

La liaison : effets de fréquence et constructions¹

1. INTRODUCTION

L'approche de la liaison proposée dans cet article est fondée sur l'intuition selon laquelle la liaison est plus couramment réalisée dans les groupes de mots qui ont une « forte cohésion syntaxique ». La solution que je propose pour définir la cohésion syntaxique est de la concevoir comme le résultat direct de la fréquence de co-occurrence : les mots qui sont souvent employés ensemble semblent liés par une plus forte cohésion.

J'ai soutenu (notamment, Bybee 1998, 2001) que les séquences de mots employées fréquemment étaient représentées dans la mémoire. Ainsi, l'existence d'une consonne de liaison dans une paire de mots mémorisée est due à la représentation de cette paire dans le lexique. J'ajouterai toutefois que de nombreux contextes de liaison ne concernent pas seulement des chaînes de mots mais aussi des constructions généralisées dans lesquelles la consonne de liaison apparaît. Comme cette consonne est tout simplement présente dans la construction, la question de savoir si elle est insérée ou supprimée ne se pose pas.

Les études sur les phénomènes de sandhi, telle la liaison en français, amènent souvent à constater des décalages entre la structure syntaxique et la structure phonologique, puisque le sandhi est parfois réalisé entre des unités qui ne font pas partie du même constituant. Je voudrais suggérer que la phonologie fournit de bons indices sur la nature des unités de stockage et de traitement. De ce fait, l'étude de ces décalages apparents n'améliore pas seulement notre compréhension de la phonologie mais aussi celle de la syntaxe².

1. Cet article a été traduit de l'anglais au français par Émilie Benoit, DESS Traduction Spécialisée, Université d'Orléans.

2. Il est à noter que le traitement de la liaison présenté ici n'est qu'illustratif et absolument pas exhaustif. Il se fonde sur des approches un peu plus complètes que l'on trouvera dans Bybee 2001a et 2001b. Je remercie Yves Charles Morin et Jean-Pierre Chevrot pour leurs commentaires et suggestions.

La théorie phonologique et morphologique adoptée ici est présentée dans Bybee 1985 et 2001b. En ce qui concerne cet article, les aspects pertinents de cette théorie sont les suivants : des mots entiers et des suites de mots sont mémorisés dans le lexique et le critère de mémorisation est la fréquence d'usage, les mots et suites de mots employés fréquemment étant mémorisés globalement (Bybee 1995, 1998). Cette mémorisation n'entraîne pas l'absence de généralisation, puisque les relations entre les items prennent la forme de connexions lexicales fondées sur des similarités phonologiques et sémantiques. À force d'être utilisés, les items fréquents voient leur force lexicale augmenter et les items qui ont une grande force lexicale sont plus facilement accessibles que les autres et résistent mieux au changement analogique.

On discerne deux effets de fréquence distincts impliqués dans la création et le maintien de la liaison et nous ne nous intéresserons ici qu'au second. Premièrement, les contextes morphosyntaxiques et lexicaux dans lesquels une liaison est maintenant établie étaient des contextes très fréquents et il s'agissait donc de séquences susceptibles d'être mémorisées. Deuxièmement, quand une liaison tend à disparaître, on constate qu'elle se maintient dans les contextes qui apparaissent le plus souvent (Ågren, 1973 ; Delattre, 1966), de la même manière que les formes fléchies irrégulières sont maintenues dans les paradigmes très fréquents.

Les contextes où se manifestent des liaisons dans la langue actuelle illustrent bien cette tendance. La liaison est censée être obligatoire entre les déterminants et les noms, les clitiques et le verbe et dans quelques cas où interviennent des suffixes de nombre ou de personne (par exemple *allon[z]-y*) (Morin et Kaye, 1982 et Tranel, 1981 : 233 entre autres). Dans tous ces cas, un morphème grammatical est présent à l'intérieur d'un constituant et nous avons donc affaire à des constructions. En outre, il s'agit toujours de constructions très fréquentes : les déterminants sont obligatoires ainsi que les pronoms clitiques et les marqueurs de personne et de nombre. De leur côté, les contextes de liaison optionnelle font intervenir soit un suffixe de pluriel entre le nom et l'adjectif ou entre l'adjectif et le nom, soit la consonne finale d'un petit nombre d'adjectifs antéposés au nom, soit certaines prépositions, certains adverbes, particules et expressions figées. Dans chacun de ces cas, c'est une catégorie lexicale restreinte qui est concernée.

On peut souligner deux points pertinents à propos de ces formes impliquées dans des contextes de liaison : d'abord, toutes sont très fréquentes, en particulier dans les constructions ou les séquences dans lesquelles la consonne de liaison apparaît ; ensuite, toutes apparaissent dans des constructions grammaticales très particulières. Aucune consonne de liaison n'apparaît indépendamment d'une construction particulière. La position de la consonne de liaison au sein de constructions précises est justement l'objet de la section qui suit.

2. CONSTRUCTIONS GRAMMATICALES ET LIAISON

Le rôle de la morphologie, celui de la syntaxe et du lexique ont été largement reconnus dans les recherches sur la liaison en français. Toutefois, on a surtout débattu de la contribution respective de chacun de ces niveaux et le rôle de la fréquence dans l'apparition et le maintien des consonnes de liaison n'a été que très

peu étudié. Quelques études ont établi qu'il était très difficile de trouver une description syntaxique unifiée des contextes dans lesquels la liaison est possible (voir Selkirk, 1974 ; Kaisse, 1985). En fait, les propositions fondées sur les hypothèses traditionnelles au sujet de la structure en constituants se sont heurtées à une difficulté : la liaison apparaît dans 98,7 % des cas dans la séquence *est [t] un + NOM* mais seulement dans environ 47 % des cas dans la séquence *je suis[z] un + NOM*, qui a manifestement la même structure de constituants. De la même façon, la séquence *est + PARTICIPE PASSÉ* a un taux de liaison extrêmement élevé qui atteint 98,6 %, alors que dans *je suis + PARTICIPE PASSÉ*, la liaison n'est réalisée que dans 57 % des cas (données tirées de Ågren, 1973).

La plupart des autres auteurs (Baxter, 1975 ; Green et Hintze, 1988 ; Klausenburger, 1984 ; Morin et Kaye, 1982 et Tranel, 1981) ont proposé des analyses qui se réfèrent à des facteurs à la fois morphosyntaxiques et lexicaux. Tout comme d'autres alternances qui ont été lexicalisées ou intégrées à la morphologie, ce qui était un changement phonologique unitaire s'est trouvé associé à des contextes grammaticaux particuliers et a perdu cette unité (Bybee, 2001b, chapitre 5).

Cette hypothèse de la lexicalisation ou de l'intégration de la liaison à la morphologie est étayée par le fait – signalé dans différentes études, mais plus explicitement dans Encrevé (1988), Morin et Kaye (1982) et Post (2000), ainsi que dans Ågren (1973) et Green et Hintze (1988) – que les consonnes de liaisons peuvent apparaître aussi bien avant qu'après une pause, avec ou sans resyllabation sur la voyelle de droite. Autrement dit, alors que la liaison dépendait à l'origine de la resyllabation, elle n'est plus contrainte aujourd'hui par les frontières des groupes créés par des pauses ou par celles d'un mot phonologique. En revanche, contrairement à la liaison, l'*enchaînement*, processus toujours d'actualité en français moderne, n'est réalisé qu'à l'intérieur d'un groupe borné par des pauses et n'enjambe donc jamais une pause.

Dans notre approche de la liaison, les unités de base sont les constructions. Puisqu'elles contiennent souvent du matériel lexical et grammatical très spécifique, nous attribuons la présence d'une consonne de liaison à la construction elle-même. Plus précisément, ces constructions sont des séquences récurrentes de morphèmes ou de mots qui sont reliés entre eux par des relations sémantiques, fonctionnelles ou statistiques particulières lorsqu'ils sont utilisés ensemble, relations qu'ils n'entretiennent pas forcément dans d'autres contextes. Les constructions sont plus ou moins conventionnelles, en fonction du degré auquel l'usage répété les a établies dans une langue particulière. Les mécanismes d'établissement des constructions sont (i) l'automatisation par répétition de « tronçons »³ de matériau linguistique (ii) la catégorisation des items qui apparaissent dans certaines positions de ces « tronçons » plus grands. Puisque la fréquence d'emploi est un facteur majeur dans la formation des constructions, il n'est *pas* obligatoire qu'elles aient une signification non-déductible des segments qui les composent : il peut s'agir simplement de « tronçons » de discours employés fréquemment. Toutefois, l'autonomie qui accompagne souvent la répétition et la fréquence d'usage peut aboutir à des significations non-transparentes.

3. Nous traduisons ici *chunk*, terme fréquent dans les théories basées sur l'usage, par l'équivalent français *tronçon*.

À l'une des extrémités du continuum formé par les constructions, on trouve des expressions figées comme *I don't know* et *c'est-à-dire*. Plus près du centre de ce continuum, se situent les constructions comportant du matériel grammatical et qui ont un schéma plus ouvert : par exemple la préposition *dans* suivi d'un SN. À l'autre extrémité du continuum, la plus générale, on trouve des constructions de type [NOM + z + ADJECTIF], qui ont deux emplacements acceptant des items issus de classes ouvertes. Il semble utile de restreindre le terme de « construction » aux séquences dont le schéma comporte au moins un emplacement plus ou moins ouvert et de considérer comme des expressions figées les séquences sans emplacement ouvert, par exemple l'expression *c'est-à-dire*. Les emplacements ouverts des constructions peuvent être catégorisés en termes de traits sémantiques (par exemple : « verbe de mouvement ») ou de traits syntaxiques (par exemple : « pronom »). Comme les constructions émergent de séquences de langage utilisées plus fréquemment, il n'est pas nécessaire que l'organisation des items qu'elles contiennent corresponde aux notions traditionnelles de la structure en constituants, comme nous le verrons en examinant quelques contextes de liaison. Cependant, il est important de garder à l'esprit que les notions traditionnelles de la composition en constituants sont également dérivables à partir des fréquences de co-occurrence, puisque les items qui vont ensemble sur le plan sémantique tendent à apparaître ensemble dans le discours (Bybee 2002).

Dans la littérature sur le phénomène de la liaison, une construction est plus particulièrement l'objet de controverses : nom pluriel suivi d'un adjectif à initiale vocalique. Dans certains cas, entre le nom et l'adjectif, apparaît un [z], une trace de la marque du pluriel qui a été supprimée avant une consonne.

- (1) (a) *enfants [z] intelligents*
(b) *des découvertes [z] inquiétantes*

Ågren (1973) signale que la liaison dans les contextes de ce type est considérée comme obligatoire dans quelques séquences fréquentes, dont certaines sont des noms propres : *affaires [z] américaines*, *Champs [z] Élysées*, *Nations [z] Unies*, *Jeux [z] Olympiques*, etc. Dans les contextes où elle est possible, la liaison n'apparaît que dans 26 % des cas où un nom pluriel est suivi d'un adjectif. La plupart des auteurs considèrent cependant cette construction avec liaison comme étant plutôt productive, à cause de l'existence d'exemples tels que ceux que présentent entre autres Morin et Kaye (1982), dans lesquels la liaison [z] du pluriel est réalisée, mais à une certaine distance de son emplacement étymologique :

- (2) (a) *des chefs d'État [z] africains*
(b) *les chemins de fer [z] anglais*

Ainsi, Morin et Kaye (1982) soutiennent que la liaison du pluriel n'apparaît pas seulement dans des expressions lexicalisées, mais s'applique également de façon productive dans ces cas.

Les données disponibles suggèrent l'existence de deux constructions pour les expressions nom-adjectif au pluriel. La plus générale contient un déterminant pluriel suivi d'un nom et d'un adjectif :

- (3)

[<i>ces</i>	NOM	ADJECTIF]
	<i>des</i>			
	<i>etc.</i>			

 pluriel

Une deuxième construction, moins courante, s'applique seulement aux adjectifs à initiale vocalique :

- (4)

les			
[ces	NOM	-z- [voyelle] - ADJECTIF]	
des			pluriel
etc.			

Mis à part le fait que la deuxième construction (4) est limitée aux adjectifs à initiale vocalique, il y a une autre différence entre les deux : la première construction (3) s'applique à davantage d'items, autrement dit, sa fréquence de type est plus élevée, ce qui la rend plus productive que la seconde. Il n'est donc pas étonnant qu'il y ait des variations dans les données : elles viennent du fait que les locuteurs choisissent le schéma le plus général dans la plupart des cas, même pour les adjectifs à initiale vocalique. Toutefois, le schéma plus spécifique avec le [z] devant les adjectifs à initiale vocalique reste disponible et il est parfois utilisé. Ainsi, la disparition de la liaison ressemble au processus de régularisation des verbes irréguliers : si l'on ne peut pas accéder facilement au schéma spécifique, alors c'est le schéma plus général, plus fort et plus facile d'accès qui est utilisé.

En plus des schémas de construction présentés en (3) et (4), il peut en exister de plus spécifiques pour les adjectifs qui sont couramment utilisés avec la liaison en [z], comme *anglais* ou *américain* :

- (5)

les			
[ces	NOM	-z- anglais]	
des			pluriel

Les exemples en (2) proviennent du fait que les noms composés *chefs d'État* et *chemin de fer* apparaissent dans la position NOM des constructions (4) ou (5).

Une autre construction plurielle donnant lieu à des « fausses liaisons » est composée d'un nombre cardinal suivi de [z] et d'un nom. Comme le mentionne Tranel (1981 : 214-16), la liaison est fréquemment réalisée avec les chiffres cardinaux *deux* [døz] et *trois* [trwaz] avant voyelle quand ces derniers font partie de la même construction que le nom suivant et quand ils indiquent la notion de pluralité. Ainsi, dans *les deux [z] amis* la liaison est réalisée, mais pas dans *le deux octobre*. L'emploi souvent signalé du nombre cardinal sans [z] final étymologique dans cette construction atteste d'ailleurs de la capacité du [z] à fonctionner comme marqueur du pluriel. Tranel (1981 : 216) donne les exemples suivants :

- (6)

(a) quatre enfants	[katzāfā]
(b) huit épreuves	[ɥizeprœv]
(c) neuf œufs	[nœfzø]
(d) vingt-cinq années	[vẽtsẽkzanel]
(e) trois mille évêques	[trwamilzevek]

Ces exemples montrent non seulement l'intrusion d'un [z] non étymologique, mais également, en ce qui concerne les exemples (c) à (e), qu'une autre consonne est présente devant ce [z] à la finale du numéral. Ainsi cette construction va au-delà de la fonction de créer une structure syllabique optimale et témoigne d'un usage véritablement morphologique où le [z] signale la pluralité.

Sur la base d'exemples du type de ceux présentés en (6), Tranel (1981), Morin et Kaye (1982) et Klausenburger (1984) préconisent une analyse dans laquelle la

consonne de liaison est introduite dans certains contextes devant voyelle, et non supprimée avant une consonne. La solution proposée ici ne prend pas position sur l'introduction ou la suppression de la consonne de liaison. Elle affirme simplement qu'il existe une construction qui contient le [z] après un numéral et avant un nom à initiale vocalique. Cette construction pourrait être formulée comme en (7) :

- (7)

[NOMBRE	-z-	[voyelle]- NOM]	
			pluriel

La construction (7) est une généralisation de l'usage normatif selon lequel *deux*, *trois*, *six*, *dix* et *quatre-vingts* sont les seuls nombres qui donnent lieu à la production d'un [z] avant les noms à initiale vocalique.

Comme nous l'avons mentionné plus haut, Ågren (1973), Morin et Kaye (1982) et Green et Hintze (1988) signalent que la consonne de liaison peut apparaître même si une pause ou une hésitation sont produites entre les deux mots de la construction. Green et Hintze (1988 : 159) en produisent un exemple avec le nombre *quatre*, issu de leurs données :

- (8) *quatre euh...[z] obligations*

De tels exemples montrent qu'il est possible de faire une pause ou d'hésiter au milieu d'une construction, tout comme il est possible de faire une pause au milieu d'un mot. Puisque les mots d'une construction sont habituellement associés à d'autres occurrences du même mot, ils sont reconnus comme étant des mots et l'emplacement entre deux d'entre eux est un endroit possible pour faire une pause. La position de la pause dans cet exemple et dans d'autres tirés de ces travaux suggère que la consonne de liaison est davantage associée avec le second mot, dont elle commence la première syllabe, qu'avec le premier, qui en est la source historique.

Une autre illustration du lien étroit qui existe entre des consonnes de liaison précises et des constructions particulières est le cas de la construction réciproque dont il est question dans Morin et Kaye (1982). Ces auteurs signalent que la liaison est facultative après *l'un*, mais seulement lorsque cet item est employé dans la construction correspondant à la réciprocité ; dans tous les autres emplois, la liaison est impossible. Ainsi, en (9), la liaison en [n] est possible, mais pas en (10) :

- (9) (a) *Il les a confondus l'un [n] avec l'autre.*
 (b) *Ils se ressemblent l'un [n] à l'autre.*
 (10) *Ils sont venus, l'un avec sa mère, l'autre avec son père.*

Cet exemple est particulièrement apte à montrer que les morphèmes grammaticaux sont très ancrés dans les constructions où ils apparaissent, et pas seulement en français. Un morphème grammatical est identifié comme tel à cause de sa présence dans certaines constructions grammaticales bien définies. L'histoire des morphèmes grammaticaux montre d'ailleurs que s'ils apparaissent dans des constructions différentes, ils divergent les uns des autres au niveau de la forme phonologique, du sens et des propriétés distributionnelles (Heine et Reh, 1984 ; Hopper, 1991 ; pour l'étude des débuts d'un processus de ce type concernant l'unité *that* de l'anglais, voir aussi Berkenfield, 2001). En français, il y a plusieurs morphèmes grammaticaux dérivant de *un*, qui à l'origine n'était que le numéral : *l'un* comme dans l'exemple (8), où la liaison est impossible ; l'item exprimant la réciprocité,

comme dans l'exemple (7), où la liaison est facultative ; et l'article indéfini, *un, une* pour lequel la liaison est considérée comme obligatoire.

3. LA DISPARITION DE LA LIAISON COMME RÉGULARISATION

À partir de cette analyse dans le cadre des constructions, il devient possible de prédire les endroits où la liaison est le plus susceptible de disparaître. Mon hypothèse est que la consonne de liaison du français, même si elle apparaît entre ce qu'il est convenu d'appeler des mots et non à l'intérieur même de ces mots, est très proche des alternances conditionnées morphologiquement et lexicalement qui ont lieu à l'intérieur des mots. C'est une motivation phonétique qui est à l'origine de la liaison, mais les alternances se sont peu à peu associées avec certains contextes morphosyntaxiques et lexicaux. Comme d'autres alternances morpho-lexicales, elle est sujette à l'extension à de nouveaux contextes dans les cas où elle est productive, mais aussi à la stabilisation ou à la disparition. Les facteurs de fréquence affectant ces changements en cours sont les mêmes que dans les cas d'alternances situées à l'intérieur des mots. Si une construction a une fréquence de type élevée, sa productivité augmente. Les alternances non productives se stabilisent ou disparaissent progressivement, les formes peu fréquentes d'abord, puis celles qui ont une fréquence d'occurrence élevée et qui résistent le plus longtemps (Bybee 1985).

Ce qui rend le cas de la liaison particulièrement intéressant est le fait que les unités dans lesquelles cette alternance apparaît sont plus grandes que ce qu'il est convenu d'appeler des mots. Si des alternances arbitraires se créent et se maintiennent dans de telles unités, alors celles-ci doivent être des unités de stockage, absolument comme le sont les mots. C'est pourquoi les études sur la liaison en français ainsi que sur d'autres cas de sandhi externe sont précieuses : elles apportent la preuve de l'existence d'unités de stockage qui vont au-delà du mot traditionnel. Les faits présentés ici suggèrent fortement que les séquences stables fréquentes sont des unités de stockage et de traitement, comme le sont les constructions contenant des morphèmes grammaticaux. Parmi ces dernières, les constructions spécifiques et les constructions générales sont impliquées dans une compétition dont l'issue est la disparition progressive de la construction la plus spécifique, qui, dans le cas qui nous intéresse, est celle où la liaison est réalisée.

J'ai déjà mentionné que les morphèmes grammaticaux étaient ancrés dans des constructions ; pour décrire cette situation, j'ai émis l'hypothèse que les constructions grammaticales contenaient ces morphèmes grammaticaux sous la forme d'un matériau phonologique explicite. Plus une construction est fréquemment employée, plus sa forme aura de chances d'être maintenue plutôt que d'être remplacée par une construction plus productive (Bybee et Thompson, 1997). Il n'est pas étonnant, alors, que certains contextes de liaison, en particulier ceux qui contiennent des articles suivis de noms ou des pronoms clitiques suivis de verbes, donnent lieu à une liaison obligatoire et ne tendent pas vers la perte de la consonne de liaison. Ces constructions sont celles dont on considère généralement qu'elles manifestent la cohésion syntaxique la plus forte, une cohésion qui pourrait être attribuée à la fréquence de co-occurrence. Nous ne disposons d'aucunes données sur les fréquences relatives de ces deux contextes qui nous permettraient

de prouver cette affirmation. Mais étant donné que presque tous les SN comportent un article défini ou indéfini et que, dans le langage parlé, les pronoms clitiques sujets et objets sont employés de façon redondante (Harris, 1988 : 231-232, 235-236), la grande fréquence de ce type de construction ne peut pas vraiment être mise en doute.

D'autres raisons nous permettent de considérer les constructions [ARTICLE + NOM] et [CLITIQUE + VERBE] comme des unités de stockage et de traitement en français moderne. Le maintien des distinctions de genre, dont l'explicitation passe principalement par l'article singulier, suggère que l'article, qu'il soit singulier ou pluriel, est mémorisé avec le nom. Le traitement particulier des mots commençant par un *h aspiré* (mots devant lesquels il n'y a pas de liaison alors qu'ils sont à initiale vocalique) indique également une représentation lexicale des articles avec leurs noms. Des études sur les usages du français parlé démontrent que les pronoms clitiques sujets et objets accompagnent presque obligatoirement le verbe et qu'ils se comportent sans doute davantage comme des préfixes que comme des clitiques, ce qui donne à nouveau une indication sur leur statut lexical (Harris 1988 : 232).

Si la fréquence de co-occurrence est le principal facteur qui régit l'apparition de la consonne de liaison, nous ne devrions pas trouver ce type de consonnes entre deux items lexicaux choisis au hasard et dont la probabilité de co-occurrence est très faible. Et en effet, le fait est qu'on n'en trouve pas. Un exemple semblerait contredire cette thèse : la petite classe des adjectifs prénominaux qui donnent lieu à une liaison avec les noms à initiale vocalique qu'ils précèdent.

- (11) (a) *un petit [t] écureuil*
 (b) *un gros [z] amiral*
 (c) *un long [g] été*

Un point important est que les constructions illustrées par ces exemples ne concernent qu'un petit nombre d'adjectifs. Beaucoup parmi ceux-ci ont des sens différents selon qu'ils sont antéposés ou placés dans la position postnominale, la plus courante, ce qui suggère l'existence de constructions spécifiques. Pourtant, la fréquence de ce type d'adjectifs et la « cohésion syntaxique » de l'adjectif et du nom qui en résulte doit être moins grande que pour certains morphèmes pleinement grammaticaux qui manifestent une liaison. On peut donc prédire que la liaison dans ce contexte sera moins maintenue que dans d'autres contextes.

Il est possible de vérifier plus concrètement ces hypothèses sur la fréquence à partir des données de Ågren (1973). Grâce à ce matériau empirique, on peut comparer le maintien de la liaison dans plusieurs formes fléchies du même mot caractérisées par des fréquences différentes ainsi que dans plusieurs occurrences du même mot intégrées à des constructions plus ou moins fréquentes. Intéressons-nous d'abord aux formes du verbe copule *être*. Ågren indique que la présence de la liaison dépend directement de la fréquence d'occurrence de ces formes. Il présente des données (tableau 1) qui explicitent le nombre de cas de liaison (L) et de non-liaison (NL) et le nombre d'emploi de chaque item (L+NL). Les formes du verbe *être* précédant une voyelle sont classées selon la proportion des cas de liaison. La correspondance entre la fréquence d'occurrence et la proportion de liaison est assez frappante.

	L	NL	Total	Liaison
est	2 591	77	2 668	97 %
sont	242	38	280	86 %
étant	22	7	29	76 %
était	272	95	367	75 %
êtes	24	10	34	71 %
étaient	36	21	57	63 %
sommes	43	31	74	58 %
suis	65	74	139	47 %
serait	17	24	41	41,4 %
soit	22	32	54	40,7 %
j'étais	6	23	29	21 %

Tableau 1. Nombre de cas de liaison pour les formes du verbe *être*

Les très grandes variations dans le comportement de la liaison dans ces formes sont particulièrement intéressantes. En effet, toutes sont des formes fléchies du même verbe et pourtant elles se comportent assez différemment en ce qui concerne les liaisons⁴. Leur usage se régularise et ce sont les formes les moins fréquentes qui sont les plus susceptibles de subir la régularisation. Ce mécanisme est parallèle à celui par lequel des formes fléchies irrégulières comme *weep/wept* (pleurer) sont régularisées. Ce verbe étant peu fréquent, son passé irrégulier peut ne pas être aussi accessible que celui d'un verbe très fréquent. Ainsi, une nouvelle forme du passé peut être formée par l'utilisation de sa forme de base et de la construction régulière du prétérit anglais.

Toutes les formes du verbe *être* listées dans le tableau 1 ont deux variantes : avec ou sans consonne finale. La variante sans consonne finale est celle qui apparaît le plus souvent. Ainsi, lors de tout emploi particulier de ces formes, deux constructions sont en compétition : une construction plus spécifique, un mot suivi d'un autre mot à initiale vocalique, et une construction plus générale, un mot suivi d'un autre mot commençant par une consonne. Cette dernière construction s'applique plus souvent puisque les mots commençant par une consonne sont plus courants que ceux commençant par une voyelle (ils sont au moins deux fois plus nombreux). La construction spécifique peut être préservée si sa force lexicale est suffisante du fait d'une fréquence assez élevée, mais la possibilité demeure d'employer la construction la plus générale, celle qui est sans liaison.

La fréquence d'occurrence du premier élément d'une séquence de deux mots ne permet pas à elle seule de prédire le comportement de la liaison entre ces deux mots, ce qui est attendu ; la variable la plus importante est plutôt la fréquence avec laquelle les deux éléments apparaissent ensemble, et peut-être la probabilité transitionnelle entre le premier et le deuxième élément. En utilisant un grand corpus de français parlé et lu, Fougeron *et al.* (2001) montrent que la fréquence du premier mot est en forte corrélation avec la présence de la liaison, tout comme la fréquence de co-occurrence des deux mots.

4. Ågren a trouvé les mêmes caractéristiques dans les formes fléchies de l'auxiliaire *avoir* et des semi-auxiliaires (comme il les appelle) *aller*, *falloir*, *pouvoir*, *devoir* et *vouloir*.

Cette hypothèse sur la co-occurrence peut elle aussi être vérifiée plus spécifiquement en utilisant un sous-ensemble des données de Ågren. Ce dernier a calculé la fréquence des liaisons suivant les auxiliaires *aller*, *falloir*, *pouvoir*, *devoir* et *vouloir*. Si on compare ces auxiliaires les uns aux autres, on constate que leur fréquence de co-occurrence ne correspond pas exactement à la proportion de liaisons. Une part de cette absence de correspondance résulte de la fréquence très élevée de certaines formes, comme *je voudrais*, qui ne donnent pas lieu à une liaison parce que (i) les formes verbales sont moins souvent le lieu de liaisons en /z/, et (ii) ces formes sont devenues invariables, à la façon d'expressions figées. La source la plus importante des différences entre les auxiliaires est en fait la variation de leur fréquence d'occurrence dans des constructions spécifiques où ils sont insérés. Tous ces auxiliaires précèdent un infinitif, le plus souvent *être*, qui comptabilise 271 occurrences sur 722 cas d'infinitifs suivant un auxiliaire. Le deuxième infinitif le plus fréquent est *avoir* qui apparaît 93 fois. On doit donc s'attendre à trouver les plus grands taux de liaison avec le verbe *être* (bien plus fréquent que *avoir*) ce qui est en effet le cas, comme on peut le constater dans le tableau 2⁵. Un effet de fréquence est clairement mis en évidence ici puisque *être* est à la fois l'infinitif le plus courant qui suit ces auxiliaires et le contexte où la liaison est la plus courante. Ces résultats sont particulièrement clairs avec *devoir* et *pouvoir* qui sont très fréquemment employés avec l'infinitif *être*.

Un autre fait intéressant dans le tableau 2 est la différence entre les cas où *être* est l'infinitif qui suit l'auxiliaire et le regroupement de tous les autres infinitifs lexicaux (dernière colonne). Alors que c'est cette dernière catégorie qui regroupe le plus d'occurrences, ni la fréquence du type ni celle de la construction dans son ensemble ne semblent être des variables favorisant la présence de la liaison. En effet, puisque le maintien de la liaison est comparable au maintien de l'irrégularité dans les formes fléchies, c'est la fréquence d'occurrence d'une séquence précise qui est à l'œuvre dans la résistance à la régularisation.

	<i>avoir</i>		<i>être</i>		<i>autres infinitifs</i>	
	L	NL	L	NL	L	NL
<i>aller</i>	3	3	7	8	8	16
<i>devoir</i>	9	7	86	5	19	16
<i>falloir</i>	2	3	11	5	42	23
<i>pouvoir</i>	27	33	116	27	91	79
<i>vouloir</i>	3	3	5	1	14	53
	44	49	225	46	171	187
% de liaison		47 %		83 %		48 %

Tableau 2. Résultats d'Ågren pour les semi-auxiliaires et les infinitifs qui les suivent

5. Le Tableau 2 est basé sur les chiffres résumés page 88 d'Ågren (1973), sauf pour les chiffres de *peut* donnés page 85, qui ont été ajoutés, et pour les occurrences des auxiliaires avant *être* + *PP* et *être* + *autre mot*, qui ont été regroupées.

4. LA COHÉSION SYNTAXIQUE COMME FRÉQUENCE DE CO-OCCURRENCE

Ces données concernant la liaison en français corroborent la thèse selon laquelle ce qui a été appelé « cohésion syntaxique » correspond en fait à la fréquence de co-occurrence, un facteur qui détermine la force de l'association entre le premier et le second éléments d'une séquence. Cette connexion est mémorisée et renforcée par un usage fréquent. Ce qui prouve cette mémorisation, c'est que les principes établis pour les alternances morpho-lexicales peuvent également être appliqués à ce qui a été considéré comme étant un plus haut niveau d'organisation. En particulier, plus la fréquence d'une séquence de mots ou d'une construction est élevée, plus il est probable que la liaison soit préservée ; plus cette fréquence est basse, plus la liaison est susceptible de disparaître au profit de l'application d'une construction plus générale.

Les collocations de mots utilisées fréquemment ont des représentations très fortes dans la mémoire. De la même façon que les mots morphologiquement complexes très fréquents sont plus autonomes par rapport à leurs propres paradigmes et par rapport aux paradigmes d'autres mots, les séquences très fréquentes se développent de façon plus autonome. Autrement dit, les connexions entre les mots et morphèmes de ce type de séquences et les autres occurrences des mêmes mots ou morphèmes dans d'autres constructions deviennent plus faibles. La probabilité que ces connexions disparaissent s'accroît du fait de changements fonctionnel ou phonétique et sémantique. L'aboutissement de ce processus est observé dans la grammaticalisation, où des parties de constructions grammaticalisées ne sont plus associées avec leurs sources lexicales (voir la difficulté des apprentis lecteurs anglophones débutants qui doivent identifier ce qu'ils connaissent sous la forme *gonna* avec les trois morphèmes *go*, *ing* et *to*). Dans la plupart des contextes de liaison, la disparition de la structure interne n'est pas si radicale, mais la « cohésion syntaxique » dont il est ici question fonctionne selon le même mécanisme : les séquences fréquentes sont traitées globalement et cette unité de traitement brise leur relation avec d'autres items.

5. PRENDRE LA PHONOLOGIE AU SÉRIEUX

C'est une pratique courante que d'essayer de prédire le comportement de la liaison en français, celui de la contraction *don't* en anglais et d'autres cas de mots qui varient selon la structure et les relations syntaxiques. Cependant, toutes ces tentatives laissent de côté certains cas et beaucoup des analyses introduisent les cas spécifiques de certains items lexicaux ou de certains morphèmes grammaticaux. Beaucoup de chercheurs en ont donc conclu que la relation entre la structure syntaxique et l'application des règles phonologiques n'était pas directe (notamment, Vogel et Kenesei, 1990). Il est intéressant de constater que rares sont ceux qui estiment que la relation entre syntaxe et phonologie est absolument directe mais que nous n'opérons pas avec la bonne syntaxe. Autrement dit, on soutient rarement l'idée que les faits phonologiques puissent nous informer sur l'existence de structures syntaxiques différentes, sauf à un niveau très superficiel. Au contraire, lorsqu'il s'agit de déterminer la structure morphologique, les preuves

fournies par la phonologie sont souvent prises en compte. Pour décider si un morphème grammatical est ou n'est pas un affixe, on s'intéresse à la fusion phonologique entre l'affixe supposé et la base. Qu'arriverait-il si on laissait la phonologie mener la danse de la syntaxe ?

Dans de nombreux cas, évidemment, ce nouveau point de vue ne changerait rien. En français, comme nous le montre le phénomène de la liaison, les déterminants vont avec les noms et les pronoms clitiques avec les verbes. La seule innovation que j'ai proposée est d'affirmer que les séquences [DETERMINANT + NOM] et [PRONOM CLITIQUE + VERBE] sont stockées dans la mémoire. De même, les séquences [ADJECTIF + NOM] et [NOM + PLURIEL + ADJECTIF] sont des constituants d'un point de vue syntaxique (Bybee 2002). Toutefois, des séquences de l'anglais de type [PRONOM + AUXILIAIRE], comme *I'll* et *I'm*, ne sont pas habituellement considérées comme des constituants. En effet, la coupure syntaxique située au plus haut niveau dans une proposition, celle qui existe entre le SN sujet et le SV, se trouve justement à l'intérieur de telles séquences. Il est pourtant indéniable que les auxiliaires forment des contractions et fusionnent avec les pronoms sujets, et non avec le verbe qui suit, même si auxiliaire et verbe font partie du même constituant. La raison que je propose pour expliquer ce phénomène (Bybee et Scheibman, 1999, Bybee, 2002, voir aussi Krug, 1998) est que des instances spécifiques de type [PRONOM + AUXILIAIRE] sont extrêmement fréquentes, beaucoup plus fréquentes que n'importe quelle séquence de type [AUXILIAIRE + VERBE]. Les constats que l'on peut faire au niveau de la phonologie et des faits d'usage conduisent à une analyse de l'anglais proche de celle du Quileute, où les formes des pronoms sujets sont déterminées par des fonctions modales (Andrade, 1933 : 203 et sv.).

Un autre cas intéressant relatif à la liaison concerne le verbe copule à la troisième personne du singulier, *est*. Dans les données d'Ågren, 47 % des emplois de *est* apparaissent dans la construction [*est* + *un* + nom]. Dans cette séquence, la liaison est réalisée dans 98,7 % des occurrences, bien plus que dans tout autre emploi de *est*. Cette tendance suggère fortement qu'il existe une construction dans laquelle *est* [*t*] *un* est un constituant qui précède un nom. On pourrait émettre une thèse comparable en anglais en affirmant que *is a* est un constituant du fait de la fréquence avec laquelle ces deux items sont utilisés conjointement. Bien sûr, dire que *est un* ou *is a* sont des constituants ne signifie pas que *un + nom* et *a + nom* ne sont pas également des constituants. Il n'y a aucune raison que deux constructions ne puissent pas se chevaucher, le chevauchement donnant alors lieu à une analyse en constituants ambiguë.

Si l'on considère que l'usage est l'élément qui détermine la structure en constituants et la hiérarchie syntaxique de telle sorte que les items employés fréquemment ensemble forment des constituants, alors la phonologie est un indicateur pertinent de la structure en constituants, puisque c'est la même propriété, le co-emploi fréquent, qui conditionne les alternances phonologiques.

6. CONCLUSION

Dans un modèle où la mémorisation ne concerne pas seulement des mots isolés mais aussi des séquences et des constructions, le lexique et la grammaire ne sont pas strictement séparés ; ils sont intégrés et contraints par les mêmes

principes organisationnels (Langacker 1987, Bybee 1998). Tout segment de discours répété peut-être mémorisé et placé dans des catégories partageant des unités identiques ou similaires. La catégorisation se fait à de multiples niveaux. Des exemplaires des mêmes mots ou séquences sont projetés sur une seule représentation. De la même manière, les occurrences d'une même construction sont reliées à une seule représentation et les items pouvant occuper les positions ouvertes de la construction contribuent à la formation de catégories fondées sur leurs propriétés sémantiques. Ainsi, les occurrences effectives de la construction *X is going to Y* de l'anglais contribuent à la formation des catégories *X* et *Y*. De la même façon, les occurrences de la construction [NOMBRE + z + [voyelle]-NOM] du français créent les catégories NOMBRE et NOM.

Cet article ne fait qu'aborder le problème des interactions entre syntaxe et phonologie, l'objectif principal étant de démontrer que les constructions partagent de nombreuses propriétés avec les mots morphologiquement complexes. Les éléments qui composent les constructions (qu'il s'agisse de phonèmes ou de morphèmes) apparaissent fréquemment ensemble et peuvent subir des réductions et des fusions phonologiques, tout comme le matériau qui compose les mots. Les alternances peuvent ainsi être préservées à l'intérieur des constructions très fréquentes. Elles peuvent être nivelées si une autre construction propose une façon plus régulière de dire la même chose. Les alternances appartenant aux constructions et aux séquences les moins fréquentes sont alors nivelées en premier et seules les alternances les plus implantées se maintiennent. Ainsi, les constructions qui incluent plus d'un mot sont mieux établies et résistent mieux au changement si elles ont une fréquence d'occurrence élevée ; elles sont plus productives si elles ont une fréquence de type élevée. Toutes ces propriétés impliquent que les constructions sont des unités de stockage et de traitement tout comme le sont les mots et les séquences figées.

Références

- ÅGREN, J. (1973). *Étude sur quelques liaisons facultatives dans le français de conversation radiophonique : fréquence et facteurs*. Uppsala : Acta Universitatis Upsaliensis.
- ANDRADE, M.J. (1933). *Quileute*. New York : Columbia University Press.
- BAXTER, A.R.W. (1975). *Some aspects of naturalness in phonological theory with special reference to Old English*. Lincoln College : Master's Thesis.
- BERKENFIELD, C. (2001). The role of frequency in the realization of English *that*. In J.L. Bybee and P. Hopper (eds.) *Frequency and the emergence of linguistic structure*. Amsterdam : John Benjamins, 281-308.
- BYBEE, J. (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam : John Benjamins.
- BYBEE, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* 10, 425-455.
- BYBEE, J. (1998). The emergent lexicon. *CLS 34: The Panels*, 421-435.
- BYBEE, J. (2001a). Frequency effects on French Liaison. In Joan Bybee and Paul Hopper (eds.) *Frequency and the Emergence of Linguistic Structure*. Amsterdam : John Benjamins, 337-359.
- BYBEE, J. (2001b). *Phonology and language use*. Cambridge : Cambridge University Press.

- BYBEE, J. (2002). Sequentiality as the basis of constituent structure. In T. Givon and B. Malle (eds.) *On the rise of language out of pre-language*. Amsterdam : John Benjamins, 107-132.
- BYBEE, J. & SCHEIBMAN, J. (1999). The effect of usage on degree of constituency: the reduction of *don't* in American English. *Linguistics* 37, 575-596.
- BYBEE, J. & THOMPSON, S. (1997) [2000]. Three frequency effects in syntax. *Proceedings of the Twenty-third Annual Meeting of the Berkeley Linguistics Society*, Berkeley : Berkeley Linguistics Society, 65-85.
- DELATTRE, P. (1966). *Studies in French and comparative phonetics*. The Hague : Mouton.
- ENCREVÉ, P. (1988). *La liaison avec et sans enchaînement : Phonologie tridimensionnelle et usage du français*. Paris : Seuil.
- FOUGERON, C., GOLDMAN, J.P. & FRAUENFELDER, U.H. (2001). Liaison and schwa deletion in French: an effect of lexical frequency and competition? *Proceedings of Eurospeech 2001 Scandinavia*. Aalborg : Denmark, 639-642.
- GREEN, J.N. & HINTZE, M.A. (1988). A reconsideration of liaison and enchaînement. *Occasional papers*. University of Essex: Department of Languages and Linguistics, 136-168.
- HARRIS, M. (1988). French. In Martin Harris and Nigel Vincent (eds.) *The Romance Languages*. Oxford : Oxford University Press, 209-245.
- HEINE, B. and REH, M. (1984). *Grammaticalization and Reanalysis in African Languages*. Hamburg : Buske.
- HOPPER, P.J. (1991). Dispersed verbal predicates in vernacular written narrative. *Berkeley Linguistics Society* 17, 402-413.
- KAISSÉ, E. (1985). *Connected speech: the interaction of syntax and phonology*. San Diego : Academic Press.
- KLAUSENBERGER, J. (1984). *French liaison and linguistic theory*. Stuttgart : Franz Steiner Verlag Wiesbaden GMBH.
- KRUG, M. (1998). String frequency: a cognitive motivating factor in coalescence, language processing and linguistic change. *Journal of English Linguistics* 26, 286-320.
- LANGACKER, R. (1987). *Foundations of cognitive grammar, vol. 1: Theoretical prerequisites*. Stanford : Stanford University Press.
- MORIN, Y.C. & KAYE, J.D. (1982). The syntactic bases for French liaison. *Journal of Linguistics* 18, 291-330.
- POST, B. (2000). *Tonal and phrasal structures in French intonation*. The Hague : Thesus.
- SELKIRK, E. (1974). French liaison and the \bar{x} -notation. *Linguistic Inquiry* 5, 573-90.
- TRANEL, B. (1981). *Concreteness in generative phonology: Evidence from French*. Berkeley and Los Angeles : University of California Press.
- VOGEL, I. & KENESEI, I. (1990). Syntax and semantics in phonology. In Inkelas and Zec (eds.), *The phonology-syntax connection*. Chicago : University of Chicago Press, 339-363.