

LA LIAISON EN FRANÇAIS ET LA THÉORIE DE L'OPTIMALITÉ

Julien Eychenne

Armand Colin | « Langue française »

2011/1 n°169 | pages 79 à 101

ISSN 0023-8368

ISBN 9782200926816

Article disponible en ligne à l'adresse :

<http://www.cairn.info/revue-langue-francaise-2011-1-page-79.htm>

!Pour citer cet article :

Julien Eychenne, « La liaison en français et la théorie de l'optimalité », *Langue française* 2011/1 (n°169), p. 79-101.

DOI 10.3917/lf.169.0079

Distribution électronique Cairn.info pour Armand Colin.

© Armand Colin. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

La liaison en français et la théorie de l’optimalité

1. INTRODUCTION

La liaison constitue, avec le « e caduc » et le « h aspiré », l’un des trois problèmes fondamentaux de la phonologie du français¹. Si elle constitue un terrain empirique de choix pour les phonologues et a permis d’éprouver ou de motiver de nombreuses propositions théoriques, elle est encore aujourd’hui l’objet de vifs débats.

Nous offrons dans cet article une analyse de la liaison dans le cadre de la théorie de l’optimalité (Prince & Smolensky 1993), un cadre grammatical utilisant des contraintes. Notre analyse met à profit un corpus de sept enquêtes issues du projet PFC, corpus dont nous dégageons les principaux patrons variationnels. Ce travail s’appuie sur diverses avancées théoriques récentes : nous proposons notamment de considérer que les consonnes de liaison (CL) sont des consonnes à exponence morphologique spécifique qui sont ciblées par des contraintes lexicales co-indexées (Pater 2006).

Nous commençons par donner dans la section 2 un panorama du phénomène de la liaison en français et offrons un aperçu des principales interprétations qui en ont été faites. La section 3 introduit la théorie de l’optimalité et donne un aperçu critique d’une approche de la liaison qui a été formulée dans ce cadre. La section 4 présente et décrit le corpus que nous avons exploité. La section 5 présente notre analyse de la liaison. Enfin, nous discutons en conclusion les apports de l’analyse proposée ainsi que ses limites.

1. Je remercie Jacques Bres, Jacques Durand, Bernard Laks ainsi que deux relecteurs anonymes de la revue *Langue française* pour leurs remarques sur une version préliminaire de ce travail. Toutes les erreurs restantes sont miennes.

2. APPROCHES CONTEMPORAINES DE LA LIAISON : UN BREF APERÇU

2.1. Faits de base

La liaison en français est la manifestation d'un phénomène de joncture (ou sandhi) entre deux unités morpho-lexicales dans un environnement linguistique donné². Nous donnons en (1) un récapitulatif des faits de base.

- (1) Réalisation de la liaison : faits de base
 - a. Masculin singulier
 - *il est grand* [ilɛgb̃ɑ̃]
 - *grand ami* [gb̃ɑ̃tami]
 - b. Féminin singulier
 - *Elle est grande* [ɛlɛgb̃ɑ̃d]
 - *grande amie* [gb̃ɑ̃dami]
 - c. Masculin pluriel
 - *Ils sont grands* [ilsɔ̃gb̃ɑ̃]
 - *grands amis* [gb̃ɑ̃zami]
 - d. Féminin pluriel
 - *Elles sont grandes* [ɛlsɔ̃gb̃ɑ̃d]
 - *grandes amies* [gb̃ɑ̃dzami]

Comme on le voit dans ces exemples, une CL apparaît en position prévocalique au masculin singulier ou au pluriel (aux deux genres). On opposera les CL, dites « latentes » (cf. 1a-c-d), aux consonnes finales « fixes » qui sont réalisées quel que soit le contexte (cf. 1b). La CL coïncide le plus souvent avec la consonne que l'on trouve au féminin et/ou en dérivation dans certains adjectifs, mais ce n'est pas toujours le cas (voir aussi Sauzet 2004) :

- (2) Liaison au masculin *vs* féminin *vs* dérivation
 - a. Consonne identique
 - forme liaisonnante du masculin : *petit* [ptit]
 - féminin : *petite* [ptit]
 - dérivation : *petitesse* [ptites]
 - b. Consonne différente
 - forme liaisonnante du masculin : *grand* [gb̃ɑ̃t]
 - féminin : *grande* [gb̃ɑ̃d]
 - dérivation : *grandeur* [gb̃ɑ̃dœv]

La question de la liaison subsume un ensemble de problématiques relativement indépendantes qu'il est utile de distinguer. Le premier problème est celui du statut de la CL : est-elle épenthétique ou lexicale ? Dans le second cas, appartient-elle au mot de gauche ou au mot de droite ? Cette question est sans doute la plus cruciale du point de vue de la théorie phonologique

2. Il existe une telle quantité de travaux sur la liaison qu'il est impossible d'être exhaustif. Nous renvoyons le lecteur intéressé à Chevrot *et al.* (2005) ; Durand & Lyche (2008) ; Dell (1985) ; Encrevé (1988) ; Tranel (1981, 1995) et aux références qu'ils citent pour un panorama représentatif.

car elle est étroitement liée à la conception que l'on se fait du lexique d'une part et des représentations phonologiques d'autre part. Le second problème est celui des contraintes linguistiques qui pèsent sur la réalisation de la liaison (Bonami *et al.* 2004 ; Selkirk, 1980 : III ; Laks 2005) : elle est généralement considérée comme catégorique lorsque le mot 1 dépend syntaxiquement du mot 2 (*ils ont ; très ensoleillé*), variable dans les verbes/auxiliaires ou quand le mot 1 régit le mot 2 (*ils sont arrivés ; des amis étrangers*³), erratique⁴ après une frontière syntaxique majeure (*son amant \cap est jeune*⁵). L'exemple canonique est *Un marchand de draps anglais*, où la liaison est normalement réalisée si *anglais* modifie *draps*, mais est jugée impossible si *anglais* modifie *un marchand de draps*, la frontière syntaxique entre *draps* et *anglais* étant plus forte dans le second cas que dans le premier (Selkirk 1980).

La troisième question est celle des contraintes extra-linguistiques qui pèsent sur la réalisation de la liaison (variable) : le phénomène est notamment étroitement corrélé avec le style et le niveau d'éducation des locuteurs, la liaison variable n'étant généralement réalisée à l'oral que par les locuteurs de niveau socio-culturel plus élevé (Armstrong 2001).

Enfin, la quatrième question est celle du rapport aux données : il existe une vaste littérature sur la question (surtout dans le paradigme générativiste) que l'on peut faire remonter en dernière analyse aux travaux normatifs de P. Delattre (1966) et P. Fouché (1956), essentiellement à l'usage des apprenants du français. Comme le rappellent J. Durand et C. Lyche (2008), même certains travaux récents analysent des « cas » de liaisons dans les adjectifs antéposés dans des formes du type *sot ami* [sotami], cas qui semblent défier l'intuition de la plupart des locuteurs natifs. En marge de cette tradition s'est constitué un ensemble de travaux plus empiriques, essentiellement dans les domaines de la sociolinguistique (voir Ågren 1973 ; De Jong 1994 *inter alia*) et de l'acquisition (voir, par exemple, Chevrot *et al.* 2007). Les travaux en acquisition sont bien sûr très utiles, mais s'il existe de nombreux arguments pour un rattachement des CL à droite à une étape précoce (éventuellement avec stockage de formes supplétives, cf. *ours* [lurs] ~ [zurs] ~ [nurs]), cela n'implique pas qu'il en aille de même chez l'adulte⁶. Quant aux travaux sociolinguistiques, J. Durand et C. Lyche (*op. cit.*) soulignent à juste titre qu'ils ne reflètent pas véritablement la réalité du français dans sa diversité et sont, pour la plupart, consacrés à des groupes de locuteurs très spécifiques (entretiens radiophoniques, discours

3. Uniquement au pluriel dans les structures Nom+Adj (cf. **soldat* [t] *anglais*).

4. Nous reprenons ici la terminologie d'Encrevé (1988), qui est préférable à la tradition normative « obligatoire », « facultative » et « interdite » (Delattre 1966).

5. Nous utilisons le symbole « \cap » pour représenter l'absence de liaison.

6. Côté (2005) note par exemple que certaines variétés de français en Ontario (Canada), dans lesquelles les occlusives à l'initiale de mot sont aspirées (*petit tamis* [ptit^hami]), présentent des consonnes de liaison resyllabifiées à droite sans aspiration (*petit ami* [pitami]) à l'instar des consonnes finales enchaînées. Ce phénomène milite clairement pour une représentation des CL distincte de celle des consonnes initiales.

d'hommes politiques, etc.). Les recherches menées au sein du projet PFC ont précisément pour but de combler cette lacune.

2.2. La représentation des CL

Une théorie adéquate de la liaison doit, dans l'absolu, être en mesure d'intégrer toutes ces dimensions. Pour ce faire, il faut tout d'abord clarifier le statut de la CL. Dans la masse des travaux sur la question, on peut identifier trois grandes tendances. La première considère que la CL est fondamentalement épenthétique : la forme de base est la forme courte (*petit* [pti]) et la CL est insérée ou non selon l'environnement. Cette approche a été développée de manière particulièrement articulée par B. Tranel (1981) dans le cadre de la phonologie générative concrète : par exemple, une consonne /C/ peut être insérée après un verbe en position prévocanique (suivi d'une frontière morphologique faible), cette consonne étant un /z/ si le verbe a le trait [- 3^e personne] (*je suis* [z] *arrivé*) et un /t/ s'il a le trait [+ 3^e personne] (*il est* [t] *arrivé*). Un adjectif comme *petit* aura, quant à lui, une forme du type /pti/ (/t/), où la consonne entre parenthèses ne fait pas partie de la représentation phonologique : c'est un marqueur phonologique spécifié (le cas échéant) pour chaque entrée lexicale (*op. cit.* : 238). Cette analyse reprend, dans l'esprit sinon dans la lettre, le travail de M. Durand (1936) sur le genre grammatical en français parlé : l'auteure, qui a inventorié de manière rigoureuse toutes les variations de genres dans les adjectifs, établit un ensemble de règles qui permettent de dériver la forme du féminin du masculin par adjonction d'une désinence. On peut ainsi distinguer les mots qui ont pour désinence /t/ (*grand~grande*), ceux qui ont /s/ (*gros~grosse*), etc. Cette analyse est tout aussi valable pour la liaison, que l'on stocke une consonne spécifique pour la liaison d'une part et pour le féminin d'autre part, ou que l'on dérive la CL à partir de la consonne du féminin (cf. Tranel, 1981 : 238-9).

Le deuxième type d'approche considère la CL comme une consonne latente. Cette approche est évoquée en passant par L. Bloomfield (1933 : 217-8) et s'est imposée comme la vision dominante dans la phonologie générativiste. Dans la phonologie linéaire, cette consonne appartient à la forme de base du mot de gauche et est tronquée sous certaines conditions. L'avènement de la phonologie non-linéaire raffinerait ce modèle en opérant un découplage entre positions prosodiques (unités de temps ou de poids) et mélodie (contenu segmental)⁷ : dans ce paradigme, la CL est considérée comme une consonne flottante. Les approches par latence ont ceci d'attrayant qu'elles s'accordent bien avec les faits diachroniques : comme le rappelle B. Laks (2005), l'évolution du latin populaire au français s'est caractérisée très tôt par un amuïssement des consonnes finales dont la liaison n'est qu'un reliquat. Cependant, les approches linéaires, pour pouvoir fonctionner, postulent des règles dérivationnelles opaques qui ne sont

7. Voir Clements & Keyser (1983) ; Hyman (1985) ; Durand (1986) ; Encrevé (1988) et Tranel (1995) pour différentes implémentations de cette intuition.

pas justifiées en surface (Tranel 1981 ; Eychenne, 2006 : chap. 2 pour une synthèse). Les approches par flottance supposent, quant à elles, un double encodage (tacite) de la linéarité : au niveau du squelette et au niveau de la chaîne segmentale (Sauzet 1999). Par ailleurs, elles ne fonctionnent que dans les cas où il y a isomorphie entre la CL et la consonne en dérivation.

La troisième famille d'approche est celle des approches que nous appellerons, faute de mieux, « lexicalisantes » : celles-ci ont en commun le fait qu'elles considèrent que les représentations lexicales peuvent être redondantes et surinformées. Les traitements supplétifs donnent, par exemple, au mot *grand* deux formes alternantes, à savoir [gʁã] et [gʁãt] (Klausenburger 1984). Ce type d'approche a connu un regain d'intérêt dans la lignée des travaux en grammaires d'usage et en grammaire cognitive (Croft & Cruse 2004 ; Langacker 2008). Ces approches voient le lexique non pas comme une liste minimalement spécifiée d'items (disons lexique régulier+exceptions), mais comme un réseau complexe, redondant, de constructions en relation les unes avec les autres. La dichotomie entre formes régulières et irrégulières est donc abandonnée au profit d'une vision graduelle de la grammaticalité, laquelle n'est plus que le résultat de la prédominance de certains schèmes (généraux) sur d'autres (spécifiques). La liaison est ici considérée comme un phénomène émergent (Bybee 2001 ; Laks 2005). Ces approches permettent d'expliquer de manière attrayante pourquoi, pour de nombreux locuteurs, la liaison est systématiquement réalisée dans *accent [t] aigu*, mais jamais dans *accent étranger* : dans le premier cas, il s'agit d'une expression figée suffisamment fréquente pour garder une trace prégnante en mémoire et conserver ainsi la liaison historique ; dans le second cas, il s'agit d'une construction moins fréquente, potentiellement construite en ligne, et qui suit le schème général des séquences Nom singulier+Adjectif. Toutefois, ces analyses ne permettent pas d'expliquer le statut spécifique des CL (Tranel, 2000 : 50-1 ; Bonami *et al.* 2005) : divers travaux ont en effet noté que les CL, contrairement aux consonnes fixes, peuvent être précédées d'une pause et enchaînées à droite (*grand ami* [gʁã // tami] vs. *grande amitié* *[gʁã // damitje]). Les phénomènes de liaison disloquée, où la CL est déportée à droite (*donne-moi-z-en* [dɔnmwazã], *bien mieux isolé* [bjɛmjɔnzɔle], cf. Côté 2005), militent également pour ce statut spécifique. Le français méridional offre un argument phonologique supplémentaire en faveur d'un traitement différencié des latentes et des fixes. Ainsi, les consonnes fixes finales se comportent comme des consonnes de coda et provoquent l'abaissement des voyelles moyennes précédentes, même si elles sont ré-enchaînées à droite (*net avantage* [nɛtava^Ntaʒɔ]). Les CL n'ont pas cet effet et se comportent comme des consonnes d'attaque (*très avantageux* [tʁɛzava^Ntaʒɔ]) : tout se passe donc comme si elles étaient extra-syllabiques, la voyelle précédente se comportant comme une voyelle en syllabe ouverte (Durand 1990). Cette asymétrie s'explique très bien dans un cadre qui reconnaît le caractère unique de ces consonnes, mais il est difficile d'en rendre compte dans un cadre supplétif. Notons par ailleurs que les approches constructionnistes semblent souffrir

en l'état actuel d'un manque de précision formelle qui rend leurs prédictions difficilement vérifiables⁸.

Les approches lexicalisantes ont le mérite de mettre en avant l'importance des idiosyncrasies dans la langue et nous invitent à abandonner une vision trop minimaliste du lexique. La position générativiste classique n'est, en effet, soutenable que si l'on se contente de considérer un sous-ensemble des données, à savoir l'ensemble des mots où il y a isomorphie entre la CL et la consonne en dérivation⁹. Il apparaît donc nécessaire d'incorporer dans les représentations sous-jacentes la richesse des idiosyncrasies lexicales tout en reconnaissant la spécificité des consonnes latentes. La proposition d'O. Bonami *et al.* (2005), formulée dans le cadre de la *Head-driven Phrase Structure Grammar* (HPSG, une théorie lexicaliste reposant sur des contraintes), nous semble représenter un compromis adéquat, et l'approche que nous développons dans la section 5 lui est apparentée. Ces auteurs distinguent deux parties dans les représentations phonologiques : le corps et l'appendice, lequel peut être vide. Dans ce modèle, le mot *petit* a pour corps <pəti> et pour appendice <t>, alors que *joli* a pour corps <ʒoli> et un appendice vide <>. Cette conception reprend, en l'intégrant dans un cadre formel explicite, le concept de désinence de M. Durand (1936). Nous en proposons une interprétation légèrement différente (mais apparentée) dans le cadre de la théorie de l'optimalité, que nous présentons dans la section suivante.

3. LA THÉORIE DE L'OPTIMALITÉ

3.1. La théorie standard

La théorie de l'optimalité (*Optimality Theory*, dorénavant OT) est issue d'une série de travaux menés aux États-Unis dans les années 90 (Prince & Smolensky 1993 ; McCarthy & Prince 1993, 1995). Elle trouve son origine dans les travaux menés par P. Smolensky et ses collaborateurs sur la mesure de l'harmonie des modèles, notamment au sein du groupe connexionniste *Parallel Distributed Processing* (Smolensky 1988). Depuis N. Chomsky et M. Halle (1968), la composante phonologique était essentiellement vue comme un ensemble de règles (généralement ordonnées) qui, appliquées à un input, produisent un output. OT conserve la distinction input/output, mais abandonne la notion de règle au profit d'un ensemble de contraintes hiérarchisées. Le modèle contient trois composants principaux : un ensemble CON de contraintes universelles, hiérarchisé différemment dans chaque langue ; un module GEN qui, étant donné un input et l'ensemble CON, génère un ensemble (théoriquement infini) de candidats et de violations de contraintes ; un module EVAL qui évalue l'ensemble des candidats

8. Ainsi, Langacker (2008) souligne-t-il, de manière très honnête, le caractère très programmatique de sa grammaire cognitive.

9. Rappelons que la consonne du féminin n'est pas nécessairement celle que l'on retrouve dans les noms dérivés (voir Bonami *et al.* 2005 pour une analyse détaillée).

en fonction de la hiérarchie de contraintes dans CON et retourne le (ou les) candidat(s) gagnant(s), c'est-à-dire le ou les output(s). Le modèle est donc un modèle génératif typique qui peut être représenté comme une double application de fonction : $EVAL(GEN(I)) = O$, où I et O représentent l'input et l'output respectivement. Il est d'usage de représenter l'évaluation d'une paire input/output au moyen d'un *tableau* où les colonnes représentent les contraintes (hiérarchisées de gauche à droite) et les lignes, les candidats : chaque cellule indique le nombre de fois que le candidat viole la contrainte correspondante, chaque violation étant représentée par une étoile. Nous donnons ci-dessous un tableau pour une mini-grammaire $C1 \gg C2 \gg C3$ ¹⁰ et trois candidats :

(3) Tableau avec trois contraintes et trois candidats

/input/	C1	C2	C3
☞ [cand1]		*	***
[cand2]	*!		*
[cand3]	*!		

Dans ce tableau, le candidat [cand1] viole la contrainte C2 une fois et C3 trois fois ; [cand2] viole C1 et C3 une fois et [cand3] viole C1 une fois. Une propriété fondamentale d'OT est le principe de domination stricte, selon lequel la satisfaction d'une contrainte plus élevée dans la hiérarchie est plus importante que la satisfaction de toutes celles qui sont situées plus bas. Dans notre exemple, [cand1] est le candidat gagnant car les deux autres violent C1 une fois alors que [cand1] satisfait cette contrainte : [cand2] et [cand3] ont donc une violation fatale de C1 (symbolisée par le point d'exclamation¹¹) et [cand1] est évalué comme le candidat optimal (symbolisé par « ☞ »), bien qu'il viole C2 une fois et C3 trois fois.

Le tableau 3 illustre une autre propriété fondamentale de la théorie, à savoir le blocage harmonique (*harmonic bounding*). En effet, l'ensemble des violations de [cand3] est strictement inclus dans l'ensemble des violations de [cand2]. En vertu du principe de stricte dominance, cela implique que [cand2] ne pourra jamais être le candidat optimal, quel que soit l'ordre des contraintes. En pratique, cela signifie non seulement que l'inclusion du candidat [cand2] dans le tableau est inutile, mais aussi que dans l'ensemble infini des candidats, il n'en existe qu'un nombre fini qui ont une chance d'émerger comme le candidat optimal. Il est donc d'usage de ne considérer que les candidats les plus raisonnables pour une grammaire donnée.

10. 'C1 \gg C2' se lit 'la contrainte C1 domine la contrainte C2'.

11. Les cases grisées signalent que ces contraintes ne sont plus pertinentes pour les candidats en question puisqu'ils ont été évalués comme sub-optimaux pour une contrainte superordonnée.

La théorie de l'optimalité est avant tout une théorie de l'interaction de contraintes. Dans le domaine phonologique, on distingue généralement deux grandes familles de contraintes : les contraintes de marque (ou contraintes structurales), qui portent sur les représentations structurales des candidats (ex. : ATTAQUE « une syllabe doit avoir une attaque »), et les contraintes de fidélité, qui pénalisent toute divergence entre l'input et le candidat. Dans cette dernière famille, les contraintes les plus importantes sont celles de la famille MAX, qui pénalisent les effacements (« maximisation des éléments de l'input dans l'output »), et DEP, qui pénalisent les épenthèses (« dépendance des éléments de l'output dans l'input »).

3.2. Analyse de la liaison en OT

Parmi les traitements de la liaison en OT, la proposition la plus articulée est celle qu'a défendue B. Tranel dans une série de travaux (voir Tranel 1996, 2000) ¹². Celui-ci propose un traitement unifié de la liaison et de l'épenthèse en OT, qui renoue en quelque sorte avec l'approche qu'il avait défendue en 1981, à savoir que les CL sont des segments épenthétiques lexicalement disponibles (notés entre accolades) qui peuvent être insérés lorsque certaines conditions syllabiques sont remplies. B. Tranel propose que la réalisation de la CL est cruciallement provoquée par une contrainte anti-hiatus *VV : dans le cas général, les consonnes latentes ne se réalisent pas sous l'effet de la contrainte DEP(L) qui pénalise l'insertion des CL. Mais, la grammaire incluant la domination *VV » DEP(L), les CL sont insérées pour éviter le hiatus (4) :

(4) Réalisation de la liaison (Tranel, 2000 : 52)

/pəti {t} ano/	*VV	DEP(L)
^{CL} pœti t ano		*
pœti ano	*!	

Cette approche est élégante dans la mesure où elle permet de connecter le problème de la liaison avec la tendance générale qu'ont les langues du monde à éviter les hiatus, en particulier en environnement dérivé. Elle permet donc d'envisager la liaison comme un cas général de sandhi qui est partiellement idiosyncrasique (la CL est spécifiée lexicalement), mais résulte de l'interaction de contraintes très générales et bien motivées d'un point de vue typologique.

L'analyse de B. Tranel pose néanmoins plusieurs problèmes. Elle soulève tout d'abord une difficulté d'ordre théorique quant à la théorie de la fidélité qui la

12. Steriade (1999) propose que la forme de liaison soit la forme du féminin : là encore, cette approche ne fonctionne que pour le sous-ensemble d'items qui ont effectivement la même forme en liaison et en dérivation, et elle requerrait de traiter le français méridional à part puisque dans la plupart de ces variétés, il existe un marqueur morphologique du féminin (cf. [pøtit] forme liaisonnante du masc. vs [pøtitə] féminin).

sous-tend. Pour le comprendre, voyons comment B. Tranel envisage le statut des CL :

[...] je ne prendrai pas position en ce qui concerne la nature de la source lexicale des consonnes de liaison. Ce qu'il est important de retenir pour mon propos, c'est que l'existence et la qualité d'une consonne de liaison ont une détermination lexicale cruciale. (Tranel, 2000 : 51)

Il apparaît donc que, quelle que soit sa nature, la CL doit être présente en input. Il est alors difficile de concevoir comment l'apparition de la CL en output peut violer une contrainte de la famille DEP qui demande à ce que tout élément de l'output ait un correspondant en input : si la CL est bien présente en input, elle ne peut pas causer de violation de DEP, au sens où cette contrainte est normalement comprise dans la *Théorie de la Correspondance* (McCarthy & Prince 1995). La position défendue par B. Tranel est d'autant plus surprenante qu'il soutient que :

DEP(C) domine DEP(L) [...] car de façon inhérente, pour tout phénomène d'épenthèse, l'utilisation phonétique d'éléments qui existent lexicalement respecte FIDELITE (sic) davantage que l'introduction d'éléments entièrement nouveaux.

Il nous est difficile de voir comment ces positions sont compatibles : soit un segment est présent en input et son apparition en output ne viole pas DEP (mais viole MAX), soit il est absent en input et son apparition en output viole DEP. Il n'est cependant pas possible qu'un élément soit présent en input et viole en même temps une contrainte DEP. Pour maintenir cette analyse, il faudrait au minimum un réaménagement important de la Théorie de la Correspondance, dont nous ne voyons pas de motivation en dehors de cette analyse de la liaison ¹³.

Au-delà de ce problème formel, et comme nous l'avons souligné plus haut, l'approche défendue par B. Tranel repose crucialement sur la contrainte antihiatus : l'intuition qui la sous-tend est qu'une CL lexicalement disponible est insérée pour éviter une structure marquée (une séquence de voyelles). Si elle prédit correctement la réalisation de la liaison dans les mots liaisonnants se terminant par une syllabe ouverte, elle n'explique pas pourquoi l'on rencontre des mots liaisonnants dont la forme courte se termine par une consonne. Observons donc ce qui se passe dans la liaison du pluriel : B. Tranel n'en discute pas dans son article, mais il y a tout lieu de penser qu'elle ne diffère pas fondamentalement des consonnes latentes des adjectifs, auquel cas la représentation de *belles* dans *belles années* est /bɛl{z}/. Considérons l'exemple de *belles années* ¹⁴ (5).

13. Un relecteur anonyme nous suggère que l'on pourrait pallier cette difficulté en introduisant, à l'instar de Tranel (1996), une contrainte qui pénalise l'ancrage des segments flottants (contrainte que Tranel appelle, de manière informelle, *Avoid Integrating Floaters*), en lieu et place de la contrainte DEP(L). Les deux analyses ne sont toutefois pas équivalentes et interchangeables, puisque dans Tranel (1996), les consonnes flottantes ne violent pas les contraintes de type PARSE et FILL qui correspondent respectivement, *mutatis mutandis*, aux contraintes de type MAX et DEP de la théorie de la correspondance. Nous renvoyons également à Eychenne (2006 : 226-227) pour une critique formelle de ce type de fidélité à l'ancrage des segments flottants.

14. Nous ignorons le /z/ final de *années* pour simplifier l'exposé.

(5) Liaison dans *belles années* selon une analyse anti-hiatus

/bɛl {z} ane/	*VV	DEP(L)
⊗ bɛlzane		*!
☞ bɛlane		

Dans ce cas, le candidat [bɛlzane], qui devrait être optimal (indiqué par « ⊗ »), est bloqué par [bɛlane]. En effet, un candidat moins fidèle à l'input ne peut être optimal que s'il est moins marqué qu'un candidat plus fidèle ; c'est l'une des propriétés les plus fondamentales d'OT. En l'occurrence, [bɛlzane] viole la contrainte DEP(L) mais n'améliore en rien la marque puisque [bɛlane] et [bɛlzane] respectent tous les deux *VV¹⁵. Inclure une contrainte sur la réalisation du morphème de pluriel /z/, comme nous le faisons nous-même à la section 5, ne résoudrait pas le problème car la grammaire générerait alors des formes comme *[bɛlzanez] où le morphème de pluriel est toujours réalisé, sauf à inclure des contraintes particulièrement *ad hoc* qui ne seraient motivées que par le besoin de sauver l'analyse. On notera, par ailleurs, que l'analyse anti-hiatus ne s'accorde pas avec les données diachroniques puisque, comme nous l'avons rappelé plus haut à la suite de B. Laks (2005), les CL sont nées d'une dynamique d'affaiblissement des codas (PASDECODA en OT) et non d'un phénomène d'épenthèse.

L'analyse que nous développons dans la section 5 permet de pallier les problèmes que nous venons de mentionner. Avant de la détailler, nous présentons le corpus que nous avons constitué pour cette étude.

4. LE CORPUS

4.1. Méthodologie et données

La base PFC peut être considérée comme un échantillon du français parlé dans l'espace francophone. Elle est d'un grand intérêt pour le linguiste dans la mesure où elle permet d'étudier le comportement de la liaison sur un ensemble massif de données attestées, et pas seulement sur des exemples construits par et pour le grammairien.

Le corpus que nous avons constitué est composé de sept enquêtes tirées de la base PFC : nous avons sélectionné différentes aires de la France métropolitaine, aussi bien dans le Nord que dans le Sud. Les points d'enquête qui ont été retenus sont les suivants (chacun est donné avec son code PFC) : Douzens (Languedoc,

15. Notons que ceci est vrai quelle que soit la syllabation de la consonne finale du mot liaisonnant (en coda, en attaque ou ambisyllabique). Le problème demeure si l'on utilise la contrainte ATTAQUE « une syllabe doit avoir une attaque » au lieu de *VV.

11a), Roanne (Loire, 42a), Brécey (Basse-Normandie, 50a), Biarritz (Pays basque, 64a), Lyon (Rhône, 69a), Paris centre (75c), Treize-Vents (Vendée, 85a).

Le corpus, qui était codé pour la liaison ¹⁶, a été étiqueté pour les parties du discours avec l'étiqueteur TreeTagger (Schmid 1994). Cet étiquetage permet d'incorporer dans une certaine mesure la dimension syntaxique, bien qu'elle ne donne aucune information sur les relations syntaxiques entre les mots : idéalement, on souhaiterait disposer d'un étiquetage syntaxique en termes de constituants ou de relations de dépendance, mais en l'absence de telles données, l'étiquetage des parties du discours permet de faire un pas en avant, aussi modeste soit-il.

Les données ont été analysées dans l'environnement statistique R ¹⁷. Les paramètres que nous avons retenus, obtenus à partir des codages liaison et des métadonnées sur les locuteurs, sont la (non-)réalisation de la liaison, la longueur du mot liaisonnant (monosyllabique ou polysyllabique), la catégorie grammaticale du mot liaisonnant, la fréquence du type (en contexte liaisonnant), l'enquête, le sexe du locuteur et sa classe d'âge (moins de 30 ans ; entre 30 et 60 ans ; plus de 60 ans). La fréquence du type a été compressée sur une échelle logarithmique de manière à atténuer les écarts. Notre analyse se base sur un modèle de classification construit à l'aide du module R « ctree » (Hothorn *et al.* 2006) ¹⁸, avec pour variable dépendante la (non-)réalisation de la liaison. L'analyse a été corroborée à l'aide d'un modèle de régression logistique.

Au total, nous avons extrait 10 934 codages de la liaison. Comme dans tout corpus, la distribution des tokens suit une distribution de puissance particulièrement prononcée (cf. Figure 1), c'est-à-dire qu'un très petit nombre de mots rend compte d'un grand nombre d'occurrences et qu'*a contrario*, il existe un grand nombre d'*hapax legomena* (on en dénombre 759, pour 981 types).

16. Voir par exemple Durand & Lyche (2008) pour un exposé des conventions de codage.

17. <http://cran.r-project.org/>

18. Le module « ctree » se situe dans la descendance des arbres de décisions du type CART (*classification and regression tree*) qui effectuent des partitions binaires récursives des données par minimisation de l'entropie. Ce module indique des p-valeurs pour chaque partition, donnant ainsi une indication sur la validité de celle-ci.

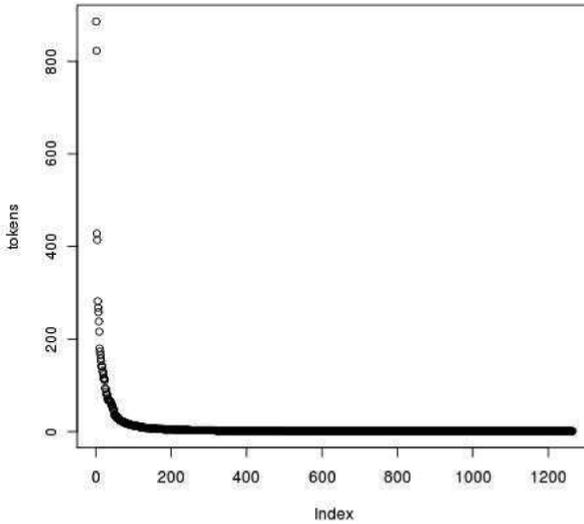


Figure 1 : Fréquence des types de mots situés en contexte de liaison

Les mots les plus souvent codés pour la liaison sont donnés en (6) :

(6) Tokens codés pour la liaison les plus fréquents

on	est	les	en	ils
886	823	428	413	282
était	des	pas	un	dans
268	258	238	216	180

Du point de vue de la répartition des codages, on compte 5 665 liaisons non réalisées (51,8 %), 5 217 liaisons réalisées (47,7 %), 20 liaisons non enchaînées, 10 incertaines et 22 considérées comme épenthétiques. Puisqu'elles représentent 99,5 % du corpus, nous nous sommes concentré sur les liaisons réalisées et non réalisées et avons écarté les autres de l'analyse quantitative. Les catégories les plus codées sont données en (7).

(7) Catégories grammaticales les plus codées

PRO	VER (pres)	ADV	NOM	VER (impf)
1 898	1 736	1 487	1 300	800
DET	CONJ	PREP	VER (inf)	NUM
604	485	420	379	358

Enfin, si l'on ignore les 9 [v] codés comme des consonnes de liaison, les CL se répartissent comme suit : z (2 388) > n (1 799) > t (1 040) > r (11) > p (6) (voir la discussion dans Durand & Lyche (2008) et Durand & Laks, ce numéro).

4.2. Résultats

Dans la mesure où le corpus que nous utilisons est un sous-corpus du corpus analysé dans J. Durand et C. Lyche (2008), J. Durand et B. Laks (ce numéro) et G. Mallet (2008), nous nous contentons d'évoquer les grandes tendances observées et renvoyons à ces auteurs pour une approche plus qualitative.

Le modèle que nous avons construit établit une asymétrie nette ($p < 0,001$) entre, d'une part, les adjectifs, déterminants (y compris les numéraux), les pronoms et les prépositions et, d'autre part, les adverbes, conjonctions, noms et verbes. Comme nous l'avons déjà souligné, cette distinction est très grossière dans la mesure où elle ne fait pas de différence, par exemple, entre les noms singuliers et pluriels, dont on sait qu'ils ont un comportement différent vis-à-vis de la liaison. Elle donne toutefois des informations utiles au niveau macroscopique : ainsi, dans la masse des données, il ressort que les adjectifs et prépositions ont tendance à se présenter plus souvent avec liaison que sans, à l'instar des déterminants et pronoms : ceci est certainement dû au fait que des adjectifs fréquents (*petit, grand*) sont souvent en position prénominale (notamment dans le texte lu, cf. *grand émoi, grand honneur*). Les verbes et noms, au contraire, tendent à se présenter plutôt sans liaison, à l'instar des conjonctions et adverbes¹⁹.

Le deuxième critère qui joue un rôle important dans la (non)-réalisation de la liaison est la taille du mot, les mots monosyllabiques présentant plus souvent une liaison. On sait que les mots monosyllabiques ont tendance à perdre leur autonomie accentuelle (Laks 2005) : ils se présentent donc comme des proclitiques au niveau prosodique, ce qui peut expliquer pourquoi ils montrent des taux de réalisation de la liaison plus élevés. À titre d'exemple, comparons les prépositions *dans* et *après*. Dans le cas de la préposition *dans*, la liaison est réalisée dans 113 cas, mais on ne relève que 4 occurrences de liaison non réalisée²⁰ : *dans* ∩ *un camping* ; *dans* ∩ *un mobil-home* [51ayp1, guidé] ; *dans* ∩ *un sens* [75cac1, guidé] ; *dans* ∩ *un truc chez moi* [75ccm1, guidé]. Dans le cas d'*après*, en revanche, on ne trouve aucune liaison réalisée (0/56). La même observation peut être faite pour les liaisons « variables » à environnement syntaxique constant : elle est réalisée dans *est* dans 45 % des cas (374/831), mais seulement dans 11,2 % des cas (30/269) dans l'imparfait *était*.

Le troisième paramètre important que l'on peut dégager est la fréquence ($p < 0,001$) : elle joue un rôle important dans les monosyllabes, les mots fréquents étant plus sujets à la liaison que les mots rares. Ce fait s'accorde avec les hypothèses des grammaires d'usage selon lesquelles la réalisation de la liaison

19. Au risque de nous répéter, il n'est pas fait de distinction entre les adverbes comme *très*, où la liaison est quasi-catégorique, et les autres (*après, vraiment*, etc.), ce qui n'est évidemment pas idéal.

20. Ces exemples nous montrent donc qu'il n'y a pas, en français parlé, de liaison proprement « catégorique » (et encore moins « obligatoire »), mais uniquement des liaisons qui tendent à être réalisées de manière systématique.

s'explique avant tout par la fréquence des constructions dans lesquelles elle se rencontre (Bybee 2001 ; Laks 2005).

Dans la hiérarchie des paramètres, la tâche a elle aussi une influence dans la réalisation de la liaison. Les trois niveaux de style qui ont été codés (entretien libre *vs* entretien guidé *vs* texte lu) étaient destinés à éliciter trois registres différents. On observe effectivement une différence claire entre les différents registres : pour les mots monosyllabiques de type {Adj, Det, Pro, Prep} peu fréquents, il y a une différence significative entre les trois tâches ($p < 0,001$), la conversation guidée étant plus proche du texte lu que de la conversation libre. Par ailleurs, dans les noms, verbes et adverbes, on observe une dichotomie nette entre texte lu d'une part et conversation (libre et guidée) d'autre part. À titre d'illustration, il est intéressant d'observer le comportement de la forme verbale *est* qui est le type le plus fréquent dans le corpus. Il faut distinguer le verbe/copule *est* du présentatif *c'est*, ce dernier étant plutôt caractéristique de l'oral (il est d'ailleurs absent du texte lu). Dans *est*, la liaison est réalisée à 98 % dans le texte lu (92/94) contre 47 % (63/135) en conversation. Dans le présentatif *c'est*, elle n'est en revanche réalisée qu'à 25 % (83/336) (conversation uniquement). On voit bien ici la gradation entre le degré de formalité et le taux de réalisation de la liaison.

Enfin, il faut relever des groupements intéressants du point de vue de l'âge des locuteurs. Dans les adjectifs et prépositions polysyllabiques, il y a une différence entre les seniors d'une part (60+ ans) et les juniors et moyens d'autre part (- de 60 ans). On observe également une différence significative dans les noms et adverbes ($p < 0.001$ dans tous les cas) : pour les polysyllabes, il existe une différence en conversation (libre et guidée) entre les seniors d'une part et les moyens et juniors d'autre part ; pour les monosyllabes, les juniors (30 ans et moins) se distinguent des moyens et seniors (> 30 ans), toujours en conversation. Dans tous les cas, nous n'observons pas d'influence de l'âge dans la lecture du texte. Ces résultats sont compatibles avec l'hypothèse que la liaison (variable) tend à perdre du terrain chez les locuteurs les plus jeunes. Le présentatif *c'est* l'illustre pleinement : la liaison y est réalisée à 92 % (22/24) chez les seniors contre 41 % (28/68) chez les moyens et seulement 26 % (10/39) chez les juniors.

Nous n'avons pas été en mesure d'établir de différence significative entre les différentes enquêtes. Il est difficile d'en inférer que la variation géographique est effectivement insignifiante : il semble clair qu'il n'y a pas de véritable différence du point de vue des liaisons « catégoriques » et « erratiques », mais on se heurte à un problème de représentativité des données pour la liaison « variable » : il est possible que des données plus massives pour chaque point d'enquête (ou idéalement chaque locuteur) fasse émerger des patrons qui sont ici noyés dans la masse des variables. En tout état de cause, il semble que, s'il y a bel et bien des différences géographiques, celles-ci sont tout à fait mineures par rapport aux autres variables. De la même manière, nous n'avons pas observé de différence du point de vue du sexe des locuteurs.

Ces résultats suggèrent, comme on peut s'y attendre, que les critères linguistiques, à savoir l'environnement syntaxique (avec les réserves que nous avons évoquées), la taille du mot et sa fréquence, sont les principaux facteurs qui conditionnent la réalisation de la liaison. Les paramètres extra-linguistiques que sont le style et l'âge expliquent, quant à eux, une grande partie de la variation restante. D'un point de vue théorique, ils mettent en exergue le caractère intrinsèquement dynamique et variable de la liaison, ce qui milite en faveur d'un cadre théorique tel qu'OT, à même d'appréhender la variation de manière satisfaisante.

5. ESQUISSE D'UNE ANALYSE EN OT

5.1. Latence et exponence morphologique

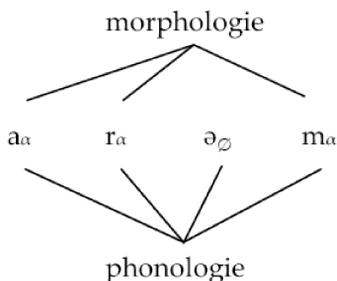
L'analyse que nous proposons s'apparente au concept d'appendice phonologique d'O. Bonami *et al.* (2005). Comme ces auteurs le suggèrent, leur approche offre une synthèse entre les approches par latence et les approches par épenthèse.

Pour le comprendre, il est nécessaire d'introduire un principe important d'OT qui, s'il a été énoncé dès les premiers travaux dans ce cadre, n'a été « redécouvert » que récemment (Oostendorp, 2005 : §8). Il s'agit du principe de cohérence de l'exponence (*Consistency of Exponence*, voir McCarthy & Prince, 1993b : 21) :

- (8) Cohérence de l'exponence : l'exponence morphologique d'un morphème phonologiquement spécifié ne peut être modifiée.

Ce principe stipule que l'affiliation morphologique (ou exponence) d'un élément ne peut être modifiée de l'input à l'output. Par exemple, un schwa épenthétique dans le néerlandais /arm/ > [arəm] 'BRAS' ne sera jamais reconnu comme faisant partie du morphème en output, même s'il est intégré prosodiquement dans celui-ci. Par commodité, on peut représenter l'exponence d'un morphème par un indice, les segments épenthétiques ayant une exponence et un indice nuls :

- (9) Exponence morphologique (modifié d'après Oostendorp 2005)



Ce principe a des répercussions profondes du point de vue de l'interaction phonologie/morphologie (cf. Oostendorp 2005 ; Eychenne 2006). Il permet également d'offrir un traitement intéressant pour les CL. Dans l'esprit d'O. Bonami *et al.*, nous proposons que les CL sont des consonnes présentes dans les formes sous-jacentes, mais qu'elles ont une exponence morphologique spécifique (que nous noterons par le diacritique λ). Comparons à cet égard la représentation de *petite* et *petit* en français non méridional (nous délaissions par ailleurs la question du schwa) :

- (10) Consonne fixe (a) *vs* latente (b)
- a. $p_{\alpha\partial\alpha}t_{\alpha}i_{\alpha}t_{\alpha}$
 - b. $p_{\alpha\partial\alpha}t_{\alpha}i_{\alpha}t_{\lambda}$

En (10a), toute la chaîne de segments a la même exponence lexicale (notée arbitrairement α) : la consonne finale est donc une consonne fixe. En (10b), en revanche, la séquence $\{p,\partial,t,i\}$ a la même exponence α alors que le $\{t\}$ final a une exponence différente (λ). Ceci étant dit, dans les deux cas, les segments sont linéairement ordonnés : le $/t/$ final de *petit* fait donc partie de la représentation lexicale et n'est pas « épenthétique »²¹. Cette approche nous semble supérieure aux approches par flottance en ce sens qu'elle permet de modéliser le caractère exceptionnel des CL sans supposer un double encodage de la linéarité (au niveau du squelette et au niveau de la mélodie, voir Sauzet 1999 ; Eychenne, 2006 : 350). Elle permet, par ailleurs, une modélisation explicite du statut CL qui s'intègre à la théorie de la correspondance.

5.2. Aspects grammaticaux

Ceci étant posé, examinons les mécanismes à l'œuvre dans la réalisation de la liaison. À la suite de B. Laks (2005), nous admettrons que les deux dynamiques à l'œuvre dans la réalisation de la liaison sont la tendance à l'enchaînement généralisé (ce qui est représenté en OT par la contrainte ATTAQUE « toute syllabe doit avoir une attaque ») et la tendance à l'amuissement des codas (PASDECODA « une syllabe n'a pas de coda »).

Si le phénomène de liaison est bien le résultat d'un phénomène de « troncation » des CL, la grammaire doit être en mesure de distinguer les CL des consonnes fixes. Avec J. Pater (2006), nous admettrons l'existence dans la grammaire d'un mécanisme d'indexation de contraintes de marque : une contrainte de marque générique peut se dédoubler en une version générique et une version spécifique. La version spécifique n'émerge dans la grammaire que s'il existe des preuves fournies par un ensemble d'items lexicaux ayant un comportement

21. Notons au passage que dans notre approche la différence entre masculin et féminin est traitée par supplétion : ceci permet d'accorder un statut différent aux CL (*i.e.* leur caractère « latent »), comme nous en avons souligné la nécessité à la section 1.

exceptionnel. Pour qu'elle émerge, la contrainte spécifique doit nécessairement dominer la contrainte générique (spécifique » générique) ; dans le cas contraire, l'effet de la contrainte spécifique ne pourrait jamais être observé et sa postulation dans la grammaire serait complètement opaque. Dans leur formulation originale, les contraintes spécifiques sont co-indexées avec les morphèmes qu'elles ciblent : nous réinterprétons ces indices comme étant l'exponence morphologique d'une unité sous-jacente. Dans le cas de la liaison, cela signifie que l'adulte, dans sa grammaire stabilisée, fait la distinction entre deux classes de consonnes finales : dans la grande majorité des cas, les consonnes finales sont réalisées dans tous les contextes, mais il existe un sous-ensemble d'unités morphologiques pour lesquelles la consonne finale ne peut apparaître en coda. Nous illustrons ce modèle en (11)²² : la version spécifique de PASDECODA (qui ne cible que les consonnes d'exponence λ) domine la contrainte de fidélité MAX(C) qui demande à ce que toute consonne dans l'input ait un correspondant dans l'output. Cette contrainte de fidélité domine à son tour la contrainte générique PASDECODA (qui, elle, cible toutes les consonnes, sans tenir compte de leur exponence), sans quoi toutes les codas seraient systématiquement éliminées. Nous incluons, par ailleurs, dans la grammaire une contrainte de marque (* λ) qui pénalise l'existence de cette structure spécifique en output (l'intuition étant qu'il est plus marqué de maintenir deux classes de consonnes qu'une seule). Dans notre approche, le candidat sans CL ne viole pas de contrainte anti-hiatus, mais viole la contrainte plus générale qui demande à ce que toute syllabe ait une attaque.

(11) *grand émoi* vs. *grand dénivelé* [64aji1, Pays basque]

/gʁãt λ emwa/	ATTAQUE	PASDECODA λ	MAX(C)	PASDECODA	* λ
☞ [gʁã.t λ e.mwa.]					*
[gʁã.e.mwa.]	*!		*		
/gʁãt λ denivəle/					
[gʁãt λ .de.ni.və.le.]		*!		*	*
☞ [gʁã.de.ni.və.le.]			*		

Le modèle que nous proposons s'accorde bien avec la productivité erratique souvent relevée (cf. Côté 2005 et les références citées) des CL, dès lors que l'on reconnaît leur caractère idiosyncrasique et exceptionnel : puisque les CL doivent être apprises au cas par cas pour chaque item, il n'y aura de CL que dans les mots qui présentent des preuves de liaison. Supposons, par exemple, que le corpus que nous avons utilisé puisse être considéré comme un échantillon d'un input moyen auquel un locuteur francophone serait soumis : dans les sept enquêtes retranscrites, nous observons en tout et pour tout 6 occurrences de l'adjectif *idiot*, mais aucune d'elles ne se trouve en position pré-nominale. Dans ce cas, il est

22. Le point dans les formes d'output représente des frontières syllabiques.

tout à fait probable que notre locuteur stocke *idiot* comme /idjo/ sans /t/ latent puisqu'il n'y a aucune preuve de liaison dans la langue ambiante. De même, la locutrice [11amg1] (18 ans, Languedoc) prononce en lecture *grand émoi* et *grand honneur* sans liaison : on peut supposer que pour elle, ce mot est stocké sans CL puisqu'il est presque exclusivement utilisé de manière attributive.

Examinons maintenant le cas du pluriel : le /z/ du pluriel se comporte comme les consonnes latentes et il est donc légitime de lui attribuer la même représentation. Nous le représenterons donc comme un /z_λ/ dans les représentations sous-jacentes. Dans le cas où le morphème de pluriel suit une consonne latente sous-jacente (*grands intellectuels* /gʁãt_λ.z_λ ẽtelektyɛl.z_λ/ [64aj1]), c'est le morphème du pluriel qui est réalisé et non la consonne finale ([gʁã.zẽtelektyɛl] et non *[gʁã.t_λẽtelektyɛl]). Dans le cadre OT, nous formaliserons ce fait par une contrainte demandant à ce que la marque fonctionnelle du pluriel soit réalisée en output ²³ :

(12) MAX(Plur) : un morphème de pluriel en input a un correspondant en output.

Cette contrainte est crucialement dominée par la version spécifique de PAS-DECODA (sans quoi l'on observerait des formes comme *petits amis* *[pœtizamiz]) mais peut, sous certaines conditions (cf. 13), décider du candidat optimal.

(13) Liaison et morphème du pluriel

/gʁãt _λ .z _λ ẽtelektyɛl.z _λ /	ATTAQUE	PASDECODA _λ	MAX(C)	PASDECODA	MAX(PLUR)
[gʁãẽtelektyɛl]	*!		*		**
[gʁãt _λ ẽtelektyɛl]			**		**!
☞ [gʁãz _λ ẽtelektyɛl]			**		*
[gʁãt _λ .z _λ ẽtelektyɛl]		*!	**	*	*

La contrainte MAX(Plur) est donc responsable de la préservation de l'information fonctionnelle en output, dès lors que la préservation de cette information ne viole pas les contraintes de marques super-ordonnées. Notre analyse résout, en outre, le problème auquel fait face l'analyse anti-hiatus (cf. 5 *supra*).

23. Voir Tranel (2000 : 57-8) sur l'expression du genre en output et Rice (2006) sur la fidélité à l'expression des catégories grammaticales.

(14) Réanalyse de la liaison dans *belles années*

/bɛl _λ .anez _λ /	ATTAQUE	PASDECODAL	MAX(C)	PASDECODA	MAX(PLUR)
[bɛl.z _λ a.nez _λ .]		*!		**	
☞ [bɛl.z _λ a.ne.]			*	*	*
[bɛ.la.ne.]			**!		**
[bɛ.l.a.ne.]	*!		**		**

Le tableau 14 montre que le candidat maximalement fidèle [bɛl.z_λa.nez_λ.] est sub-optimal puisqu'il viole la contrainte spécifique PASDECODAL qui domine MAX(C). En d'autres termes, vis-à-vis des CL, la grammaire privilégie les outputs non marqués aux outputs fidèles. La forme [bɛ.l.a.ne.] viole quant à elle la contrainte dominante ATTAQUE responsable de l'enchaînement généralisé en français. Enfin, la forme [bɛ.la.ne.] est éliminée par le candidat optimal puisqu'elle viole à deux reprises la contrainte de fidélité MAX(C) alors que le candidat optimal ne la viole qu'une fois ²⁴.

Nous concluons cette discussion en étudiant les cas de liaison disloquée : M.-H. Côté (2005) souligne à juste titre que l'on ne peut pas traiter ces cas de liaison à distance comme de simples erreurs de performance et que, si les locuteurs peuvent produire ces formes, c'est que leur grammaire les y autorise. OT offre un cadre particulièrement intéressant pour traiter ces phénomènes : on reconnaît, en effet, depuis J. McCarthy et A. Prince (1995) l'existence de contraintes de linéarité qui demandent à ce que l'ordre linéaire des segments de l'input soit préservé en output. Ces contraintes ont, notamment, pour rôle de bloquer les métathèses. Si la contrainte de linéarité vient à être crucialement dominée par la contrainte ATTAQUE et par la contrainte qui bloque l'épenthèse ({ATTAQUE ; DEP(?)} » LINÉARITÉ), on voit alors émerger dans la grammaire ces cas de liaison disloquée :

(15) Liaison disloquée en OT

/dɔnez _λ mwa ā/	ATTAQUE	NOCODAL	DEP(?)	LINÉARITÉ	MAX(C)
[dɔ.nez _λ mwa. ā]	*!	*			
[dɔ.ne.mwa. ā]	*!				*
☞ [dɔ.ne.mwa.z _λ ā]				***	
[dɔ.ne.mwa.ʔā]			*!		

24. La position de Max(Plur) n'est pas déterminée de manière exacte et la hiérarchie, comme c'est souvent le cas, n'est pas strictement transitive. Du point de vue de notre argument, ce qui importe est que la grammaire satisfasse les conditions suivantes : NOCODAL » MAX(C) » NoCODA et NOCODAL » MAX(Plur).

Dans cet exemple, bien que la contrainte de linéarité soit violée trois fois (les segments [m], [w] et [a] séparent [z_λ] de sa position en input) dans [dɔ.ne.mwa.z_λɑ̃], c'est lui qui émerge comme optimal en vertu du principe de domination stricte : une contrainte superordonnée a toujours précédence sur une contrainte subordonnée. Pour rendre compte de l'asymétrie entre consonnes fixes et consonnes latentes, on introduira dans la grammaire une contrainte d'alignement sur le bord droit d'un mot du type :

- (16) ALIGN-D : ALIGN(morph, droite, σ, droite) : le bord droit de toute unité morphologique doit correspondre avec le bord droit d'une syllabe.

où « unité morphologique » désigne une séquence de segments dont l'exponente morphologique est identique. Cette contrainte sera satisfaite lorsqu'une consonne fixe finale est enchaînée (auquel cas elle est ambisyllabique), mais elle sera violée si la consonne est disloquée. Dans le cas des consonnes latentes, en revanche, cette contrainte ne permet pas de départager un candidat avec consonne disloquée d'un candidat avec consonne enchaînée, les deux violant la contrainte (bien que le locus soit différent).

Nous nous sommes concentré ici sur la caractérisation des contraintes nécessaires pour rendre compte de la réalisation de la liaison en OT. Or, comme nous l'avons souligné à la section 4, la liaison est un phénomène éminemment variable, ce qui en OT se traduit par la génération d'outputs différents. Le modèle le plus intéressant à l'heure actuelle est sans doute la théorie de l'optimalité stochastique développée par P. Boersma (cf. Boersma & Hayes 2001). Il s'agit d'un modèle paramétrique où les contraintes sont représentées par des gaussiennes réparties sur une échelle arbitraire, la moyenne de la gaussienne représentant le rang de la contrainte et l'écart-type un bruit qui est ajouté aléatoirement au moment de l'évaluation : selon la valeur des paramètres, le modèle permet d'obtenir des effets de fréquence du type « la contrainte C1 domine C2 dans 80 % des cas ». Ce modèle possède un algorithme d'apprentissage attrayant (le *gradual learning algorithm*, GLA) qui permet d'obtenir de bonnes approximations de fréquences en corpus : il pourrait être appliqué au nôtre pour construire une grammaire probabiliste qui puisse générer les patrons variationnels observés. Cela nécessiterait toutefois un codage des données qui prenne en compte les aspects syntaxiques de manière plus fine, ce qui dépasse le cadre de ce travail.

6. CONCLUSION

Dans cet article, nous avons traité de la liaison en français. Au niveau empirique, nous avons montré que les corpus PFC apportent de nouvelles données qui reflètent les dynamiques synchroniques et diachroniques en cours dans la réalisation de la liaison.

Au niveau théorique, nous avons proposé une approche en théorie de l'optimalité qui, en exploitant le principe de cohérence de l'exponence, offre une synthèse des approches qui ont été proposées jusqu'alors. À l'instar des approches constructionnistes, notre modèle est non représentationnel et reconnaît le caractère idiosyncrasique et émergent de la liaison. Comme dans les approches par épenthèse, il permet de rendre compte du statut spécifique des CL, qui sont distinctes des consonnes fixes, notamment dans le cas des liaisons disloquées. Enfin, notre modèle reste dans la tradition des approches par latence qui reconnaissent l'attachement des CL au mot gauche, sans pour autant que ne se pose le problème de la flottance.

Dans son approche en grammaire de construction, B. Laks soutient que la liaison est le résultat de l'interaction complexe de quatre dynamiques fondamentales :

- (17) Les quatre dynamiques de la liaison (Laks 2005)
- (i) la dynamique d'enchaînement, qui tend à resyllabifier les consonnes finales à droite au sein d'un groupe prosodique ;
 - (ii) la dynamique de simplification des codas, qui s'est manifestée dans l'histoire du français et dont la liaison est un reliquat ;
 - (iii) la dynamique graphique, qui tend à fixer en mémoire l'association entre la consonne de liaison et le mot précédent ;
 - (iv) la dynamique morphosyntaxique, qui tend à protéger les CL qui sont des marqueurs fonctionnels (marqueurs verbaux, pluriel).

Au niveau phonologique, notre approche repose sur les dynamiques (i) et (ii), formalisées par les contraintes de marque ATTAQUE et PASDECODA. Bien que nous n'en ayons pas proposé de modélisation explicite, il est tout à fait possible que la dynamique (iii) soit effectivement responsable de l'ancrage à gauche des CL : de fait, dans les variétés de français où l'influence de la graphie est moins prégnante (notamment en Amérique du Nord, cf. Côté 2005), les CL semblent être plus autonomes du mot 1 et l'on observe des phénomènes tels que la généralisation de [t] comme CL (par exemple, avec l'adjectif *gros*). Enfin, nous avons montré que la dynamique (iv) était effectivement à l'œuvre dans la réalisation du morphème du pluriel : celle-ci est régie par la satisfaction d'une contrainte de fidélité qui demande à ce qu'un morphème de pluriel en input soit réalisé en output. OT fournit donc un cadre formel robuste et explicite pour modéliser la liaison, qui n'est que le résultat émergent d'un ensemble de dynamiques contradictoires et de leur hiérarchisation singulière en français.

Références bibliographiques

- ÅGREN J. (1973), *Enquête sur quelques liaisons facultatives dans le français de conversation radiophonique*, Uppsala : Acta Universitatis Upsaliensis.
- ARMSTRONG N. (2001), *Social and Stylistic Variation in Spoken French: a comparative approach*, Amsterdam: John Benjamins.
- BLOOMFIELD L. (1933), *Language*, Londres: Allen and Unwin.

- BOERSMA P. & HAYES B. (2001), "Empirical tests of the Gradual Learning Algorithm", *Linguistic Inquiry* 32, 45-86.
- BONAMI O., BOYÉ G. & TSENG J. (2004), "An integrated approach to French liaison", in G. Jäger et al. (eds.), *Proc. of Formal Grammar 2004*, Nancy, 29-45.
- BONAMI O., BOYÉ G. & TSENG J. (2005), « Sur la grammaire des consonnes latentes », *Langages* 158, 89-100.
- BYBEE J. (2001), *Phonology and Language Use*, Cambridge: Cambridge University Press.
- CLEMENTS N. & KEYSER S. (1983), *CV Phonology*, Cambridge (MA): MIT Press.
- CÔTÉ M.-H. (2005), « Le statut lexical des consonnes de liaison », *Langages* 158, 66-78.
- CHEVROT J.-P., CHABANAL D. & DUGUA C. (2007), « Pour un modèle de l'acquisition des liaisons basé sur l'usage : trois études de cas », *Journal of French Language Studies* 17, 103-129.
- CHEVROT J.-P., FAYOL M. & LAKS B. (éds) (2005), *Langages n°158 : Nouvelles approches de la liaison : phonologie et cognition*, Paris : Larousse/Armand Colin.
- CHOMSKY N. & HALLE M. (1968), *The Sound Pattern of English*, New York: Harper and Row.
- CROFT W. & CRUSE D. A. (2004), *Cognitive linguistics*, Cambridge: Cambridge University Press.
- DELATTRE P. (1966), *Studies in French and Comparative Phonetics*, La Haye: Mouton & Co.
- DELL F. (1985), *Les règles et les sons : introduction à la phonologie générative*, Paris : Hermann.
- DE JONG D. (1994), « La sociophonologie de la liaison orléanaise », in C. Lyche (ed.), *French Generative Phonology: Restrospective and Perspectives*, Salford: AFLS/ESRI, 95-130.
- DURAND J. (1986), "French liaison, floating segments and other matters in a dependency framework", in J. Durand (ed.), *Dependency and Non-Linear Phonology*, Londres: Croom Helm, 161-201.
- DURAND J. (1990), *Generative and Non-Linear Phonology*, Londres/New-York: Longman.
- DURAND J. & LYCHE C. (2008), "French liaison in the light of corpus data", *Journal of French Language Studies* 18 (1), 33-66.
- DURAND M. (1936), *Le genre grammatical en français parlé à Paris et dans la région parisienne*, Paris : d'Artrey.
- ENCREVÉ P. (1988), *La liaison avec et sans enchaînement. Phonologie tridimensionnelle et usages du français*, Paris : Éditions du Seuil.
- EYCHENNE J. (2006), *Aspects de la phonologie du schwa dans le français contemporain. Optimalité, visibilité prosodique, gradience*, Thèse de doctorat de l'Université de Toulouse II- Le Mirail.
- FOUCHÉ P. (1956), *Traité de prononciation française*, Paris : Klincksieck.
- HYMAN L. (1985), *A Theory of Phonological Weight*, Dordrecht: Foris.
- HOTHORN T., HORNIK K. & ZEILEIS A. (2006), "Unbiased Recursive Partitioning: A Conditional Inference Framework", *Journal of Computational and Graphical Statistics* 15 (3), 651-674.
- KLAUSENBURGER J. (1984), *French liaison and linguistic theory*, Wiesbaden: F. Steiner.
- LAKS B. (2005), « Phonologie et construction syntaxique : un test de figement et de cohésion syntaxique », *Linx* 53, 155-171.
- LANGACKER R. (2008), *Cognitive Grammar. A Basic Introduction*, Oxford: Oxford University Press.
- MALLET G.-M. (2008), *La liaison en français : descriptions et analyses dans le corpus PFC*, Thèse de doctorat de l'Université Paris Ouest Nanterre La Défense.
- MCCARTHY J. & PRINCE A. (1993a), "Generalized Alignment", *Yearbook of Morphology*, Dordrecht: Kluwer, 79-153.

- McCARTHY J. & PRINCE A. (1993b), *Prosodic Morphology: constraint interaction and satisfaction*, Rapport technique RuCSS-TR-3, New Brunswick: Rutgers University Center for Cognitive Science.
- McCARTHY J. & PRINCE A. (1995), "Faithfulness and reduplicative identity", *University of Massachusetts Occasional Papers in Linguistics*, Amherst: GLSA, 249-384.
- OOSTENDORP M. van (2005), "The Theory of Faithfulness", Manuscrit, Meertens Institute.
- PATER J. (2006), *The locus of exceptionality: morpheme-specific phonology as constraint indexation*, Manuscrit, Amherst.
- PRINCE A. & SMOLENSKY P. (1993), *Optimality Theory: Constraint Interaction in Generative Grammar*, New Brunswick (NJ): Rutgers University Center for Cognitive Science.
- RICE C. (2006), "Nothing is a phonological fact: gaps and repairs at the phonology/morphology interface", *Proceedings of the North East Linguistic Society* 36, 27-38.
- SAUZET P. (1999), « Linéarité et consonnes latentes », *Recherches Linguistiques de Vincennes* 28, 59-86.
- SAUZET P. (2004), « La singularité phonologique du français », *Langue française* 141, 14-35.
- SCHANE S. A. (1968), *French Phonology and Morphology*, Cambridge (MA): MIT Press.
- SCHMID H. (1994), "Probabilistic part-of-speech tagging using decision trees", *Proceedings of the International Conference on New Methods in Language Processing*, Manchester, 44-49.
- SELKIRK E. (1980), *The Phrase Phonology of English and French*, New-York/Londres: Garland Publishing.
- SMOLENSKY P. (1988), "Information processing in dynamical systems: foundations of harmony theory", in D. E. Rumelhart, J. L. McClelland & le groupe de recherche PDP (éds), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition – Vol. 1: Foundations*, Londres/Cambridge (MA): MIT Press, 194-281.
- STERIADE D. (1999), "Lexical conservatism in French adjectival liaison", in B. Bullock, M. Authier & L. Reed (eds), *Formal Perspectives in Romance Linguistics*, Amsterdam: John Benjamins, 243-270.
- TRANEL B. (1981), *Concreteness in Generative Phonology. Evidence from French*, Berkeley: University of California Press.
- TRANEL B. (1995a), "Current issues in French phonology: liaison and position theories", in J. Goldsmith (ed.), *The Handbook of Phonological Theory*, Cambridge (MA): Blackwell, 798-816.
- TRANEL B. (1995b), "French final consonants and nonlinear phonology", *Lingua* 95, 131-167.
- TRANEL B. (1996), "French liaison and elision revisited: a unified account within Optimality Theory", in C. Quicoli, M. Saltarelli & M. L. Zubizarreta (eds), *Aspects of Romance Linguistics*, Washington: Georgetown University Press, 433-355.
- TRANEL B. (2000), « Aspects de la phonologie du français et la théorie de l'optimalité », *Langue française* 126, 39-72.