Frequency Effects on French Liaison

Joan Bybee
University of New Mexico

1. Phonology-syntax interface

A mainstay of the debates concerning the phonology-syntax interface are phenomena of external sandhi, that is, phonological alternations whose conditioning environment is across a word boundary. [1]A recurrent problem in this area is the fact that it is usually impossible to motivate a purely syntactic account of such alternations. This has led to the widespread consensus that the relation between syntax and phonology is indirect and often seemingly arbitrary (Inkelas and Zec 1995, Nespor and Vogel 1986, Vogel and Kenesei 1990). In this paper, I propose a solution to the problem of predicting the sites for external sandhi, and I pose certain questions that are seldom raised in such discussions, in particular, how is it that syntactic structures can condition phonological alternations, and how is it that such alternations can develop and be maintained?

It is important first to distinguish between phonetically conditioned processes that operate across word boundaries and those lexicalized instances that occur only within fixed phrases or constructions.[2] Phonetically-conditioned processes are observable within pause groups, wherever their conditioning environment occurs, both inside of words and across boundaries, allowing for some variation. Examples are Spanish s-aspiration at early stages of development, the spirantization of voiced stops in Spanish, vowel coalescence in Spanish, English flapping of coronal stops, French *enchaînement* (resyllablification), just to name a few. Phonetically-motivated processes sometimes give rise to word-level alternations, that is, cases of a single morpheme with two variants in two different words. I have argued in Bybee 2000a that while morphemes commonly develop alternations, words tend NOT to develop variants; rather there is a strong tendency for a single word to have a single variant or small range of variation. Yet sometimes alternations DO develop such that a single word will have more than one stable variant. In this case, we get the second type of external sandhi, which is no longer phonetically-conditioned, but rather applies across word boundaries in particular constructions. I will claim that this only happens in high frequency phrases and constructions, and that the establishment of such alternations provides evidence that, indeed, these phrases and constructions are stored in memory just as invariant words are.

---

[2] Kaisse 1985 and Hayes 1990 also make this distinction.

In the case of French liaison, which we will discuss here, it is often claimed that liaison occurs most commonly in phrases with 'greater syntactic cohesion' and yet, there are no definitions of this cohesion that correctly predict all cases of liaison. I will demonstrate that this syntactic cohesion is a direct result of frequency of co-occurrence: words that are used together more often tend to seem more fused and also tend to have more liaison. I will argue against a mismatch between syntactic and phonological structure, and will argue instead that the phonology provides good evidence for storage and processing units. Since I do not see the need to posit any grammatical structure that is independent of processing and storage, I will argue that the phonology provides excellent evidence for the nature of the syntactic structure.

2. French liaison

One of the best-studied cases of alternations between versions of the same word under putative syntactic conditions is the case of French liaison. Liaison is the name for the appearance of a word-final consonant before a vowel-initial word in words that in other contexts end in a vowel. Thus, the third singular copula *est* is pronounced [t] in example (1a and b) and as [   ] in example (2a and b) (the *s* is never pronounced). Examples from Green and Hintze 1988.

(1) (a) …le climat est [t] également très différent.
         'The climate is also very different'
    (b) C'est [t] encore un refuge de notables.
         'It's still a refuge for famous people'

(2) (a) C'e(st) le meurtre.   'It's murder'
      (b) le Conseil Régional qui e(st) donc son assemblée délibérante…
         'The Regional Council which is thus their deliberative assembly'

The phonological conditions for the appearance of the liaison is before a vowel-initial word, but only under certain syntactic conditions. In (3) and (4), the presence of the plural liaison is obligatory for the definite article *les.* However, in the noun phrase in (3), the plural morpheme on the noun may variably appear before a vowel-initial adjective, while in (4) the presence of [z] on this same noun is not possible, as the construction involves an NP subject and its verb.

(3) le[z] enfant[z] intelligent   'the intelligent children'

(4) le[z] enfan(ts) arrivent.   'the children arrive'

Moreover, in cases such as (1) and (3) where liaison is possible, there is currently considerable variability. In the databases studied by Ågren 1973 and by

Green and Hintze 1988, speakers at times also omitted the consonant before a vowel. Such omissions are taken as evidence that liaison is disappearing in some contexts.

In the next section I will examine the evolution of liaison from its initiation as a phonetically-conditioned consonant deletion to its present state as attested in conversation, in which it is highly lexically and morphologically governed, as argued by Baxter 1975, Green and Hintze 1988, Klausenberger 1984, Morin and Kaye 1982 and Tranel 1981. I will argue that the morphosyntactic and lexical contexts in which liaison became established occurred with high frequency and were thus sequences that could be stored in memory. Currently, as liaison is being lost, we see that it is maintained in the contexts that are most frequently occurring (Ågren 1973, Delattre 1966). Please note that the treatment of liaison offered here is illustrative only and not by any means exhaustive.

3. Final Consonant Deletion in French

The source of the liaison alternations is the deletion of word-final consonants before another consonant. It is important to note that this was only a specific instance of the more general deletion of syllable-final consonants, which was entirely phonetically conditioned. The first wave of such deletion occurred very early in French, when consonants that had been final in Latin were lost, e.g. Latin *pontem* 'bridge' and *caput* 'chief' lost their final consonants. A subsequent development was the loss of final post-tonic vowels, which created another full set of final consonants (giving e.g. *pont, chef* from the Latin words given above). The stops and fricatives in final position developed two or sometimes three alternates in the environments before a pause, a consonant or a vowel. Harris 1988:213 gives the example of *dix* 'ten' which is pronounced [dis] before a pause, [di] before a consonant (*dix femmes* 'ten women) and [diz] before a vowel (*dix élèves* 'ten pupils). More commonly today, where the alternations persist, the only two variants are the presence and absence of the consonant. However, it is interesting that while this consonant deletion was in progress, the preconsonantal and prepausal conditions were distinguished. The deletion occurred earlier in preconsonantal position than before pause (Klausenberger 1984). This fact suggests that an important phonetic condition for the deletion was the masking provided by a following consonant, and that the spread of deletion to prepausal position was due to the restructuring of the lexical representations.

The result of this phonetic change was that many words, notably nouns and some adjectives, lost their final consonants completely. For instance, nouns such as *haricot* 'kidney bean', *buffet* 'sideboard', *bois* 'forest', *goût* 'taste', *tabac* 'tobacco' and *sirop* 'syrup' are pronounced without a final consonant. However,

words that occurred frequently in particular grammatical or idiomatic conditions that placed them before a vowel, tended to develop an alternation. Word-internally such conditions existed before the feminine suffix, which was vocalic and thus yielded alternations between masculine and feminine nouns and adjectives, such as found in [p   ˌ'small (masc.)' and [p   ˌ'small (fem.)', which today, with the loss of final schwa, yields the alternation [p   ˌˌ p   t ˜

      In French of the 16th and 17th centuries, when final consonant deletion was being implemented, there was a strong tendency, as there is today (Green and Hintze 1988), for forward resyllabification in case a final consonant was followed by a vowel. This process, known as *enchaînement,* makes a final consonant syllable-initial when a vowel follows within the same pause group. As I argued in Bybee 2000a, there is a strong tendency towards a single representation for individual words, the result of which was that many words simply lost their final consonants. However, for grammatical words and grammatical morphemes multiple representations according to the constructions they frequently occur in are possible, so that many such words or morphemes that frequently occurred in constructions that put them in prevocalic position maintained their liaison consonant in those constructions (Berkenfield, this volume, Bybee and Scheibman 1999, Jurafsky et al. this volume). Examples are shown in (5) – (12), based on Morin and Kaye 1982 and Tranel 1981:233. The use of the liaison consonant in (5), (6) and (7a&b) are considered obligatory, while in the remaining contexts, liaison is variable in spoken French.

(5) Determiners

      a. *vos [z] enfants*   'your children'
      b. *les [z] autres*   'the others'
      c. *un [n] ancien ami*   'an old friend'

(6) Clitic pronouns

      a. *nous [z] avons*   'we have'

      b. *ils [z] ont*   'they have'

(7) Person/number endings

      a. *allons [z]-y*   'let's go'

      b. *chante-t-il*   'does he sing?'

      c. *nous vivons [z] à Paris*   'we live in Paris'

      d. *ils chantent [t] en choeur* 'they sing in chorus'

(8)   Plural /-z/ in noun-adjective constructions

     a. *enfants [z] intelligents*   'intelligent children'

     b. *des découvertes [z] inquiétantes*   'worrisome discoveries'

(9) A small set of masculine singular adjectives that occur before their nouns

     a. *un petit [t] écureuil*   'a little squirrel'

     b. *un gros [z] amiral*   'a fat admiral'

     c. *un long [g] été*   'a long summer'

(10) The plurals of the same

     a. *Deux petites [z] histoires.*   'Two short stories'

     b. *Quelques [z] années plus tôt* 'A few years earlier'

(11) Prepositions, adverbs, particles

    a. *dans [z] un mois*   . .   in a month'
    b. *pendant [t] un mois*   . . .   'for a month'

(12) Fixed phrases

     c'est [t] a dire   'that is to say'

     pas [z] encore   'not again'

Despite the variety of morpheme or word types included in this list, it is important to note that there are no purely lexical, or open class items, that exhibit liaison; rather the items range from suffixes, e.g. for plural, to grammatical classes, such as prepositions and small closed classes, i.e. the prenominal adjectives (see section 5 for further discussion of this small class). Two relevant points can be made about such forms: first, they are all of relatively high frequency, especially in the constructions or phrases in which the liaison consonant appears; and second, they all occur in very specific grammatical constructions. No liaison consonant appear independently of a specific construction. The situation of liaison consonants within specific constructions will be discussed in the next section.

4.   Grammatical constructions and liaison

The role of morphology, syntax and lexicon have been widely recognized in accounts of French liaison, but the relative contribution of each has been debated in the literature and very little has been said about the role of frequency in establishing and maintaining liaison consonants. Perhaps the most monolithic approach is that of Selkirk 1974 which attempts to derive liaison contexts by reference to the placement of word boundaries (#, ##) according to the principles proposed in Chomsky and Halle 1968, which place single word boundaries around members of lexical categories, but not members of grammatical categories. These principles define a phonological word as the material between instances of two word boundaries (##). Selkirk proposes that liaison occurs only within the phonological word. Because of the way #'s are placed, the effect of these principles is to say that liaison occurs when grammatical and not lexical morphemes are involved. This analysis works well for most cases, but because it treats all instances of liaison as involving grammatical morphemes, it leaves open the question of how prenominal adjectives will be treated, since they are technically lexical in generative theory. Kaisse 1985 proposes that liaison takes place in a sequence *ab* if *b* is the head of the phrase and c-commands *a* (that is, if *a* is in the phrase of which *b* is the head). The data we will examine in this section and the next shows that the variability of liaison is highly affected by the very specific location of the grammatical element in a construction, and that all cases of liaison do not have the same status in terms of their productivity and degree of entrenchment, indicating that a single syntactic principle is not likely to be successful in predicting liaison contexts.

The syntactic principles proposed so far are adequate for the obligatory liaison in examples (5) and (6), which involve determiners with their nouns and clitic pronouns with their verbs, but it is not possible to successfully extend either of these proposals to cases where liaison is considered variable. Indeed, any syntactic proposal referring to the head of a phrase is going to have trouble applying to both noun-adjective combinations (example [7]) and adjective-noun combinations (examples [9] and [10]) (de Jong 1990). In fact, proposals based on traditional assumptions about constituent structure run into difficulties with the fact that liaison occurs 98.7% of the time in a sequence *est [t] un + NOUN* '3rd Sg. is a NOUN' but about 47% of the time in the sequence *je suis un + NOUN* 'I am a NOUN', which presumably has the same constituent structure; similarly *est + PAST PARTICIPLE* has an extremely high rate of liaison, 98.6%, while *je suis + PAST PARTICIPLE* has only 57% liaison (data from Ågren 1973).

Another approach is to postulate a level of prosodic organization and stipulate that liaison applies within units so organized, i.e. phonological words or phrases (Selkirk 1986, de Jong 1990). Such proposals were tested in the experiments of Post 2000, but her attempts to find a correspondence between the prosodic unit of phonological word, and the occurrence of liaison failed to produce significant results.

Most other authors (Baxter 1975, Green and Hintze 1988, Klausenburger 1984, Morin and Kaye 1982 and Tranel 1981) have offered an analysis that refers to both morpho-syntactic and lexical factors. Like other alternations that have become lexicalized or morphologized, what was a unitary phonological change has become associated with particular grammatical contexts and is no longer unitary (cf. Bybee 2001, Chapter 5). The account offered here is based on these previous treatments and will only mention some of the liaison environments, as the main point will be to underscore the role of frequency and phonological material situated in constructions. I will assume that frequency of use played a major role in the establishment of these alternations, and I will argue on the basis of modern data that frequency of use plays a major role in preserving liaison alternations. In particular, I will argue that the 'degree of syntactic cohesion' that is often mentioned in studies of liaison is a direct result of the frequency with which the two items surrounding the liaison consonant occur in sequence.

Evidence in support of the view that liaison is morphologized or lexicalized is the fact, pointed out in various studies, most explicitly in Encrevé 1983. Morin and Kaye 1982 and Post 2000, but also in Ågren 1973 and Green and Hintze 1988, that liaison consonants can occur both before and after a pause, or with and without forward syllabification or *enchaînement*. In other words, while liaison originally depended upon forward resyllabification, it is not now restricted to occurring within a phonological word. On the other hand, *enchaînement*, which is still a viable process in Modern French, occurs only within pause groups and not across them.

The current approach to liaison takes the construction as the basic unit, and since constructions often contain very specific lexical and grammatical material, attributes the liaison consonant to the construction itself. Constructions are repeated sequences of morphemes or words which bear a particular semantic or functional relation to one another when used together in a construction which they do not necessarily have outside that construction. Constructions have different degrees of conventionalization, as they come to be established in a language through repeated use. The mechanisms for the establishment of constructions are (i) automation of chunks of linguistic material due to repetition, and (ii) categorization of the items occurring in particular positions in these larger chunks. Because repeated use is a major factor in the formation of constructions, it will NOT necessarily be the case that constructions have unpredictable meaning—they can simply be oft-used chunks of language. However, owing to the autonomy that accompanies repetition and frequency of use, constructions will often take on non-transparent meanings.

On one end of a continuum involving constructions are fixed phrases, such as *I don't know* and *c'est à dire* 'that is to say', nearer the middle are

7

constructions with some grammatical material and a slot that is more open, e.g. the preposition *dans* with its NP object, and on the most general end, a construction such as [NOUN + PLURAL + ADJECTIVE], with two slots which take open class items. It seems useful to restrict the term 'construction' to sequences that include a more-or-less open slot and to classify phrases without open slots, such as *c'est à dire* as fixed phrases. The open slots in constructions are subject to categorization in terms of semantic features (such as 'motion verb') or grammatical features (such as pronoun). Since constructions arise from frequently-used stretches of speech, it is not necessary for the organization of items in a construction to correspond to traditional notions of constituency, as we will see in examining some liaison contexts. It is important to bear in mind, however, that traditional notions of constituency are also derivable from frequency of co-occurrence, since items that go together in a semantic sense tend to occur together in discourse.

One construction that is much discussed in the liaison literature involves a plural noun followed by an adjective which begins with a vowel. In some cases, a [z] occurs between the noun and adjective, a remnant of the plural marking that has been deleted when a consonant follows. The examples from (7) are repeated here as (13).

(13)   a. *enfants [z] intelligents*   'intelligent children'

        b. *des découvertes [z] inquiétantes*   'worrisome discoveries'

Agren 1973 reports that liaison in such contexts is considered obligatory in certain frequent phrases, some of which are proper nouns: *affaires [z] américaines* 'American affairs', *Champs [z] Elysées, Nations [z] Unies* 'United Nations', *Jeux [z] Olympiques* 'Olympic Games' and so on. Among the optional contexts, liaison only occurs in 26% of cases with a plural noun followed by an adjective. Most authors regard this construction with liaison as somewhat productive, however, owing to the existence of examples such as these offered by Morin and Kaye 1982, in which the [z] liaison occurs for plural but at some remove from its etymological site:

(14)   (a) *des chefs d'Etat [z] africains*   'African heads of state'

        (b) *les chemins de fer [z] anglais*   'the English railways'

Morin and Kaye 1982 argue that the plural liaison does not just occur in lexicalized expressions, but also applies productively in these cases.

The data suggest two constructions for plural noun-adjective expressions. The more general one contains a plural determiner followed by an unmarked noun and adjective.

(15)   *les*

[*ces*   NOUN   ADJECTIVE]

       des   plural

       etc.

A second, more restricted construction, applies only to vowel-initial adjectives:

(16)   *les*

[*ces*   NOUN   -z-   [Vowel]-ADJECTIVE]

       des   plural

       etc.

Besides the second construction (16) being restricted to vowel-initial adjectives, there is another difference between them: the first construction (15) applies to more items, that is, it has a higher type frequency, which make it more productive than the second one. Thus it is not surprising that there is variation in the data resulting from speakers choosing the more general schema even for vowel-initial adjectives in most cases. Still the more specific schema with the [z] before vowel-initial adjectives is available and is sometimes used. Thus the loss of liaison resembles regularization of irregular verbs: if the specific schema is not easily accessed, then the more general one, which is stronger and easier to access, is used. (see section 7.4).

In addition to the schemas in (15) and (16) there might also be more specific schemas for adjectives that are frequently used with the [z], such as *anglais,* or *américain:*

(17)   *les*

[*ces*   NOUN   -z-   *anglais*]

       des   plural

The examples in (14) result from the phrasal nouns *chefs d'Etat* and *chemin de fer* occurring in the NOUN position in the constructions (16) or (17).

Another plural construction that results in cases of 'false liaison' consists of the cardinal numbers plus [z] and a noun. As reported in Tranel 1981:214-16, liaison is frequently maintained with the cardinal numbers *deux* [d   'two' and *trois* [trwaz] 'three' before vowels when in the same construction as the following noun and when plurality is indicated. Thus *les deux [z] amis* 'the two friends' has liaison, but *le deux octobre* 'October 2nd' does not. The viability of the [z] as a plural marker in this construction is evidenced in the widely-reported use of

cardinal numbers without etymological final [z] in this construction. Tranel 1981:216 gives the following examples:

(18) (a) *quatre enfants*   [katz   ]   'four children'
      (b) *huit épreuves*   ['eight events'
      (c) *neuf oeufs*   ['nine eggs'
      (d) *vingt-cinq années* [twenty-five years'
      (e) *trois mille évêques* [trwamilzev   k] 'three thousand bishops'

Not only do these examples show the intrusion of a non-etymological [z], but examples (c) through (e) also show that the other final consonant of the number is present. Thus this construction goes beyond the function of creating optimal syllable-structure to a truly morphological use, where the [z] is signaling plurality.

On the basis of examples such as those in (18), Tranel 1981 and Morin and Kaye 1982 and Klausenberger 1984 argue for an analysis which inserts the liaison consonant in certain contexts before a vowel, rather than deleting it before a consonant. The solution proposed here is neutral with regard to insertion or deletion. It simply states that a construction exists which contains the [z] after a number and before a vowel-initial noun. This construction could be formulated as in (19):

(19)

$$\left[\ \text{NUMBER} \quad \text{-z-} \quad \text{[vowel]- NOUN} \ \right]$$
                                                    plural

The construction in (19) is a generalization from the conservative usage, in which *deux, trois, six* and *dix* were the only numbers that had [z] before vowel-initial nouns. The forms in (18), then, provide positive evidence for the tendency to extract generalized or more schematic constructions from more specific instances of use.

As mentioned above, Ågren 1973, Morin and Kaye 1982, and Green and Hintze 1988 all report that the liaison consonant can appear even if a pause or hesitation syllable occurs between the two words of the construction. Green and Hintze 1988:159 found an example with the number *quatre* in their data:

(20) *quatre euh…[z] obligations*   'four uh… obligations'

Such examples show that it is possible to pause or hesitate in the middle of a construction, just as it is possible to pause in the middle of a word. Since the words of a construction are usually associated with other instances of the same word, their identity as words is known, and the point between two words is a

possible place to pause. The position of the pause in this and other examples reported in these works suggests that liaison consonant is more associated with the second word, where it begins the first syllable, than with the first, which was its historical source.

Another illustration of the close association of particular liaison consonants with particular constructions is in the reciprocal construction, discussed in Morin and Kaye 1982. These authors report that liaison is optional after *l'un* 'the one', but only when used in the reciprocal construction; elsewhere liaison is not possible. Thus in (21) liaison with [n] is an option, but in (22) it is not:

(21)   (a) *Il les a confundus l'un [n] avec l'autre.*
                'He took them for each other'
         (b) *Ils se ressemblent l'un [n] à l'autre.*
                'They look like each other'

(22) *Ils sont venus, l'un avec sa mère, l'autre avec son père.*
        'They came, one with his mother, the other with his father'

This example in particular shows that grammatical morphemes are highly entrenched in the constructions in which they appear, not just in French, but in all cases. A grammatical morpheme is identified as such because of its appearance in certain well-defined grammatical constructions. The history of grammatical morphemes shows that if they occur in different constructions, they move away from one another in phonological shape, meaning and distributional properties (Heine and Reh 1984, Hopper 1991; see also Berkenfield, this volume, for a study of the beginnings of such a process with English *that*). In French there are several grammatical morphemes deriving from *un,* which originally was only the numeral 'one': *l'un* 'the one', as in example (20), where liaison is not possible; the reciprocal, as in (19), where liaison is optional; and the indefinite article, *un, une,* in which use liaison is considered obligatory.

5.   Loss of liaison as regularization

Studies of optional liaison, such as Ągren 1973, show a tendency for the loss of liaison in many contexts. As mentioned above, the observation is often made that liaison is maintained in cases of 'tighter syntactic cohesion' (Tranel 1981); however, no one has offered a definition of this syntactic cohesion that is detailed enough to make correct predictions across the numerous constructions involved in liaison. It has been noticed and amply documented in Ągren 1973 that uses with higher frequency maintain a higher level of optional liaison than those that are less frequently used (see also Booij and de Jong 1987). However, no one

has yet zeroed in on frequency of use as a causal factor in the establishment, maintenance or loss of liaison.

My proposal is that French consonant liaison, though it takes place between traditional 'words' rather than word-internally, is very similar to morphologically and lexically conditioned alternations that occur word-internally: it was established with an original phonetic motivation, and the alternations gradually came to be associated with certain morpho-syntactic and lexical contexts. Like other morpho-lexical alternations it is subject to both extension to new contexts in cases of productivity, and leveling or loss of the alternations. The frequency factors affecting these ongoing changes are the same as in the cases of word-internal alternations. High type frequency of a construction spurs productivity. Unproductive alternations are gradually leveled or regularized, with low frequency forms being leveled first, and high token frequency resisting the leveling for the longest time.

What makes this case of special interest is the fact that the units in which the alternations occur are larger than traditional words. For arbitrary alternations to become established and to be maintained in such units, these units must constitute units of storage, just as words do. Thus the facts of French liaison, and other cases of external sandhi are valuable in that they provide evidence for the nature of storage units beyond the traditional word. The evidence presented so far strongly suggests that frequent fixed phrases are storage and processing units, as are constructions containing grammatical morphemes. Among the latter, more specific and more general constructions compete, leading to the gradual loss of the more specific construction, which in this case is the one with liaison.

I have already mentioned that grammatical morphemes are entrenched in constructions; to describe this situation, I have proposed that grammatical constructions contain these grammatical morphemes as explicit phonological material. Thus the same grammatical morphemes in different constructions are independent of one another. The more frequently used a construction is, the greater likelihood that its form will be maintained, rather than being replaced by some more productive construction (Bybee and Thompson 1987). It is not surprising, then, that certain liaison contexts, in particular those involving articles and their noun (see examples [5]) and those involving clitic pronouns and their verb (examples [6]), are obligatory by all accounts and not tending towards loss of liaison. These constructions are those that are apparently regarded as having the tightest syntactic cohesion, a cohesion that could be attributed to frequency of co-occurrence. No relative frequency counts are available to prove this point, but given the fact that almost all NPs have either a definite or indefinite article and in spoken language subject and object clitic pronouns are used redundantly (Harris 1988:231-232, 235-236), the high frequency of these construction cannot really be in doubt.

Other reasons exist for regarding [ARTICLE + NOUN] constructions and [CLITIC + VERB] constructions as storage and processing units in Modern French. The maintenance of gender distinctions, which are overtly signaled primarily in the singular article suggest the storage of the article, both singular and plural, with the noun. Studies of spoken French usage demonstrate that the subject and object clitic pronouns are now almost obligatory accompaniments to the verb, behaving perhaps more like prefixes than clitics, again suggesting lexical status (Harris 1988:232).

The special treatment of articles with *h-aspiré* words also points to lexical representation of articles with nouns. These are words which are vowel-initial but do not take a liaison consonant, even in the obligatory contexts. Thus *des haricots* 'the green beans' is pronounced [d hariko] in standard French. If *h-aspiré* words behaved as though they were consonant initial with respect to liaison from all sources and with respect to elision (vowel deletion at the end of the preceding word), then it would make sense to treat them as though they were consonant-initial. However, Tranel 1981 reports that the exceptional status of these words is maintained most strongly in those contexts in which the syntactic constituency is tighter (p. 300-301, n. 4).

If frequency of co-occurrence is the main factor governing the appearance of the liaison consonant, then we would not expect to find such consonants between any two randomly selected lexical items, whose probability of co-occurrence is extremely low. Indeed, we do not. One case that might appear to contradict this claim is the small class of prenominal adjectives that link to a following vowel-initial noun, even in the masculine.

(23)   a. *un petit [t] écureuil*   'a little squirrel'

       b. *un gros [z] amiral*   'a fat admiral'

       c. *un long [g] été*   'a long summer'

The important point about the construction represented by these examples is that it is restricted to a small set of adjectives, many of which have a different meaning when used pre-nominally rather than in the more common post-nominal position. They are, in a sense, partially grammaticized and not fully lexical in this construction. Still, the frequency of such adjectives in this construction, and their resulting 'syntactic cohesion', must be less than some of the other fully grammatical morphemes exhibiting liaison. It is thus predicted that liaison in this context will be maintained less than in other contexts.

(24) Prenominal adjectives that condition liaison

13

bon/bonne		'good'		
long/longue		'long'		
nouveau/nouvelle		'new'		
mauvais/mauvaise		'bad'		
grand/grande		'great, big'		
gros/grosse		'fat'		
petit/petite		'little'		

A real test of the frequency hypothesis is possible with the data reported in Ągren 1973, where different inflectional forms of the same word with different frequencies and the same word in different constructions with different frequencies can be compared for the maintenance of liaison. First consider the forms of the copular verb, *être*. Ągren points out that the presence of liaison is directly related to the token frequency of these forms. He gives the data in Table 1, which shows the number of cases of liaison (L) and non-liaison (NL) and the number of times each item was used in the data he analyzed, listed according to the percentage of cases of liaison.

Table 1. Number of instances of liaison for the forms of the verb *être*.

	L	NL	total	percentage of liaison
est (3rd Sg. Pres. Ind.)	2591	77	2668	97%
sont (3rd Pl. Pres. Ind.)	242	38	280	86%
étant (Pres. Part.)	22	7	29	76%
était (3rd Sg. Impf.)	272	95	367	75%
êtes (2nd Pres. Ind.)	24	10	34	71%
étaient (3rd Pl. Impf.)	36	21	57	63%
sommes (1st Pl. Pres. Ind.)	43	31	74	58%
suis (1st Sg. Pres. Ind.)	65	74	139	47%
serait (3rd Sg. Fut.)	17	24	41	41.4%
soit (3rd Sg. Pres. Subj.)	22	32	54	40.7%
j'étais (1st Sg. Impf.)	6	23	21	21%

The correspondence between token frequency and percentage of liaison is quite close, except for two important exceptions, both of which are explained by Ągren. The first is the high percentage of liaison for the not-so-frequent Present Participle, *étant.* Here Ągren points out the high percentage of occurrence of liaison in the construction *étant + Past Participle*, which has liaison in seven out of eight cases in his data. Four of these cases are the fixed phrase *étant entendu* 'being understood, given that'.

The other exception is the 1st Singular Present Indicative form *suis,* which has fewer instances of liaison than predicted by its token frequency. For this case, Ågren observes that it is common to reduce the sequence *je suis* 'I am' to [. In fact, further reduction of this sequence is often noted, even to [~ In Ågren's data this reduced form tends to occur without liaison. In other words, the new contraction of *je suis* does not end in [z], possibly because the reduction of this phrase, which is originally a casual speech phenomenon, is not compatible with liaison, which is more common in more formal styles. This example shows that the reduction of high frequency sequences in casual speech can lead to the establishment of competing constructions. The reduced form can then spread to contexts in which it would not have originally occurred, i.e. before a vowel-initial word.

In addition, there is a general tendency for liaison in verbs to involve the consonant [t], and liaison for noun plural to involve [z]. Thus all verb forms ending in [z] tend to have a lower percentage of liaison than those of comparable frequency involving [t]. This fact suggests the possibility of a more schematic representation that associates liaison [t] with verbs and [z] with noun plurals (Klausenberger 1984).

The wide range of variance for liaison with the forms in Table 1 is especially interesting because these forms are all inflected forms of the same verb, and yet they behave quite differently under liaison conditions.[3] Their usage is regularizing, with the low frequency forms more likely to undergo regularization than the high frequency forms. The mechanism by which this occurs is parallel to the way in which irregular inflected forms such as *weep/wept* regularize. As a low frequency verb, its irregular Past may not be as easy to access as a high frequency verb would be. Thus a new Past can be made for it by using the base form and the regular Past tense construction.

All of the forms of *être* listed in Table 1 occur in two variants, one with and one without a final consonant. The variant without a final consonant is the more commonly occurring. Thus in any given use of these forms there is competition between a construction that is more specific—the one for a word before a vowel-initial word—and the more general construction—the one for the word before a consonant-initial word. The latter construction will apply more often, since consonant-initial words are more common than vowel initial ones (by at least two to one). The more specific construction can be preserved by frequency which increases its lexical strength, but there is always the option of using the more general construction, the one without liaison.

---

[3] Ågren found the same pattern among the inflected forms of the auxiliary *avoir* 'to have' and the semi-auxiliaries (as he calls them) *aller* 'to go', *falloir* 'to be necessary', *pouvoir* 'to be able to', *devoir,* 'to have to', and *vouloir* 'to want'.

The token frequency of the first element alone will not predict the occurrence of liaison, nor should we expect it to; rather the important variable is how often the two elements that are linked occur together, and perhaps also the transitional probability between the first and second element. For instance, Ǻgren counted the frequency of liaison with the auxiliaries *aller* 'to go', *falloir* 'to be necessary', *pouvoir* 'to be able to', *devoir*, 'to have to', and *vouloir* 'to want'. When these are compared to one another, their frequency of occurrence does not correspond neatly with their percentage liaison. Part of this lack of correspondence is due to certain high frequency forms such as *je voudrais* not participating in liaison because (i) final *s* in verb forms shows less liaison, and (ii) because as a fixed expression it has become invariable. However, much of the variation among the auxiliaries is due to different rates of occurrence in specific constructions. All of these auxiliaries occur with a following infinitive; the most frequent infinitive to follow them is *être* 'to be' accounting for 226 out of 604 infinitives after an auxiliary. The second most frequently-occurring infinitive was *avoir* 'to have', which occurred 71 times. We would thus expect the highest rates of liaison to occur with *être*, and indeed this is what is found, as shown in Table 2. In fact, occurrences of *être* can be divided into those that comprise the passive construction and those that are more copular in function. Ǻgren found a different percentage of liaison in the two cases.[4]

Table 2. Ǻgren's findings for the auxiliaries and following infinitives

	avoir		être + PP		être + other word		other Infs	
	L	NL	L	NL	L	NL	L	NL
aller	3	3	5	2	2	6	5	16
devoir	9	7	47	3	39	2	19	16
falloir	2	3	1	1	10	4	42	23
pouvoir	18	20	41	4	40	13	60	59
vouloir	3	3	3		2	1	14	53
	35	36	97	10	93	26	140	167
% liaison	50%		91%		78%		46%	

A frequency effect is evident here, in that *être* is the most common infinitive to follow these auxiliaries, and the most common site for liaison. These findings are particularly clear with *devoir* and *pouvoir* which occur very frequently with *être*.

The very high percentage of liaison with *devoir* and *pouvoir* with *être + Past Participle* suggests very specific constructions for these modals and the passive. The differences between this and the other uses of *être* confirms our

---

[4] In Ǻgren's chart *être + Past Participle* is distinguished from *être + autre mot* 'other word'. I am assuming that most of the cases of the latter are not also passives, though they could be in case the 'other word' is followed by the Past Participle.

statement that grammatical morphemes are very much entrenched in the particular constructions in which they occur. Considerations of function also play a role. When *devoir* and *pouvoir* are used with the passive, their subjects are not the agents of the main verb, and therefore, their subjects cannot be the agents for whom obligation or ability is being predicated. Consider (25) from Ạgren 1973:83:

(25)   *Marie-Claire, est-ce que vous pensez que l'homme et la femme doivent [t] être placés sur le même plan intellectuel et social?*

        'Marie-Claire, do you think that man and woman should be put into the same intellectual and social level?'

This example illustrates the 'root obligation' sense, as no specific source for the obligation is expressed. It expresses only a very general sense of obligation. Similarly (26) expresses 'root possibility'—general conditions exist for the possibility of completing the predication (Ạgren 1973:86). Both root obligation and root possibility are more grammaticized functions than the obligation and ability meanings from which they arise (Bybee, Perkins and Pagliuca 1994, Nordquist 1999).

(26) *Ca prouve, enfin qu'il y a, qu'il y a des choses extraordinaires qui peuvent [t] être faites encore.*

        'This proves finally that there are, there are extraordinary things that could be done still.'

Note that in these cases, the auxiliary will ordinarily be in the third person. Thus the sequences [*doi[t] être + PP*] and [*peu [t] être +PP*] and their plural counterparts are constructions which have particular semantic readings that differ slightly from those of the same modals in active constructions. These functions may reinforce the autonomy of the construction. The viability of the liaison [t] in these constructions is supported by the overgeneralization reported in Morin and Kaye 1982:324:

(27)   *Ca doit bien t-être cuit, maintenant*   'It must be cooked by now'

In this example the modal is fulfilling an epistemic functions, which is a further development from the root obligation reading.

        Another interesting difference to observe in Table 2 is the difference between the cases where *être* is the following infinitive and those cases, which are pooled together, in which a variety of lexical infinitives occur. While, in all, there are more of the latter, neither type frequency nor frequency of the construction as

a whole are the relevant variables. Since the maintenance of liaison is comparable to the maintenance of irregularity in inflected forms, it is the token frequency of the particular sequence that is operable in resisting regularization.

On the other hand, liaison IS still used with lexical infinitives about half the time in the data analyzed. This means that forms such as 3rd Singular *doit* or *peut* occur in constructions that supply the liaison consonant in case the next word begins with a vowel. That is, there are two constructions for *doit* (besides the ones mentioned above)—(28a) is the more general construction that is used with a greater variety of infinitives, while (28b) is the less general one. The more general schema is gradually taking over and replacing the less general one, except in very specific sequences with high token frequency.

(28)   (a)   [[dwa] INFINITIVE   ]
                                                    obligation

         (b)   [[dwa] -t- [vowel]- INFINITIVE   ]
                                                    obligation

6. Transitional probability

A possibility to consider in this case of two-word combinations is that transitional probability might be a better predictor of cohesion and thus liaison than simple string frequency (Bush, this volume, Jurafsky et al. this volume). Transitional probability is the probability that in any occurrence of one word, a particular second word will follow. It is calculated by dividing the number of occurrence in a text of XY by the number of occurrences of X (Saffran, Newport and Aslin 1996). Very often, string frequency and transitional probability make the same predictions, but in some cases where string frequency is high, but transitional probability is not so high due to the occurrence of a wide variety of other elements after X.

A chance to distinguish string frequency and transitional probability is afforded by the data in Table 2, which may point to transitional probability as an important factor in cohesion. As shown in Table 3 (based on Table 2), the rate of liaison before *être* is lower after *pouvoir* than after *devoir*, especially when *être* is followed by a non-Past Participle word. This could be related to the fact that *pouvoir* is followed by other infinitives much more often than *devoir* is. The fact that *pouvoir* is used with many other infinitives lowers the transitional probability for *être* after *pouvoir*.

Table 3. Item following *devoir* and *pouvoir* (Ågren 1973)

	être + PP		*être + other word*		*other infinitives*		*total*
	L	NL	L	NL	L	NL	

18

devoir	47	3	39	2	28	23	142
pouvoir	41	4	40	13	78	79	254

We can use the figures in this table to calculate the transitional probability for the combinations *devoir + être* and *pouvoir + être*, as shown in Table 4. Note that the number of tokens of these sequences are approximately the same, but that *pouvoir* is used much more often before other infinitives than *devoir* is.[5]

Table 4. Comparison of the rate of liaison and transitional probability for *devoir* and *pouvoir*

	être + PP		*être + other word*	
	percentage of L	*TP*	*percentage of L*	*TP*
devoir	94%	.35	95%	.29
pouvoir	91%	.18	75%	.21

The lower transitional probability for *être + PP* after *pouvoir* does not have much of an effect on this construction, possibly because of its grammatical status as passive. But for *pouvoir* with *être* followed by some other word, the low transitional probability relative to *devoir* corresponds to a lowered rate of liaison, though the differences in transitional probability are small and may not be significant given the number of instances available.

What would be the causal mechanism that is reflected in the relation between liaison and transitional probability? Two possibilities present themselves. First, the frequent co-occurrence of X+Y leads to a strong sequential connection between these units, but if X also occurs before other items frequently (thus having a lowered transitional probability before Y), the connection between X and Y often has to be suppressed, which could weaken the connection. A second possibility refers more to type frequency than transitional probability. Since an item such as *peut* occurs before many different infinitives, the construction used with consonant-initial infinitives will have a higher type frequency and thus be more likely to cause regularization than the construction with *doit,* which would have a lower type frequency. Further evidence for transitional probability must be sought in cases that would differentiate between these mechanisms.

Other evidence for the workings of transitional probability in French liaison is the fact that the coordinating conjunction *et* never conditions liaison. This is a very high frequency morpheme, and undoubtedly occurs in some high frequency combinations. However, the fact that virtually any word of the

---

[5] The reason for this seems to be that *falloir*, which is very similar in meaning to *devoir* is commonly used with other infinitives instead of the third person forms of *devoir*, see Table 2.

language can follow *et* gives any particular word a rather low transitional probability, possibly explaining why there is no liaison with this item.

7. Syntactic cohesion as frequency of co-occurrence

The data examined here concerning French liaison supports the view that what has been called 'syntactic cohesion' is frequency of co-occurrence, the factor which determines the strength of the association between the first element and the second one. These connections are stored in memory and reinforced by frequent use. The evidence for their memory storage is that the principles that we have established for morpho-lexical alternations are operable at what has been taken to be this higher level of organization as well. In particular, the higher the frequency of the phrase or construction, the more likely it is to preserve liaison; the lower its frequency, the more likely it is to lose liaison by the application of a more general construction.

Collocations of words that are used frequently have strong memory representations. Just as morphologically complex words that are high frequency are more autonomous from their own paradigms and paradigms of other words, so high frequency phrases grow more autonomous. That is, the connections between the words and morphemes of such phrases and other instances of the same words or morphemes in other constructions becomes weaker. The potential loss of association is heightened by phonetic and semantic or functional change. The extreme outcome of this process is seen in grammaticization, where parts of grammaticized constructions are no longer associated with their lexical sources, e.g. the difficulty English-speaking children have when they begin to read of identifying the form they know as *gonna* with the three morphemes, *go, ing* and *to*. In most liaison contexts this loss of internal structure is not so extreme, but the 'syntactic cohesion' referred to operates by this same mechanisms: frequent sequences are processed together and this unity breaks down their relations with related items.

8. Taking the phonology seriously

It is common practice to try to predict French liaison, English *don't* reduction, and other cases of words that have variants by reference to syntactic constituency and relations. However, all such attempts leave some cases unaccounted for, and many such analyses still require special mention of certain lexical items or grammatical morphemes. Many researchers have thus concluded that the relation between syntactic structure and phonological rule application is indirect (Vogel and Kenesei 1990 among others). Interestingly, a position that is not often taken is that the relation between syntax and phonology is quite direct,

but that we are not operating with the correct syntax. In other words, it is not usually argued that phonological evidence suggests different syntactic structures, except at the most surface level. In contrast, in determining morphological structure, phonological evidence is often taken into account. Deciding whether or not a grammatical morpheme is an affix or not often involves some consultation of the phonological fusion between the proposed affix and stem. What would be the consequences of letting the phonological tail wag the syntactic dog?

In many cases, of course, nothing would change. In French, as liaison indicates, determiners go with nouns and clitic pronouns go with verbs. The only innovation I have proposed is that [DETERMINER + NOUN] and [CLITIC PRONOUN + VERB] sequences are stored in memory. Similarly, [ADJECTIVE + NOUN] and [NOUN + PLURAL + ADJECTIVE] sequences are constituents syntactically. However, an English sequence of [PRONOUN + AUXILIARY], such as *I'll* and *I'm* are not usually considered constituents. In fact, the highest-level syntactic break within a clause—that between subject NP and VP occurs within this sequence. Yet it is undeniable that auxiliaries contract and fuse with subject pronouns, not with the following verb, even though the auxiliary and verb are in the same constituent. The reason proposed (Bybee and Scheibman 1999, see also Krug 1998) is that specific instances of [PRONOUN + AUXILIARY] are of extremely high frequency, much higher than any particular sequences of [AUXILIARY + VERB]. The phonological and usage facts, then suggest an analysis of English much like that of Quileute, where the forms of subject pronouns are determined by modal functions (Andrade 1933:203ff).

Another interesting case of French liaison concerns the copular verb in 3rd Singular, *est*. In Ągren's data, 47% of the uses of *est* occur in the construction [*est + un* + noun] 'is a noun'. In this sequence, liaison occurs 98.7% of the time, much more than with any other uses of *est*. This strongly suggests a construction in which *est [t] un* is a constituent that precedes a noun. A comparable claim about English would be to say that *is a* is a constituent because of the frequency of use of these two items together. Of course, to say that *est un* or *is a* is a constituent is not to say that *un + noun* and *a + noun* are not also constituents. There is no reason why two constructions cannot overlap, giving ambiguous constituent analyses in cases such as these.

If we take usage as the determinant of constituency and syntactic hierarchy such that items frequently used together are constituents, then phonology is a valid indicator of constituency, since the same property, frequent co-use, conditions the phonological alternations.

9. Conclusion

In a model in which memory storage includes not just individual words, but also phrases and constructions, lexicon and grammar are not strictly separated, but are integrated and subject to the same organizational principles (Langacker 1987, Bybee 1998). Any repeated stretch of speech can be stored in memory and placed into categories with identical and similar units. Categorization occurs at multiple levels. Exemplars of the same word or phrase are mapped onto a single representation. Tokens of the same construction are similarly mapped onto a representation, and the items in the variable positions of the construction contribute to the formation of categories based on their semantic properties. Thus in the English construction *X is going to Y*, the occurring tokens contribute to the formation of the categories *X* and *Y*. In the French construction [NUMBER + z + [vowel]-NOUN], the occurring tokens create the categories NUMBER and NOUN.

While this paper is superficially about phonology-syntax interactions, the main goal has been to argue that constructions have many of the same properties as morphologically-complex words. Elements (both phonemes and morphemes) within constructions frequently co-occur and can undergo phonological reduction and fusion just as material inside of words can. Then alternations can be preserved inside of constructions which are of high frequency. Alternations are also subject to leveling if some other more general construction produces a new, regularized way of saying the same thing. Alternations in lower frequency constructions and phrases are then leveled first, with the more entrenched alternations remaining. Thus constructions that encompass more than one word are more entrenched (resistant to change) if they have high token frequency, and more productive if they have high type frequency. These properties all imply that constructions are storage and processing units just as words and fixed phrases are.

References

Ågren, John. 1973. *Etude sur quelques liaisons facultatives dans le français de conversation radiophonique: frequence et facteurs*. Uppsala: Acta Universitatis Upsaliensis.

Andrade, Manuel J. 1933. 1933. *Quileute.* New York: Columbia University Press.

Bush, Nathan. This volume. The predictive value of transitional probability for word-boundary palatalization in English.

Bybee, Joan. 1998. The emergent lexicon. *CLS 34: The Panels,* 421-435.

Bybee, Joan. 2000a. Lexicalization of sound change and alternating environments. In Michael Broe and Janet Pierrehumbert (eds.) *Papers in Laboratory Phonology V: Acquisition and the Lexicon.* Cambridge: Cambridge University Press, pp. 250-268.

Bybee, Joan. 2000b. Phonology of the lexicon. In Michael Barlow and Suzanne Kemmer (eds.) *Usage-based models of language.* Standord: CSLI, pp. 65-85.

Bybee, Joan. 2001. *Phonology and language use.* Cambridge: Cambridge University Press.

Bybee, Joan and Joanne Scheibman. 1999. The effect of usage on degree of constituency: the reduction of *don't* in American English. *Linguistics* 37.575-596.

Booij, Geert and Dann de Jong. 1987. The domain of liaison: theories and data. *Linguisitcs* 25.1005-1025.

Chomsky, Noam and Morris Halle. 1968. *The sound pattern of English.* New York: Harper and Row.

Delattre, Pierre. 1966. *Studies in French and comparative phonetics.* The Hague: Mouton.

Encrevé, Pierre. 1988. *La liaison avec et sans enchaînement: Phonologie tridimensionnelle et usage du français.* Paris: Seuil.

Green, John N. and Marie-Anne Hintze. 1988. A reconsideration of liaison and enchainement. *Occasional Papers.* University of Essex: Department of Languages and Linguistics, pp. 136-168.

Harris, Martin. 1988. French. In Martin Harris and Nigel Vincent (eds.) *The Romance Languages.* Oxford: Oxford University Press, pp. 209-245.

Hayes, Bruce. 1990. Precompiled phrasal phonology. In Inkelas and Zec (eds.) pp. 85-108.

Inkelas, Sharon and Draga Zec. 1990. *The phonology-syntax connection.* Chicago: University of Chicago Press.

Inkelas, Sharon and Draga Zec. 1995. Syntax-phonology interface. In J. Goldsmith (ed.) *The Handbook of Phonological Theory*. Cambridge, MA: Blackwell, pp. 535-549.

Kaisse, Ellen. 1985. *Connected speech: the interaction of syntax and phonology.* San Diego: Academic Press.

Klausenberger, Jürgen. 1984. *French liaison and linguistic theory.* Stuttgart: Franz Steiner Verlag Wiesbaden GMBH.

Krug, Manfred. 1998. String frequency: a cognitive motivating factor in coalescence, language processing and linguistic change. *Journal of English Linguistics* 26.286-320.

Langacker, Ronald. 1987. *Foundations of Cognitive Grammar, Vol. 1. Theoretical Prerequisites.* Stanford: Stanford University Press.

Morin, Yves-Charles and Jonathan D. Kaye. 1982. The syntactic bases for French liaison. *Journal of Linguistics* 18.291-330.

Nespor, Marina and Irene Vogel. 1986. *Prosodic Phonology.* Dordrecht: Foris.

Nordquist, Dawn. 1999. A synchronic study of *have to* and *got to* with diachronic implications. Paper presented at the Second Annual High Desert Linguistic Society Student Conference, University of New Mexico.

Post, Brechtje. 2000. *Tonal and phrasal structures in French intonation.* The Hague: Thesus.

Saffran, J. R, E. L. Newport and R. N. Aslin. 1996. Word segmentation: the role of distributional cues. *Journal of memory and language* 35.606-621.

Selkirk, Lisa. 1974. French liaison and the X -notation. *Linguistic Inquiry* 5.573-90.

Tranel, Bernard. 1981. *Concreteness in generative phonology. Evidence from French.* Berkeley and Los Angeles: University of California Press.

Vogel, Irene and István Kenesei. 1990. Syntax and semantics in phonology. In Inkelas and Zec (eds.), pp. 339-363.