

# La liaison facultative en français : étude de grands corpus combinant approche automatique relâchée et jugement perceptif

Mathilde Hutin<sup>1,\*</sup>, Caihong Weng<sup>2</sup>, Martine Adda-Decker<sup>3</sup>, Lori Lamel<sup>1</sup>

<sup>1</sup> Université Paris-Saclay, CNRS UMR 9015, LISN, 91400, Orsay, France

<sup>2</sup> Université Paris Cité, 75013, Paris, France

<sup>3</sup> Université Paris 3 / Sorbonne Nouvelle, CNRS UMR 7018, LPP, 75005, Paris, France

**Résumé.** Les systèmes d'alignement automatique de la parole sont aujourd'hui très performants pour produire segmentations et étiquetages automatiques, notamment grâce aux variantes incluses dans leurs dictionnaires de prononciation, par exemple en autorisant des variantes avec et sans consonne de liaison (ex. *est* prononcé [e] ou [et]). La liaison en français est un phénomène de sandhi externe qui, lorsqu'elle est dite « facultative », se caractérise par le fait que la consonne de liaison peut être réalisée ou non de façon également grammaticale, ce qui représente un défi pour les systèmes. Nous nous concentrons ici sur la liaison dans le contexte « verbe *être* conjugué + *un/une* » et analysons plus de 100h de français parlé pour (i) analyser les erreurs d'alignement, et (ii) dresser un panorama de la réalisation de la liaison facultative en français contemporain. Nous montrons, d'abord, que les erreurs d'alignement ont souvent des causes non-identifiables à l'oreille et que, lorsque la cause est établie, elles sont essentiellement dues aux variations prosodiques et aux disfluences, typiques de l'oralité mais encore mal prises en compte par les systèmes. Ensuite, nous montrons que la réalisation de la liaison est influencée par le style de parole mais pas par le sexe des locuteurs et que son taux a baissé de 30% en 50 ans.

**Abstract. Optional *liaison* in French: Study of large corpora using loose automatic alignment and perceptual judgement.** Automatic alignment systems have achieved a high level of efficiency in segmenting and tagging phonemic units, in part thanks to the variants in their pronunciation dictionaries, for instance by allowing variants with and without *liaison* consonants (ex. Fr. *est*, 'is', aligned as [e] or [et]). The French *liaison* is an external sandhi phenomenon that can, in some cases, be optional, meaning that its realization or not are both grammatical, which is a challenge for ASR systems. We focus here on the optional *liaison* after the verb *être*, "be" when conjugated and followed by an

---

\* Corresponding author: [mathilde.hutin@lisn.upsaclay.fr](mailto:mathilde.hutin@lisn.upsaclay.fr)

indefinite article *un/une*, “a (masc./fem.)”. Over 100h of aligned French speech are used to: (i) analyze alignment errors and (ii) provide an overview of the realization of optional *liaison* in contemporary French. We show, first, that the sources of many alignment errors are not identifiable by ear but that, when they are, the errors most often stem from prosodic variations and dysfluencies, which, while common in spontaneous speech, remain challenging for automatic systems. Second, we show that the realization of optional French *liaisons* is influenced by speech style but not by speaker gender, and that the rates of realizations have dropped by 30% in 50 years.

## 1 Introduction

Les travaux en phonétique et en phonologie de laboratoire ont connu ces dernières décennies un essor de taille grâce à l’apport de nouvelles méthodologies issues de l’informatique. Ces méthodologies impliquent des processus automatisés permettant l’analyse rapide et efficace de très grands corpus de parole, comprenant jusqu’à plusieurs centaines d’heures, provenant de différentes sources et donc représentatifs de la parole naturelle sous toutes ses formes, évitant ainsi les écueils des données de laboratoire ou de terrain, souvent orientées (Wu & Adda-Decker 2021). Réciproquement, les résultats de ces recherches d’un nouveau genre ont permis d’améliorer les technologies existantes, par exemple en mettant en lumière l’importance de la finesse du dictionnaire de prononciations fourni au système de reconnaissance (Dutrey *et al.* 2016). C’est dans la lignée de ce cercle vertueux entre phonétique des grands corpus et technologies du langage que s’inscrit cet article.

L’un des principaux apports de cette collaboration est sans doute la mise en lumière de l’éminente variation de la parole, qui représente un défi à la fois pour les linguistes qui souhaitent comprendre le fonctionnement du langage et pour les informaticiens qui tâchent de développer des technologies pertinentes qui ne commettent pas d’erreurs. Ainsi, l’oralité se caractérise par de la variation au niveau prosodique : les locuteurs parlent plus ou moins vite (Jacewicz *et al.* 2010), ou augmentent le pitch, l’intensité ou la durée d’un phone ou d’un mot, ce qui est une source reconnue d’erreurs de reconnaissance (Siegler & Stern 1995, Shinozaki & Furui 2001, Fosler-Lussier & Morgan 1999 sur le débit ; Goldwater *et al.* 2009 sur le pitch, l’intensité et la durée). L’oralité se caractérise aussi par la présence de marqueurs discursifs révélateurs de la (ré-)organisation du discours (Maschler & Schiffrin 2015) et de disfluences inhérentes à la production orale (pauses silencieuses, pauses remplies, hésitations, répétitions...) (Shriberg 1994) et qui tous deux présentent des défis à la reconnaissance du langage par les robots (Vasilescu *et al.* 2010 sur les marqueurs discursifs, Goldwater *et al.* 2009 sur les marqueurs discursifs et les disfluences). Enfin, la parole se caractérise aussi par des phénomènes de réductions segmentales parfois spectaculaires (Ernestus & Warner 2011, Wu & Adda-Decker 2021) qui représentent autant de challenges pour la reconnaissance vocale (Adda-Decker & Lamel 1998, Adda-Decker *et al.* 2005, Goldwater *et al.* 2009). Dans ce cadre, évaluer la performance des systèmes permet d’améliorer les outils et, en retour, analyser les erreurs commises par les systèmes est riche d’enseignement sur la langue.

Le présent article propose d’utiliser une telle approche pour une étude préliminaire sur un peu plus de 100 heures de français alignées automatiquement pour mettre en lumière les erreurs commises par le système sur un processus catégoriel du français : la liaison. Ce choix sera explicité en Section 2. Notre opération a deux buts : d’une part, il s’agira d’analyser certaines causes des erreurs d’alignement pour mettre en lumière de nouvelles hypothèses de travail ; d’autre part, il s’agira d’établir les motifs de réalisation de la liaison facultative en français et, notamment, de comparer les données contemporaines avec des données passées pour faire un état des lieux de l’évolution de l’usage.

L’article est construit comme suit. Dans un premier temps, nous revenons brièvement sur la question de la liaison, sa définition et pourquoi nous l’avons choisie comme cas d’étude

pour cette recherche (Section 2). Dans la Section 3, nous présentons nos corpus et notre méthodologie. Dans la Section 4, nous établissons une typologie des erreurs d'alignement et les nouvelles questions linguistiques qu'elle soulève. La Section 5 présente les résultats concernant l'usage actuel de la liaison et une comparaison avec des données similaires des années 1960 (issues d'Ågren 1973) afin de tirer des conclusions sur la diachronie. La Section 6 sera consacrée à la conclusion et la discussion de l'article.

## 2 La liaison dite facultative en français

Dans cette section, nous revenons brièvement sur la définition de la liaison et pourquoi elle est intéressante pour cette étude particulière dédiée aux erreurs d'alignement.

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. Cette consonne latente peut apparaître entre deux mots si Mot2 commence par une voyelle, comme dans *les + amis* = [lezami], mais pas dans *les + copains* = \*[lezkopɛ̃].

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 » (ex. *un [n] ami*), « facultative » (ex. *amis [z/0] intimes*) et « interdite » (ex. *les [0] héros*).

Cependant, les analyses plus récentes du corpus *Phonologie du Français Contemporain*, (désormais, PFC) ont amené à remettre en question les appellations comme « obligatoire » ou « facultative » (Durand *et al.* 2011). Dans des contextes jusqu'à présent reconnus comme étant des contextes de liaison obligatoire, notamment le contexte « Mot invariable monosyllabique + X » (ex. *dans [z] une heure*), 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 dépendant de phénomènes morphologiques et catégoriels.

Cette variation peut être due à de nombreux facteurs, sociolinguistiques ou extralinguistiques. Ainsi, Delattre (1947) notait déjà une forte différence dans la fréquence des liaisons facultatives selon les registres utilisés par les locuteurs. Entre le registre familier et le registre plus soutenu, la fréquence de production 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 facultative, des études (Malécot 1975, Ashby 1981) montrent que les locuteurs plus âgés réalisent plus de liaisons, surtout en conversation spontanée. Quant au sexe des locuteurs, les résultats sont controversés : les analyses de Malécot (1975) ou Booij & 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 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. Ce point est crucial car il signifie que (i) il y a de la variation dans la réalisation de la liaison, mais que (ii) même pour un locuteur humain, la production ou non de la liaison est également acceptable. Dans la mesure où, comme nous l'avons montré plus haut, liaison « obligatoire » et liaison « facultative » sont toutes les deux variables en parole spontanée (bien que la première le soit moins que la deuxième), cette intuition linguistique des locuteurs est donc ce qui distingue la liaison facultative de la liaison obligatoire. Nous attendons donc que la liaison

soit bien réalisée de façon catégorielle (elle est réalisée ou non) mais de façon non généralisée (pas dans tous les sites potentiels de liaison). Étant donnée la méthode que nous avons adoptée, et que nous expliquons dans la Section 3 juste ci-dessous, le choix de la liaison facultative permet d'utiliser l'alignement automatique comme un proxy pour un locuteur humain et de proposer de nouvelles hypothèses sur le traitement de la liaison (et plus largement de la langue) chez les systèmes automatisés. En retour, les données nous permettront de dresser un panorama de l'usage de la liaison facultative dans le courant des années 2010 et d'établir les variables qui l'influencent (style de parole, genre du locuteur...).

Pour cette étude préliminaire, nous restreignons l'étude au contexte « verbe *être* conjugué suivi de l'article indéfini singulier *un* ou *une* ». Ce choix a été fait afin de restreindre les facteurs de confusion et de permettre de combiner l'analyse quantitative à une analyse qualitative des erreurs d'alignement qui soit humainement réalisable.

## 3 Corpus et méthodologie

### 3.1 Les corpus

Un phénomène aussi spécifique et multifactoriel que la liaison facultative en français peut être observé de façon plus fiable grâce à de très grands corpus (Coleman *et al.* 2016). Outre l'observation à grande échelle qui permet des résultats statistiquement significatifs, l'avantage des grands corpus repose aussi sur le fait que la parole y est plus naturelle que dans des enregistrements opérés lors d'expériences en laboratoire ou d'enquêtes sur le terrain, et qu'ils peuvent représenter un échantillon intéressant de situations variées de communication permettant de produire des résultats contrastifs, plus complets et plus détaillés.

Dans cette étude, nous avons utilisé trois grands corpus représentatifs chacun d'un style de parole différent. Le corpus *Évaluation des Systèmes de Transcription enrichie d'Émissions Radiophoniques*, désormais ESTER (Galliano *et al.* 2005), comprend 80 heures de parole journalistique caractéristique du discours préparé ou semi-préparé, soigné, et potentiellement prompté, donc lu à haute voix. Le corpus *Évaluation en Traitement Automatique de la Parole*, désormais ETAPE (Gravier *et al.* 2012), contient 13,5 heures de radio et 29 heures de télévision en français, notamment des débats et des conversations, et est donc représentatif de monologues et dialogues semi-préparés à deux ou plusieurs interlocuteurs. Enfin, le *Nijmegen Corpus of Casual French*, désormais NCCFr (Torreira *et al.* 2010), contient 31 heures d'interaction *de visu* entre amis et est donc représentatif du discours informel spontané.

### 3.2 La méthodologie

La méthodologie se divise en deux phases : l'alignement automatique avec variantes de prononciation, et l'évaluation auditive.

#### 3.1.1 L'alignement automatique avec variantes de prononciation

Les corpus étaient tous trois munis d'une transcription manuelle dans l'orthographe du français. Ils ont ensuite été traités suivant la méthode décrite dans Gauvain (2002) et Hallé & Adda-Decker (2007, 2011). Le système de reconnaissance automatique de la parole du LISN a utilisé la transcription en français pour opérer un alignement avec des transcriptions phonétiques. Pour cela, le système dispose (i) de modèles de phones pour chaque phonème du français, qui lui permettent d'identifier les phones sur un spectrogramme, et (ii) de son dictionnaire de prononciation, qui lui donne accès aux prononciations possibles de chaque mot de la langue française.

Dans le cadre de notre étude, le dictionnaire a été enrichi de façon à autoriser l'une des quatre consonnes de liaison facultative /t, n, z, p/ devant n'importe quel mot commençant par une voyelle. Le /r/ n'a pas été ajouté à l'étude car il présente beaucoup de variation à travers les variétés de français (voir Webb 2009 pour la revue de littérature) et ne représente de toute façon qu'une liaison extrêmement rare (0,04% des liaisons dans PFC selon Durand & Lyche 2008). Ainsi, la séquence *était une* par exemple pouvait être alignée avec les transcriptions canoniques [eteyn] ou [etetyyn], mais aussi avec les transcriptions non-canoniques [etezyn], [etenyn] ou [etepyn], selon que la machine estimait qu'une consonne était réalisée ou non entre Mot1 et Mot2. Nous avons ensuite nettoyé les données afin d'assurer que la construction syntaxique était bien du type « verbe conjugué + objet » (par ex. *était un homme*), et non « verbe conjugué + adverbe » (par ex. *était un peu*). Au total, nous avons conservé 5049 occurrences du verbe *être* conjugué suivi de *un/une*.

Parmi ces 5049 occurrences, 731 présentaient une inadéquation entre les consonnes de liaison alignées par le système et l'orthographe du mot (par exemple *est une* aligné [ezyn], [enyn], ou [epyn]). Parmi les 4320 occurrences restantes, l'écoute aléatoire de 964 occurrences a montré que 71 occurrences étaient mal alignées, soit à peu près 7,36% du corpus. Ainsi, si l'on extrapole, on peut imaginer que 7,36% des 4320 occurrences, soit environ 318, étaient mal alignées : additionnées aux 731 déjà repérées, on obtient un total de 1049 occurrences mal alignées, soit environ 20% du corpus (contre moins de 5% avec un dictionnaire moins permissif, cf. Adda-Decker *et al.* 2012 sur NCCFr). Quelles sont donc les causes qui expliquent un tel taux d'erreurs ?

### 3.1.2 L'évaluation auditive

Lors de la deuxième étape, nous avons constitué un sous-corpus à partir de ces 5049 occurrences. La seconde autrice a écouté un total de 1695 occurrences, dont les 731 repérées préalablement par l'inadéquation avec l'orthographe et les 964 de contrôle. Les 1695 occurrences correspondant à 893 liaisons bien alignées et 802 liaisons mal alignées par le système. Pour chacun des 802 alignements erronés, elle a spécifié la cause probable de l'erreur :

- les causes techniques :
  - o la bande passante est insuffisante : l'audio a parfois été enregistré sur bande téléphonique, coupée sous 4 kilohertz, ce qui est suffisant pour extraire les formants (au moins F<sub>0</sub>, F<sub>1</sub>, F<sub>2</sub> et F<sub>3</sub>) mais provoque une inadéquation avec les modèles acoustiques appris sur bande large (= 8 kHz),
  - o l'audio est inaudible : il arrive que le son n'ait pas été correctement capté et qu'il ne soit pas bien présent dans le fichier-son ;
- les causes liées à la situation discursive :
  - o il y a du bruit de fond,
  - o des disfluences (pauses, hésitations, répétitions, autocorrections, bégaiements, rires) se trouvent à l'intérieur de la séquence Mot1-Mot2,
  - o deux personnes ou plus parlent en même temps,
  - o le locuteur hypo-articule, et le Mot1 ou le Mot2 n'est pas clair,
  - o la parole présente des caractéristiques de production particulière (chant ; voix trop basse, trop forte, craquée...).

Notons ici que nous n'avons pas pris en compte le débit de parole, car ce dernier se confond aisément avec les réductions de prononciation (hypo-articulations).

Dans ce qui suit, nous présentons deux analyses : une typologie des erreurs d'alignement, consacrée au groupe des liaisons mal alignées (Section 4) et un état des lieux de l'usage de la liaison facultative au début du XXI<sup>e</sup> siècle, sur l'ensemble des données vérifiées auditivement (Section 5).

## 4 Résultats : typologie des erreurs d'alignement

Dans cette section, nous nous concentrons sur les 802 occurrences d'alignement erroné, dont le détail par corpus et par genre est donné dans le Tableau 1.

**Tableau 1.** Répartition du nombre d'occurrences mal alignées par corpus et par genre du locuteur.

Corpus	Genre	Occurrences mal alignées	Total
ESTER	Femmes	82	260
	Hommes	178	
ETAPE	Femmes	70	398
	Hommes	324	
	Non-connu	4	
NCCFr	Femmes	70	144
	Hommes	74	
Total	Femmes	222	802
	Hommes	576	
	Non-connu	4	

Comme on le voit dans le Tableau 1, les groupes ne sont pas équilibrés par corpus ni par genre. Cela est dû en partie aux données, puisque chaque corpus ne présente pas un nombre d'heures égal et que les femmes et les hommes y sont inégalement représentés, mais aussi à la méthodologie de constitution de ce sous-corpus (cf. Section 3).

Dans ce qui suit, nous présentons les erreurs faites par le système et en dressons une typologie afin de répondre à diverses questions : quel type d'erreur le système commet-il ? quels phones sont les plus touchés par quel type d'erreur ? les erreurs ont-elles les mêmes causes dans les trois corpus ? entre les femmes et les hommes ?

### 4.1 Les erreurs d'alignement

Parmi les 802 occurrences du sous-corpus, on observe les erreurs détaillées en Tableau 2. La nature de la liaison effectivement réalisée est corrélée de façon significative à quelle consonne est alignée erronément par le système ( $\chi^2 = 205.68$ ,  $df = 12$ ,  $p < 0.0001$ ).

**Tableau 2.** Répartition des consonnes mal alignées selon la liaison effectivement réalisée.

Consonne alignée	Consonne réalisée				Total
	inaudible	0	t	z	
0	0	-	34	0	34
t	0	41	-	0	41
z	12	65	125	-	202
n	12	92	70	3	177
p	12	55	281	0	348
<b>Total</b>	<b>36</b>	<b>253</b>	<b>510</b>	<b>3</b>	<b>802</b>

Lorsqu'on observe le Tableau 2 par ligne, on constate que, lorsque le système n'a pas identifié de liaison alors qu'il y en avait une, la consonne non-repérée est toujours un [t], et inversement, lorsque le système a identifié un [t] de façon erronée, dans tous les cas, il n'y a en réalité pas de liaison produite. La réalisation d'un [t] est aussi le cas dans lequel il y a le plus d'erreurs d'alignement en [z] (62% des alignements en [z] viennent d'un [t]) et en [p]

(81% des alignements en [p] viennent d'un [t]). Les alignements erronés en [n], en revanche, se trouvent majoritairement dans les cas où il n'y a pas de consonne de liaison réalisée (52% des cas), suivi par les cas où un [t] était réellement produit (40%).

Inversement, lorsqu'on observe le Tableau 2 par colonne, on peut voir que dans les cas où l'audio était difficilement audible, le système fait autant d'erreurs d'alignement en [z], [n] et [p]. Dans les cas où il y a un [z] réalisé, la seule erreur est l'alignement avec [n] : [z] n'a que ces trois erreurs d'alignement, ce qui est probablement dû au fait qu'il s'agit d'une fricative dont la durée est, de fait, plus élevée que celle d'une occlusive comme [t]. En revanche, lorsqu'il y a en réalité un [t], l'erreur la plus courante est l'alignement avec un [p] (55% des cas). Dans la colonne des totaux, on constate d'ailleurs que [p] est la consonne la plus alignée erronément par le système (43%), suivi des liaisons en [z] (25%) et [n] (22%).

Les liaisons en [n] ont ceci de surprenant que /n/ est une continuante, dont la durée devrait jouer un rôle dans la détection, comme [z], mais qu'on n'attend pas comme consonne de liaison dans ce contexte : l'erreur d'alignement est donc doublement surprenante. Il aurait été possible que la forme masculine du Mot2 influence l'alignement erroné de la consonne de liaison, par exemple si la nasalité de la voyelle est anticipée par le locuteur.

**Tableau 3.** Comptes (nb) et proportions (%) de consonnes alignées en fonction de Mot2.

Consonne alignée:	0		t		z		n		p	
	nb	%	nb	%	nb	%	nb	%	nb	%
un	24	4,72	27	5,31	95	18,70	100	19,69	262	51,57
une	10	3,40	14	4,76	107	36,39	77	26,19	86	29,25

Cependant, si l'effet de Mot2 sur l'alignement est significatif ( $\chi^2 = 48.985$ ,  $df = 4$ ,  $p < 0.0001$ ), le Tableau 3 montre que *un* favorise nettement l'alignement avec [p] (51,57%) et *une* favorise, bien que moins nettement, l'alignement avec [z] (36,39%).

## 4.2 Les causes des erreurs d'alignement

Parmi les 7 causes d'erreurs repérées dans les données lors de la vérification auditive, on observe les proportions par consonne dans le Tableau 4.

**Tableau 4.** Causes des erreurs selon la consonne alignée.

Cause	0	t	z	n	p	total	%
bande passante	2	0	16	22	68	108	<b>13,47</b>
audio inaudible	0	0	8	5	5	19	<b>2,24</b>
bruit de fond	10	0	33	19	25	87	<b>10,85</b>
disfluences	3	22	21	30	34	110	<b>13,72</b>
chevauchement de parole	11	2	42	26	26	107	<b>13,34</b>
hypoarticulation	6	0	6	13	6	31	<b>3,87</b>
prosodie inhabituelle	1	17	10	47	69	144	<b>17,96</b>
aucune identifiable	1	0	66	15	115	197	<b>24,56</b>
<b>total</b>	<b>34</b>	<b>41</b>	<b>202</b>	<b>177</b>	<b>348</b>	<b>802</b>	<b>100</b>

Dans le Tableau 4, on constate qu'une grande partie des occurrences n'a pas de causes identifiables à l'oreille (197 occurrences, soit 24,56% des alignements erronés). Des études

futures tenteront de déterminer si des paramètres acoustiques fins dans ces occurrences peuvent expliquer les erreurs d'alignement.

Outre cette catégorie inexplicable, la source d'erreur la plus répandue est l'usage par les locuteurs d'une prosodie inhabituelle (voix soudainement plus forte, plus basse, chantée...), qui représente 18% des erreurs et favorise l'alignement avec un [p] (48% des erreurs liées à la prosodie résultent en l'alignement erroné d'un [p]). La deuxième source d'erreur est la présence de disfluences dans la séquence Mot1-Mot2, qui représente près de 14% des erreurs. Ces chiffres montrent une nouvelle fois, s'il en était besoin, que les variations prosodiques et les disfluences font partie intégrante de l'oralité et sont traitées sans problème par les humains mais restent problématiques pour les machines. Ce résultat prêche en faveur d'études plus poussées sur l'effet des disfluences au sens large sur les systèmes de reconnaissance, dans la lignée de Shriberg (1994) ou Goldwater *et al.* (2009).

Les problèmes de bande passante (13,47% des alignements erronés) favorisent également l'alignement d'un [p] (dans 63% des cas) mais défavorise l'alignement d'un [t] (0%), tandis que le chevauchement de parole (13,34% des alignements erronés) favorise à égalité l'alignement de [p] ou de [n] (dans 24,30% des cas) et défavorise toujours [t] (1,87%). Que l'audio soit inaudible ou que le locuteur hypo-articule, en revanche, sont peu source d'erreurs, ce qui est sans doute lié à la nature de nos corpus (parole journalistique ou de laboratoire).

Enfin, il est à noter que les alignements avec 0 ou avec [t] sont peu nombreux et causés par un nombre restreint de sources d'erreur : 32% des alignements avec 0 sont dus à du chevauchement de parole et 29% à du bruit de fond, tandis que 53% des [t] sont dus aux disfluences et 41% à une prosodie inhabituelle.

Les sources d'erreurs identifiées sont corrélées de façon significative à la consonne alignée par le système ( $\chi^2 = 259.35$ ,  $df = 28$ ,  $p < 0.0001$ ).

### 4.3 Les facteurs sociolinguistiques

Dans cette dernière sous-section, nous présentons quelles sources d'erreurs sont favorisées ou défavorisées par des facteurs stylistiques, c'est-à-dire sont fonction du corpus (4.3.1) ou du genre des locuteurs (4.3.2).

#### 4.2.1 Le style de parole

Dans la mesure où les conditions d'enregistrement, mais aussi les styles de parole, ne sont pas les mêmes dans nos trois corpus, il est attendu que les sources des erreurs seront représentées en proportions différentes entre ESTER, ETAPE et NCCFr (cf. Tableau 5).

**Tableau 5.** Cause des erreurs selon le corpus.

Cause	ESTER	ETAPE	NCCFr	Total
bande passante	97	11	0	108
audio inaudible	3	11	4	18
bruit de fond	40	47	0	87
disfluences	20	43	47	110
chevauchement de parole	6	95	6	107
hypoarticulation	9	19	3	31
prosodie inhabituelle	20	57	67	144
aucune identifiable	65	115	17	197
<b>total</b>	<b>260</b>	<b>398</b>	<b>144</b>	<b>802</b>

Il est peu surprenant de constater que les erreurs dans NCCFr (style informel mais amical) sont liées essentiellement au cadre dialogique : des prosodies inhabituelles (46,53% de NCCFr), des disfluences (32,64%) et, dans une moindre mesure, un son moins audible (si les locuteurs se sont trop éloignés du micro par exemple), quelques chevauchements de parole et quelques instances d'hypo-articulation.

Les résultats d'ESTER, en revanche, sont plus surprenants. La première source d'erreurs dans ESTER est le problème de bande passante (37,31%) alors qu'il s'agit d'enregistrements professionnels. Il est donc fort probable que ces erreurs soient liées à la parole d'invités téléphoniques dans les émissions. Les 15,38% d'erreurs dues au bruit de fond sont sans doute liés à la même cause.

Enfin, les résultats d'ETAPE ont ceci de surprenant qu'une grande proportion des erreurs n'a pas de source identifiée (28,89% des alignements erronés dans ETAPE) alors qu'il s'agit, là encore, d'enregistrements professionnels. Par ailleurs, les grandes proportions de chevauchement de parole (23,87%), de prosodie inhabituelle (14,32%) et de disfluences (10,80%) sont sans doute dues à la présence de débats dans le corpus.

Les sources des erreurs d'alignement sont corrélées de façon significative aux corpus ( $\chi^2 = 397.13$ ,  $df = 14$ ,  $p < 0.0001$ ).

#### 4.2.1 Le genre des locuteurs

Pour ce paragraphe sur le genre, nous avons supprimé les données des 4 locuteurs non-identifiés, ce qui nous amène à 222 occurrences prononcées par des femmes et 576 prononcées par des hommes.

**Tableau 6.** Causes des erreurs selon le genre du locuteur.

Cause	Femmes		Hommes	
	Compte	%	Compte	%
bande passante	39	<b>17,57</b>	69	<b>11,98</b>
audio inaudible	3	<b>1,35</b>	15	<b>2,60</b>
bruit de fond	28	<b>12,61</b>	59	<b>10,24</b>
disfluences	34	<b>15,32</b>	76	<b>13,19</b>
chevauchement de parole	17	<b>7,66</b>	89	<b>15,45</b>
hypoarticulation	8	<b>3,60</b>	23	<b>3,99</b>
prosodie inhabituelle	48	<b>21,62</b>	96	<b>16,67</b>
aucune identifiable	45	<b>20,27</b>	149	<b>25,87</b>
total	222		576	

Dans le Tableau 6, on peut voir que le taux d'erreurs dont la source n'est pas identifiable à l'oreille est légèrement plus haut chez les hommes que chez les femmes ( $\Delta = 5,6\%$ ). Ce résultat suggère que, lorsqu'il n'y a pas de raison manifeste pour mal aligner, les systèmes alignent mieux les femmes que les hommes, conformément à ce qui a été décrit dans la littérature (Adda-Decker & Lamel 2005).

Par ailleurs, la cause principale pour un alignement erroné est l'usage de prosodies inhabituelles, chez les femmes comme chez les hommes ( $\Delta = 4,95\%$ ), suivi des problèmes de bande passante chez les femmes, mais par le chevauchement de parole chez les hommes.

Les différences dans les sources d'erreurs d'alignement sont, bien que relativement petites, corrélées au genre des locuteurs de façon significative ( $\chi^2 = 17.797$ ,  $df = 7$ ,  $p < 0.015$ ).

## 5 Résultats : état des lieux de la liaison en français contemporain

Dans cette section, nous prenons en considération les 1695 occurrences de verbe *être* conjugué suivi de *un/une* qui ont été vérifiées auditivement, et nous proposons un état des lieux de la réalisation de la liaison facultative dans ce contexte sur la base des consonnes identifiées par la seconde autrice. Seuls les 36 fichiers-sous étiquetés comme inaudibles ont été mis de côté, ce qui nous amène à un total de 1659 occurrences.

### 5.1 Les contextes de réalisation de la liaison

#### 5.1.1 Les contextes gauche et droit

Tout d'abord, établissons quelles sont les formes conjuguées du verbe *être* qui donnent lieu au plus grand nombre de liaisons.

**Tableau 7.** Proportions de liaisons selon la forme du verbe *être*.

Mot 1	Effectif total	Pas de liaison	Liaison	% Liaison
est	1386	494	892	<b>64,36</b>
était	120	71	49	<b>40,83</b>
suis	40	16	24	<b>60,00</b>
sera	25	23	2	<b>8,00</b>
serait	20	13	7	<b>35,00</b>
es	19	15	4	<b>21,05</b>
sont	17	10	7	<b>41,18</b>
êtes	9	6	3	<b>33,33</b>
sommes	6	2	4	<b>66,67</b>
étaient	4	1	3	<b>75,00</b>
étais	4	3	1	<b>25,00</b>
fut	4	1	3	<b>75,00</b>
seras	2	2	0	<b>0,00</b>
serai	1	1	0	<b>0,00</b>
serais	1	0	1	<b>100,00</b>
seront	1	1	0	<b>0,00</b>
total	1659	659	1000	<b>60,28</b>

Dans le Tableau 7, de nombreuses formes ne sont pas assez représentées pour justifier des conclusions représentatives. En revanche, on constate que, dans nos trois corpus confondus, on trouve en général 60% de liaisons.

Il est intéressant de noter que les deux formes les mieux représentées, *est* et *était*, ne présentent pas les mêmes tendances : le présent de l'indicatif favorise nettement la liaison tandis que l'imparfait le défavorise ( $\Delta = 24\%$ ), similairement aux résultats de Durand & Lyche (2008).

En revanche, le Mot2 n'influence pas du tout la réalisation de la liaison (59,80% après *un*, 60,97% après *une* :  $\Delta = 1,17\%$  ;  $\chi^2 = 0.28851$ ,  $df = 2$ ,  $p = 0.8657$ ).

### 5.1.2 Les liaisons fautives

Parmi les 1659 liaisons observées ici, on repère 8 liaisons fautives, c'est-à-dire des liaisons réalisées alors qu'il n'y a pas de consonne graphique (*sera* prononcé *sera[t]*) ou des liaisons réalisées avec une consonne agrammaticale dans ce contexte (ex. *es* prononcé *es[t]*).

**Tableau 8.** Compte des liaisons fautives selon la forme du verbe *être*.

Mot 1	0	t	z
es	15	2	2
est	484	891	1
êtes	6	1	2
sera	23	2	0
sommes	2	1	3
suïs	16	1	23

Malgré le peu d'occurrences, on constate que, dans 7 cas sur 8, les erreurs proviennent de la réalisation d'un [t] alors que la consonne attendue est un [z] (5 cas) ou qu'aucune consonne n'est attendue (2 cas). Cela montre qu'il y a sans doute une pression paradigmatique en faveur de [t] comme consonne de liaison après un verbe conjugué, du moins après le verbe *être*, mais que cette pression reste légère.

### 5.2 Le style de parole

Il est communément admis depuis Delattre (1947) qu'entre le registre familial et le registre plus soutenu, la fréquence de production des liaisons variables augmente. Eychenne *et al.* (2014) montre notamment qu'entre les trois modalités plus ou moins surveillées prévues dans le protocole PFC, un pourcentage plus élevé de liaison est produit en lecture, puis en conversation guidée, puis en conversation spontanée.

**Tableau 9.** Comptes et proportions de liaisons selon le corpus.

Corpus	non-liaison	liaison	% liaison
ESTER	128	400	<b>75,76</b>
ETAPE	245	519	<b>67,93</b>
NCCFr	286	81	<b>22,07</b>
Total	659	1000	<b>60,28</b>

Les résultats du Tableau 9 confirment les résultats de Delattre (1947) et Eychenne *et al.* (2014). Plus le style est soutenu, comme dans le discours préparé, voire prompté, d'ESTER, plus on trouve de liaisons, et plus le style est informel, comme dans NCCFr, moins on en trouve ( $\Delta = 53,69\%$  entre ESTER et NCCFr). Cette corrélation entre corpus et réalisation de la liaison est significative ( $\chi^2 = 297.86$ ,  $df = 4$ ,  $p < 0.0001$ ).

### 5.3 Le genre des locuteurs

Qui des hommes ou des femmes réalise le plus de liaisons ? Comme précisé en Section 2, les résultats concernant le sexe des locuteurs sont controversés : pour certains ce sont les femmes (Malécot 1975, Booij & De Jong 1987), pour d'autres ce sont les hommes (Ashby 1981)

tandis que pour les plus récents, il n'y a pas de différence statistiquement significative entre les deux (Durand *et al.* 2011).

**Tableau 10.** Comptes et proportions de liaisons par genre et par corpus.

Corpus	Femmes				Hommes			
	0	t	z	% liaison	0	t	z	% liaison
ESTER	54	132	1	<b>71,12</b>	74	258	9	<b>78,30</b>
ETAPE	93	149	8	<b>62,80</b>	152	349	9	<b>70,20</b>
NCCFr	145	39	3	<b>22,46</b>	141	36	3	<b>21,67</b>

Dans le Tableau 10 portant sur 1655 hommes et femmes (les non-identifiés n'ont pas été pris en compte), on peut constater que les hommes font plus de liaison que les femmes dans ESTER ( $\Delta = 7,18\%$ ) et ETAPE ( $\Delta = 7,40\%$ ) mais que la tendance s'annule dans le discours plus spontané de NCCFr ( $\Delta = 0,79\%$ ).

La tendance (tout juste) significative en faveur de la liaison chez les hommes dans ETAPE ( $\chi^2 = 3.8693$ ,  $df = 1$ ,  $p = 0.04918$ ) soutient les résultats d'Ashby (1981) mais dans l'ensemble, nos résultats ne sont significatifs ni dans ESTER ( $\chi^2 = 3.0069$ ,  $df = 1$ ,  $p = 0.08291$ ) ni dans NCCFr ( $\chi^2 = 0.0032815$ ,  $df = 1$ ,  $p = 0.9543$ ), ce qui soutient les résultats de Durand *et al.* (2011).

## 5.4 Étude diachronique

Ågren, dans son *Étude sur quelques liaisons facultatives dans le français de conversation radiophonique* (1973), fournit, entre autres, des chiffres sur la liaison après le verbe *être* conjugué provenant de 134 émissions de radio des années 1960 et 1961. Son corpus étant comparable à notre corpus ETAPE, bien que nous nous limitons aux cas où le verbe *être* est suivi de l'article indéfini singulier, nous proposons ici une comparaison entre ses données et les nôtres, afin d'estimer l'évolution de l'usage entre le début des années 1960 et les années 2010.

**Tableau 11.** Comptes et proportions de liaisons par forme du verbe *être* conjugué dans Ågren (1973) et ETAPE (Gravier *et al.* 2012).

Mot1	Ågren			ETAPE		
	non liaison	liaison	% liaison	non liaison	liaison	% liaison
suis	6	7	<b>53,85</b>	14	11	<b>44,00</b>
est	15	1240	<b>98,80</b>	192	464	<b>70,73</b>
sommes	0	1	<b>100,00</b>	0	4	<b>100,00</b>
êtes	2	1	<b>33,33</b>	2	3	<b>60,00</b>
sont	1	5	<b>83,33</b>	6	7	<b>53,85</b>
était	27	126	<b>82,35</b>	23	18	<b>43,90</b>
fut	1	2	<b>66,67</b>	1	1	<b>50,00</b>
serait	8	6	<b>42,86</b>	8	3	<b>27,27</b>
<b>Total</b>	<b>60</b>	<b>1388</b>	<b>95,86</b>	<b>246</b>	<b>511</b>	<b>67,50</b>

D'une manière générale, comme on peut le voir dans le Tableau 11, la tendance est à la baisse, puisque, dans la parole radiophonique des années 1960, près de 96% des liaisons sont réalisées après le verbe *être* conjugué, contre 67,50% dans les années 2010 ( $\Delta = 28,36\%$ ).

Les quelques formes pour lesquelles la tendance ne se confirme pas (*sommes* et *êtes*) sont trop peu représentées pour remettre les résultats en question.

## 6. Conclusion et discussion

Le présent article avait deux objectifs.

D'abord, il s'agissait de présenter une typologie des erreurs commises par l'alignement automatique lorsque le système y est autorisé. Nous avons montré, entre autres, que les modulations de la voix, ou variations prosodiques, sont une source majeure d'erreurs, et qu'elles méritent qu'on leur consacre davantage de recherche que ce qui a été proposé jusqu'à présent, dans les cadres théoriques comme dans les applications pratiques. Nous avons aussi montré que les erreurs ne sont pas dues aux mêmes causes dans les différents corpus, ce qui était tout à fait attendu, et confirmé que, toutes choses égales par ailleurs, les systèmes de reconnaissance sont plus performants sur les voix de femmes.

Ensuite, nous avons présenté un rapide panorama de la liaison facultative en français, dans le contexte spécifique « verbe *être* conjugué suivi de *un* ou *une* ». Cette étude a montré que les formes prises par le verbe *être* favorisent la liaison à des degrés divers, ce qui peut être dû soit à un effet de facteurs syntaxiques (temps, mode...) soit à un effet de fréquence. Le peu de liaisons fautives montre en tous cas une légère pression paradigmatique en faveur de [t], mais avec seulement 8 liaisons fautives sur 1659 occurrences, on peut suggérer sans risque que cette pression reste marginale. Pour ce qui est des facteurs individuels, le style de parole joue en effet un rôle dans la réalisation de la liaison, mais pas le sexe des locuteurs. Enfin, la réalisation de la liaison a notablement baissé ces 50 dernières années.

Cette étude préliminaire a mis en lumière de nouvelles questions de recherche. Entre autres, nous souhaitons analyser plus finement les variations prosodiques et les types d'erreur qu'elles favorisent, et comparer leur traitement par la machine au traitement opéré par les humains. Nous souhaitons aussi nous pencher plus en détail sur les paramètres acoustiques fins des fichiers audio (durées, valeurs formantiques, intensité, centre de gravité, etc.) afin d'établir si de telles caractéristiques expliquent, du moins en partie, les erreurs d'alignement, notamment dans le cas des 24,56% d'erreurs pour lesquelles aucune cause n'a pu être identifiée à l'oreille. Enfin, cette étude préliminaire s'est avérée fructueuse et nous espérons l'étendre à d'autres contextes de liaison.

Cette recherche a été partiellement financée par le prix Excellence de l'Institut DATAIA et la MSH Paris-Saclay, par le projet SON-DISOURS dans le cadre de l'appel Émergence de l'IdEX U. de Paris (ANR-18-IDEX-0001) et par le Labex EFL (ANR-10-LABX-0083).

## Références bibliographiques

- Adda-Decker, M. & Lamel, L. (1998). Pronunciation variants across systems, languages and speaking style. *Proceedings of Modeling Pronunciation Variation*, pp. 1-6
- Adda-Decker, M., Boula de Mareüil, Ph., Adda, G. & Lamel, L. (2005). Investigating syllabic structures and their variation in spontaneous French. *Speech Communication*, vol. 46, n° 2, p. 119-139
- Adda-Decker, M. & Lamel, L. (2005). Do speech recognizers prefer female speakers? *Interspeech 2005*: 2205-2208
- Adda-Decker, M., Fougeron, C., Gendrot, C., Delais-Roussarie, E. & L. Lamel. (2012). La liaison dans la parole spontanée familière : une étude sur grand corpus, *Revue Française de Linguistique Appliquée*, vol. 17, n° 1, p. 113-128, 2012.
- 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 William W. Cressy and Donna Jo Napoli (ed.). *Linguistic Symposium on romance languages 9*. Washington D.C.: Georgetown University Press, 46-57.

- Booij, G. & De Jong, D. (1987). The domain of liaison: theories and data. *Linguistics*, 25, 1005-1025
- Coleman, J., Renwick, M., & Temple, R. (2016). Probabilistic underspecification in nasal place assimilation. *Phonology*, 33(3), 425-458.
- 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. <https://doi.org/10.1017/S0959269507003158>
- Dutrey, C., Adda-Decker, M., Yamaguchi, N. (2016). Alignement de séquences phonétiques pour une analyse phonologique des erreurs de transcription automatique. *Actes de La Conférence Conjointe JEP-TALN-RECITAL 2016*. Volume 1 : JEP. pp. 46–54
- Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants [Editorial]. *Journal of Phonetics*, 39(3), 253–260. [https://doi.org/10.1016/S0095-4470\(11\)00055-6](https://doi.org/10.1016/S0095-4470(11)00055-6)
- 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
- Fosler-Lussier, E. & Morgan, N. (1999). Effects of speaking rate and word frequency on pronunciations in conversational speech. *Speech Communication* 29, 137–158
- 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 Newshase II Evaluation Campaign for the Rich Transcription of French Broadcast News. *Interspeech 2005*, 2453–2456
- Gauvain, J.-L., Lamel, Lori & Adda, G. (2002). The LIMSI broadcast news transcription system. *Speech communication* 37 (1-2), 89–108
- Goldwater, S., Jurafsky, D. & Manning, C. D. (2009). Which words are hard to recognize? Prosodic, lexical, and disfluency factors that increase speech recognition error rates. *Speech Communication* 52 (2010) 181–200
- 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. *Proceedings of LREC Eighth international conference on Language Resources and Evaluation*
- Hallé, P. & Adda-Decker, M. (2007). Voicing assimilation in journalistic speech. *16th International Congress of Phonetic Sciences*, 2007, 493–496
- 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
- Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America*, 128(2), 839–850. <https://doi.org/10.1121/1.3459842>
- 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
- Maschler, Y. & Schiffrin, D. (2015). Discourse markers: Language, meaning, and context. *The Handbook of Discourse Analysis*, 2<sup>nd</sup> edition. Chichester, UK: John Wiley & Sons, Ltd. 189-221
- Shinozaki, T. & Furui, S. (2001). Error analysis using decision trees in spontaneous presentation speech recognition. *Proceedings of ASRU 2001*
- Shriberg, E. E. (1994). *Preliminaries to a Theory of Speech Disfluencies*. PhD dissertation. University of California at Berkeley
- Siegler, M. & Stern, R. (1995). On the effects of speech rate in large vocabulary speech recognition systems. *Proceedings of ICASSP*
- Torreira F., Adda-Decker M. & Ernestus M. (2010). The Nijmegen Corpus of Casual French. *Speech Communication*, Elsevier: North-Holland, 2010, 52 (3)
- Vasilescu, I., Rosset, S. & Adda-Decker, M. (2010). On the role of discourse markers in interactive spoken question answering systems. *Proceedings of LREC'10*
- Webb, E. (2009). Minimalism and French /R/: Phonological representations in phonetically based phonology. *Journal of French Language Studies*, 19(1), 87-115. doi:10.1017/S095926950800358X
- Wu, Y. & Adda-Decker, M. (2021). Réduction des segments en français spontané : apports des grands corpus et du traitement automatique de la parole. *Corpus*, n° 22, 2021