	6% ¹⁹	77
<i>est</i>	97%	2668	69%	1692	44%	636	86%	148	33%	2825	85%	278
<i>c'est</i>	-	-	-	-	28%	1470	87%	143	-	-	-	-
<i>sommes</i>	58%	74	71%	28	44%	27	11%	9	31%	49	33%	6
<i>êtes</i>	71%	34	0%	11	31%	16	75%	4	18%	33	0%	1
<i>sont</i>	86%	280	46%	200	19%	208	69%	16	24%	326	88%	26
<i>étais</i>	21%	29	5%	76	-	-	16%	19	6%	18	80%	10
<i>était</i>	75%	367	19%	212	8%	343	73%	174	5%	1041	82%	28
<i>c'était</i>	-	-	-	-	-	-	63%	52	-	-	-	-
<i>étaient</i>	63%	57	21%	34	11%	71	82%	27	10%	99	50%	2

détriment de liaisons en /z/ et /t/, consonnes de liaisons connues pour être liées à des cas de liaisons variables.

En distinguant les liaisons invariables (selon les propositions de Côté, 2017 : déterminant + adjectif/nom, proclitique + proclitique/verbe, verbe + enclitique et *en* + X), qui sans surprise, sont stables (en moyenne réalisées à 95%), des liaisons variables, on observe que ces dernières chutent de façon régulière de 51% (316/950) à 18% (145/1169) entre la première et la dernière décennie, pour atteindre des taux proches de ceux observés dans la langue quotidienne (21,39% (n = 389) pour le corpus des ESLO ; Dugua et Baude, 2017: 49 et 19% (4296/22568) pour PFC ; Barreca, 2015: Annexe 2). Nous nous alignons sur les observations de Laks (2009) : le changement diachronique peut provenir de glissements diaphasiques, diastratiques ou encore diamésiques. Ici, le changement en cours dans les chansons, à première vue diachronique, semble relever d'un alignement progressif de la langue chantée sur la langue parlée, les chansons

¹⁹L'étude de chansons pose des problèmes lors des statistiques, car un même énoncé peut être prononcé un très grand nombre de fois dans une même chanson voire être répété par une seconde voix ou à l'aide d'un effet d'écho. Ici, les cooccurrences *t'es O_K* et *t'es in* de la chanson *T'es OK* (Ottawan, 1980 [1980]), qui ne liaisonnent pas, influencent énormément les chiffres présentés : en excluant ces données (n = 52, on aurait par exemple 20% de réalisation au lieu de 6%). Comme expliqué auparavant, nous n'avons exclu aucun contexte à initiale vocalique, car les auditeurs du français apprennent grâce à leur input quels M2 peuvent être liaisonnés ou non et dans une linguistique fondée sur l'usage, les contextes non liaisonnés font autant partie du système que les autres. Une façon de pallier ce problème serait notamment de travailler avec un large jeu de données (*big data*).

Tableau 3. Liaisons après des adverbes polysyllabiques

	Ågren 1973 (radio)		De Jong 1994 (Orléans)		Laks, 2009 (HPOL1)		Pustka 2017 (livres audio)		PFC 2020		Tops 1	
	%	n	%	n	%	n	%	n	%	n	%	n
<i>assez</i>	62%	108	16%	49	0%	1	0%	2	3%	66	–	0
<i>beaucoup</i>	72%	71	–	–	33%	3	0%	1	7%	81	0%	8
<i>complètement</i>	36%	11	–	–	–	–	20%	5	17%	12	–	0
<i>jamais</i>	14%	57	6%	64	73%	11	13%	30	1%	126	68%	31
<i>tellement</i>	46%	26	–	–	100%	1	17%	18	0%	39	0%	2
<i>souvent</i>	6%	16	0%	1	50%	2	50%	2	2%	42	0%	4
<i>toujours</i>	4%	61	–	–	73%	11	4%	23	0%	294	0%	43

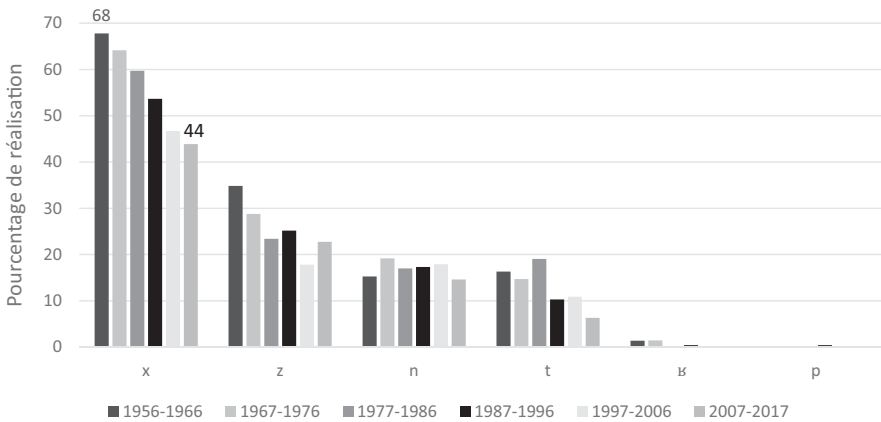


Figure 3. Evolution de la réalisation des liaisons dans les tops 1.

passant de la distance à la proximité. Ceci serait à vérifier pour d'autres phénomènes linguistiques caractéristiques de productions distantes, en observant par exemple l'évolution du taux de réalisation du *il* à valeur impersonnelle ou du *ne* de négation.

Il y a tout lieu de penser que ces changements affectent des constructions spécifiques, de façon plus ou moins visible. On peut par exemple regarder ce qui se passe pour les liaisons après *est* (Tableau 4), où un changement semble en cours depuis les trois dernières décennies ou encore après les prépositions²⁰ (Tableau 4).

C'est donc dans des contextes spécifiques, morphosyntaxiques ou lexicaux et à différents niveaux d'analyse qu'il semble falloir aller chercher une explication

²⁰PRP dans l'annotation en POS de Treetagger, qui n'a pas été manuellement corrigée.

Tableau 4. Diachronie de la liaison dans deux constructions : est + X et PRP + X

	1956–1966		1967–1976		1977–1986		1987–1996		1997–2006		2007–2017	
	%	n	%	n	%	n	%	n	%	n	%	n
est +X	100%	30/30	95%	37/39	100%	73/73	87%	13/15	75%	58/78	60%	26/43
PRP+X	100%	45/45	96%	51/53	89%	34/38	96%	27/28	89%	42/47	91%	40/44

à l'évolution diachronique du taux de liaison global. Les liaisons variables dans de tels contextes rendent les productions étudiées distantes, cependant en diachronie certaines de ces liaisons sont moins produites ce qui rend les chansons actuelles plus proches de réalisations authentiques en matière de liaison.

CONCLUSION

Le corpus que nous avons développé est le premier pour l'étude de la liaison à être constitué de chansons issues de l'industrie du disque. Il vient également apporter un éclairage nouveau à l'étude diachronique de la liaison, pour laquelle on ne disposait que de données de locuteurs des villes de Tours et d'Orléans ainsi que de personnalités politiques. Notre corpus est temporellement clos sur la période étudiée et de fait, nous avons accès à des sujets qui ne sont pas équilibrés comme dans un corpus variationniste et à une population (les chansons) de la même façon non équilibrée en termes de genres musicaux. Nous avons pris le temps de lister ses limites afin d'aborder nos résultats avec la prudence nécessaire. Nous avons également présenté un outil prometteur pour traiter des corpus d'étude de la liaison plus rapidement qu'avec un repérage des contextes et une annotation des liaisons manuels, même s'il nous reste à affiner notre codage.

Les liaisons dans notre corpus de tops 1 sont globalement plus réalisées que dans un corpus interactionnel, la part des liaisons en /t/ y est supérieure et certaines constructions spécifiques, comme les liaisons après *suis*, *étais* ou *jamais*, semblent être caractéristiques de productions distantes (personnalités politiques, livres audio pour enfants). C'est surtout dans sa dimension diachronique que notre corpus révèle sa transition d'une mise en scène plutôt distante, avec l'utilisation de liaisons plus rares dans des constructions spécifiques, à une proximité croissante. De 1956 à 2017, le taux global de liaison a en effet diminué de façon importante, ce qui n'avait jamais été noté même chez des locuteurs professionnels, ce qui est évidemment imputable aux seules liaisons variables. Des corpus tels que le nôtre ont pu amener certains linguistes à parler de chute de la liaison en français contemporain. Nous avons montré que celle-ci se fait à travers des constructions spécifiques, synonymes de distance, et à divers niveaux d'analyse. La convergence progressive de la langue chantée vers la langue parlée dans les tops 1 sur la période étudiée nous semble principalement illustrer une variation diaphasique moindre au fil du temps.

Notre corpus reste limité par sa taille. Étant exhaustif, il pourrait être agrémenté au fil des années avec les nouveaux titres accédant en tête des charts, voire être étendu aux tops 2 et 3. Nous avons présenté ici une étude relativement

préliminaire, les pistes futures qui s'offrent à nous sont : la prise en compte des variables sociolinguistiques, des différents genres musicaux et notamment de questions de métrique et de text-setting, une distinction des temps chantés et déclamés, tout comme on pourrait s'intéresser aux trajectoires individuelles de chanteurs en particulier.

Abréviations. Voir les abréviations du jeu d'étiquettes de Treetagger, disponibles sur <https://cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/french-tagset.html>.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Ågren, J.** (1973). *Étude sur quelques liaisons facultatives dans le français de conversation radiophonique : fréquences et facteurs*, Acta Universitatis Upsaliensis, Uppsala: Kå-We Tryck.
- Ashby, W.** (2003, September). *La liaison variable en français parlé tourangeau: une analyse en temps réel*, Paper presented at the Colloque de l'AFLS (Association For French Language Studies) : Le français aujourd'hui : problèmes et méthodes, Tours, France.
- Barreca, G.** (2015). L'acquisition de la liaison chez des apprenants italo-phones : des atouts d'un corpus de natifs pour l'étude de la liaison en français langue étrangère (FLE). Thèse de doctorat : Sciences du langage, Université Paris Ouest Nanterre La Défense, Università Cattolica del Sacro Cuore di Milano.
- Bidan, M., et Cuccagna, M.** (2016, March 29). Comment l'argot de Grigny a envahi la France. Retrieved July 7, 2020, from <https://www.streetpress.com/sujet/1459251456-comment-argot-grigny-envahi-france>
- Brugman, H., Russel, A., et Nijmegen, X.** (2004). Annotating Multi-media/Multi-modal resources with ELAN. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, Portugal: European Language Resources Association.
- Bybee, J.** (2006). From usage to grammar: the mind's response to repetition. *Language*, 82(4): 711–733.
- Cerisara, C., Mella, O., et Fohr, D.** (2009). JTrans, an open-source software for semi-automatic text-to-speech alignment. In: *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*, Brighton, UK.
- Chauvin, C., et Colletta, J.-M.** (2003). La gestualité dans les jeux chantés du folklore enfantin : description, transcription et analyse. In: G. Barrier et N. Pignier (dir.), *Sémiotiques non verbales et modèles de spatialité : textes du congrès Sémio 2001*, Limoges: Presses Universitaires de Limoges, pp. 39–62.
- Chigarevskaja, N. A.** (1973). *Traité de phonétique française : cours théorique*, 3e édition, 1982. Moscou: Vyssšaja škola.
- Côté, M.-H.** (2017). La liaison en diatopie : esquisse d'une typologie. *Journal of French Language Studies*, 27(1): 13–25.
- De Jong, D.** (1994). La sociophonologie de la liaison orléanaise. In: C. Lyche (dir.), *French Generative Phonology: Retrospectives and Perspectives*. AFLS/ESRI: Salford, pp. 95–130.
- Descamps, F.** (2005). *L'historien, l'archiviste et le magnétophone: de la constitution de la source orale à son exploitation*, Paris: Institut de la gestion publique et du développement économique.
- Dugua, C. et Baude, O.** (2017). La liaison à Orléans, corpus et changement linguistique : une première étude exploratoire. *Journal of French Language Studies*, 27(1): 41–54.
- Dugua, C., Baude, O., Badin, F. et Siccardi, A.** (2017, March). *La liaison dans l'environnement langagier des enfants : Vers une annotation commune ?*, Paper presented at FLORAL (Français Langue Orale et Recherches Avancées en Linguistique) 2017 : Accessibilité, représentations et analyses des données, Orléans, France.
- Durand, J., Laks, B., Calderone, B. et Tchobanov, A.** (2011). Que savons-nous de la liaison aujourd'hui ? *Langue française*, 169(1): 103–135.
- Durand, J., Laks, B. et Lyche, C.** (2009). Le projet PFC (Phonologie du Français Contemporain) : une source de données primaires structurées. In: J. Durand, B. Laks et Ch. Lyche (dir.), *Phonologie, variation et accents du français*. Paris: Hermès, pp. 19–61.
- Encrevé, P.** (1977). Présentation : Linguistique et sociolinguistique. *Langue Française*, 34: 3–16.
- Encrevé, P.** (1983). La liaison sans enchaînement. *Actes de la recherche en sciences sociales*, 46(1): 39–66.

- Eshkol-Taravella, I., Baude, O., Maurel, D., Hriba, L., Dugua, C. et Tellier, I. (2011). Un grand corpus oral «disponible» : le corpus d'Orléans 1968-2012. In: B. Sagot et N. Bel (dir.), *Traitement Automatique Des Langues, ATALA 2011, Ressources Linguistiques Libres*, 52(3), 17–46.
- Hornsby, D. (2019). Variable liaison, diglossia, and the style dimension in spoken French. *French Studies*, 73(4): 578–597.
- Koch, P. et Oesterreicher, W. (2001). Langage parlé et langage écrit. *Lexikon Der Romanistischen Linguistik*, 1(2): 584–627.
- Laboratoire Ligérien de Linguistique – UMR 7270 (LLL). (2017). *ESLO [Corpus]*, ORTOLANG (Open Resources and TOols for LANGuage)-www.ortolang.fr. Retrieved from <https://hdl.handle.net/11403/eslo/v1>
- Labov, W. (1994). *Principles of Linguistic Change: Volume 1: Internal Factors*, réimpression, 1999. Malden: Oxford: Blackwell.
- Laks, B. (2009). Dynamiques de la liaison en français. In: L. Baronian et F. Martineau (dir.), *Le français d'un continent à l'autre : mélanges offerts à Yves Charles Morin*. Québec: Presses de l'Université Laval, pp. 237–267.
- Laks, B. et Peuvergne, J. (2017). La liaison en français contemporain dans la parole publique (1999–2015). *Journal of French Language Studies*, 27: 55–72.
- Le Gorju, T. (2014, February 2). Les Français et la musique. Retrieved June 24, 2019, from <https://www.ipsos.com/fr-fr/les-francais-et-la-musique>
- Léon, P. (1992). *Phonétisme et prononciations du français : avec travaux pratiques d'application et corrigés*, 6e édition, 2011. Paris: Armand Colin.
- Liégeois, L., Etienne, C., Parisse, C., Benzitoun, C. et Chanard, C. (2015). Using the TEI as a pivot format for oral and multimodal language corpora. Paper presented at the Text Encoding Initiative Conference and Member's meeting 2015, Lyon, France, pp. 28–31.
- MacWhinney, B. (2000). *The CHILDES project: Tools for Analyzing Talk*, 3e édition, 2014, Vol. I: Transcription format and programs. New York; Hove: Psychology Press.
- Maisonneuve, S. (2009). *L'invention du disque 1877-1949 : Genèse de l'usage des médias musicaux contemporain*. Paris: Éditions des archives contemporaines.
- Mallet, G.-M. (2008). La liaison en français : descriptions et analyses dans le corpus PFC. Thèse de doctorat : Sciences du langage, Université Paris Ouest Nanterre La Défense.
- Max Planck Institute for Psycholinguistics. (2018). ELAN [Programme informatique] (Version 5.2), Nijmegen. Retrieved from <https://archive.mpi.nl/tla/elan>
- Modèles, Dynamiques, Corpus – UMR 7114 (MoDyCo). (2016). *teicorpo [Outil]*. Retrieved from <https://hdl.handle.net/11403/teicorpo/v1>
- Modèles, Dynamiques, Corpus – UMR 7114 (MoDyCo), et Université de Groningen (RUG). (2017). *PFC – Phonologie du Français Contemporain [Corpus]*, ORTOLANG (Open Resources and TOols for LANGuage)-www.ortolang.fr. Retrieved from <https://hdl.handle.net/11403/pfc/v1>
- Morin, Y.-C. (1982). De quelques [l] non étymologiques dans le français du Québec : notes sur les clitiques et la liaison. *Revue Québécoise de Linguistique*, 11(2): 9–47.
- Nardy, A., Chevrot, J.-P. et Chauvin-Payan, C. (2014). La liaison facultative dans les formes récitées du folklore enfantin. In: Ch. Soum-Favaro, A. Coquillon et J.-P. Chevrot (dir.), *La liaison : approches contemporaines*. Bern: Peter Lang.
- Pustka, E. (2017). L'écrit avant l'écriture : la liaison dans les livres audio pour enfants. *Journal of French Language Studies*, 27(2): 187–214.
- Russo, M. (2014, January). *Les origines de la liaison en proto-français : graphies et oralité*, Paper presented at the Colloque de la SIDF (Société Internationale de Diachronie du Français) : L'histoire du français : Nouvelles approches, nouveaux terrains, nouveaux traitements. Cambridge, UK.
- Schmid, H. (1999). Improvements in part-of-speech tagging with an application to German. In: S. Armstrong, K. Church, P. Isabelle, S. Manzi, E. Tzoukermann et D. Yarowsky (dir.), *Natural Language Processing using Very Large Corpora*, Vol. 11. Dordrecht: Springer, pp. 13–25.
- Söll, L. (1974). *Gesprochenes und geschriebenes Französisch*. Berlin: Eric Schmidt Verlag.
- Syndicat national de l'édition phonographique (SNEP). (2019). A propos des certifications. Retrieved from <http://snepmusique.com/les-certifications/a-propos-des-certifications/>
- Thomas, A. (1998). La liaison et son enseignement : des modèles orthoépiques à la réalité linguistique. *The Canadian Modern Language Review*, 54(4): 543–552.

RÉFÉRENCES DISCOGRAPHIQUES

- Aznavour, C. (1963). *La Mamma*, Barclay.
- Booba. (2014). *OKLM*, Tallac/AZ/Capitol/Universal.
- Coeur de pirate, et Doré, J. (2010). *Pour un infidèle*, Grosse Boîte.
- Farmer, M. (2010). *Oui mais... non*, Polydor/Universal Music.
- Hallyday, J. (1986). *Je te promets*, Philips/Phonogram.
- Hallyday, J. (2008). *Ça n'finira jamais*, Warner Music France.
- Jordy. (1992). *Dur dur d'être bébé !*, Columbia Records.
- Lady Gaga. (2009). *Bad Romance*, RedOne.
- Ottawan. (1980). *T'es OK*, Carrère/Zagora/Bleu Blanc Rouge.
- Sébastien, P. (2008). *Ah... si tu pouvais fermer ta gueule...*, Faisage Music.
- Shakira. (2011). *Je l'aime à mourir*, Sony Music Entertainment.