

Phonological change and grammaticalization in HPSG: the case of French final consonants

Jesse Tseng

Cognition Langues Langage Ergonomie UMR 5263

CNRS & University of Toulouse, France

Jesse.Tseng@univ-tlse2.fr

This paper explores the use of HPSG for modeling historical phonological change, focusing on the evolution of word-final consonants in French. The stages of development are presented as a series of overlapping grammars that show how the loss of final consonants, originally a phonological development in Middle French, gave rise to the grammaticalized external sandhi phenomenon known as consonant liaison in modern French.

A striking aspect of French orthography is the prevalence of silent letters, particularly at the ends of words. Final written consonants can be very roughly divided as follows (leaving aside unassimilated loanwords):

- (1) a. usually silent: s, x, t, d, g, p, b
- b. usually pronounced: c, f, l, r
- c. nasals: n, m (generally silent but nasalize preceding vowel)

Silent letters are an indication of the conservative nature of French spelling, but silent final consonants are retained also because there are synchronic grammatical processes that appear to reactivate the corresponding sounds.

Inflectional morphology and the phenomenon known as “consonant liaison” provide many examples of $\emptyset \sim C$ alternations:

- (2) a. grand [gʁɑ̃] ‘big.*msg*’ vs. grande [gʁɑ̃d] ‘big.*fsg*’
- b. tout nu [tuny] ‘totally naked’ vs. tout autre [tutotʁ] ‘totally different’

In (2a), the written final “d” is silent in the masculine form of the adjective, but it corresponds to a pronounced [d] in the feminine. In (2b), the adverb *tout* is pronounced [tu] before consonant-initial *nu*, but the final “t” is pronounced as a liaison consonant ([tut]) before vowel-initial *autre*.

The synchronic analysis of such phenomena has been the subject of active debate, much of which is centered on the question of abstractness in phonological representation. The data in (2) can be dealt with by assuming underlying phonemic forms containing a final consonant, which must be deleted to derive the appropriate “truncated” forms. The alternations are not always as straightforward as in these examples, however, and more concrete approaches assuming representations closer to the surface forms prove to be more adequate. The analyses proposed in the HPSG literature naturally tend to be surface-oriented (although capturing empirical generalizations always involves some degree of abstractness); see for example Bonami *et al.* (2004) for consonant liaison and Bonami and Boyé (2003) for prenominal adjectives.

Analyses of French phonology in the generative tradition are often deeply inspired by knowledge of diachronic development. A good example of this kind of approach is Casagrande (1984), who proposes highly abstract underlying forms that are basically Latin words, and then a series of ordered rules formalizing the historical sound changes that eventually produced the modern French forms. For a psychologically plausible synchronic analysis, this is obviously taking things too far, but it is nonetheless useful to ask to what extent and how diachronic evidence should be taken into account. And of course the description and analysis of the historical development can be an objective in its own right.

The hypothesis adopted in this paper is that while the grammar of a language can change radically within the space of a few generations, this global change is the sum of smaller, individual changes

that can be modeled in terms of successive, overlapping alternative grammars corresponding to periods of variation (i.e. coexisting, competing analyses) eventually leading to reanalysis (cf. the approach to grammaticalization of Harris and Campbell 1995).

1 Middle French

The general evolution of final consonants in French is relatively well understood (Bourciez and Bourciez, 1967; Zink, 1986). Certain classes of consonants were lost before the end of the Old French period (up to the 13th century); these consonants—primarily nominal case endings and 3rd person verb endings—left no trace in the orthography and have no real relevance in later stages. A more significant development during this period is the loss of most final vowels (other than *-a*), which exposed a large number of new final consonants going into Middle French (“MidFr”, 14th–16th centuries).

As mentioned above, French spelling is highly conservative, and the presence of a final consonant in a modern written form is a good indication that the word had a pronounced final consonant in the 13th century (with isolated exceptions, like *et* ‘and’). The precise identity of the grapheme may not have remained stable (for example, “z” is often replaced by “s”, “s” by “x”, and “t” by “d”) but final consonant letters are rarely eliminated altogether, even if the corresponding sounds subsequently disappeared from the pronunciation. This means, unfortunately, that the abundant texts of the Middle French period give us very little direct evidence for the sound changes under consideration here. Indirect evidence can be gleaned from rhymes in poetry, but this highly specific context does not reveal changes in the pronunciation of words in connected speech.¹

Starting in the 16th century, however, the first published grammars of French provide explicit descriptions of actual usage as well as normative recommendations (which become increasingly influential in the following centuries). There are also contemporary proposals for radically reforming French orthography and replacing it with a more transparently phonetic system (e.g. Jacques Peletier in 1550 and Giles Vaudelin in 1713).

From these various sources, it can be established that in MidFr a weakening process affected all final consonants, to varying degrees, subject to the following conditions (Thurot, 1883):

- (3) a. Final consonants were lost before a following consonant-initial word.
- b. Final consonants were preserved before a following vowel-initial word.
- c. Final consonants were preserved before a pause.

The first two contextual conditions are of course the origin of the phenomenon of liaison in modern French. However, the alternation is widely assumed to have started out as a purely phonological process of final consonant deletion in pre-C environments. Thus, while the examples provided by the early grammarians already exhibit many aspects of modern French pronunciation, we also observe some striking differences.

- (4) a. *mortel peril* = *morteperil* (Palsgrave) vs. ModFr [mɔʁtɛlɾɛvɛil] ‘mortal danger’
- b. *quel monstre* = *qué’ monstre* (Duez) vs. ModFr [kɛlmɔ̃stʁ] ‘what a monster’
- (5) *Vous me dites tousiours que vostre pays est plus grand de beaucoup et plus abundant que le nostre* = *Vous me dite touiours que votre pays est plu gran de beaucoup et plus abandon que le nostre* ‘You always tell me that your country is much larger and more abundant than ours’ (H. Estienne) vs. ModFr *toujours* [tuʒuʁ], *pays* [peɪ], *beaucoup* [boku]

This change has been explained as the result of a weakening of the boundaries between words, such that a -C#C- boundary came to be treated like a medial consonant cluster (and thus subject to reduction), while in a -C#V- sequence, the consonant could be resyllabified as the onset of the following syllable. Unfortunately there is insufficient evidence from the relevant period to reveal the precise progression

¹The presence or absence of a final consonant does not affect the number of syllables, so the scansion of lines of poetry is not affected.

of this evolution. For example, we might expect some final consonants to have been preserved for a time before words starting with “r” or “l” (since consonant-liquid sequences such as “tr-” and “pl-” are possible syllable onsets). Also, while it is unsurprising for the final “s” in -Vs#V- contexts to be pronounced as [z] (like a word-internal, intervocalic “s”), we might expect it to remain unvoiced at first in -Cs#V- contexts, since this has always been the pronunciation of medial “s” in most words like *verser* and *penser*. These hypothetical early steps cannot be confirmed by available data.

Therefore, instead of speculating about and attempting to model the original mechanisms of the change at the level of phrasal phonology, we focus on a later stage at which the alternation is somewhat grammaticalized, so that the alternating forms and the contextual conditions listed in (3) are encoded directly in the grammar. A number of formal approaches are possible in HPSG. We adopt a variant of the PHON-CONTEXT model of Asudeh and Klein (2002), which allows the phonological realization of each daughter within a syntactic construction to vary as a function of the grammatical properties of the immediately following daughter.

- (6) P-CONTEXT model (simplified and modified)

construction →

$$\left[\text{DTRS} \left\langle \left[\text{PHON} \mid \text{P-CTXT } \boxed{1} \right], \boxed{1} \left[\text{PHON} \mid \text{P-CTXT } \boxed{2} \right], \dots, \boxed{n} \left[\text{PHON} \mid \text{P-CTXT } p\text{-ctxt} \right] \right\rangle \right]$$

To model the realization of final consonants, the words in question are assigned lexical entries with complex PHON specifications:

$$(7) \left[\text{PHON} \left[\begin{array}{l} \text{SEGS } \boxed{1} \\ \text{P-CTXT } \textit{nil} \end{array} \right] \vee \left[\begin{array}{l} \text{SEGS } \textit{s-z}(\boxed{1}) \\ \text{P-CTXT } \left[\text{SEGS } \langle \textit{vow}, \dots \rangle \right] \end{array} \right] \vee \left[\begin{array}{l} \text{SEGS } \textit{trunc}(\boxed{1}) \\ \text{P-CTXT } \left[\text{SEGS } \langle \textit{cons}, \dots \rangle \right] \end{array} \right] \right]$$

In this analysis, the pre-pausal form (null P-CTXT) is taken as the basic form (corresponding to the historically original form), the pre-V form is identical except that final [s] undergoes voicing,² and the pre-C form is derived by truncation of the final consonant. The adverb *plus*, for instance, has the three realizations [plys], [plyz], and [ply], respectively, while *tout* has the full form [tut] and the truncated form [tu].

While the lexical entry in (7) is appropriate for many words in Middle French, it is an idealization. The transition from the earlier Old French stage (where roughly speaking each word had a single, context-independent pronunciation) involved a long period during which competing forms co-existed, with diverse factors determining their variation. By the time of the first grammars (16th century), some words (or entire classes of words) had not yet reached the stage represented by (7), while others had already undergone further evolution. This partly explains the distribution sketched in (1) above.

2 Modern French

In the transition to modern French (“ModFr”, 17th century onwards), two major changes must be accounted for. First, the basic or citation form of alternating words is no longer the form with the final consonant, but the “truncated” form. And second, the liaison alternation becomes highly grammaticalized (losing its purely phonological character and becoming subject to lexical and syntactic constraints).

It is evident from contemporary descriptions that by the end of the MidFr period, condition (3c) was no longer systematically respected. The explanation may be phonological: the weakening of final consonants that first led to truncation before #C- expanded (for whatever reason) to the environment ____]PhonP. It may be analogical: truncated forms were more frequent than non-truncated forms in phrase-internal contexts. And perhaps most significantly, the prevalence of truncated forms diminished the status of the final consonant, which was no longer part of the core phonological content of the word. It became unnecessary for determining lexical identity, so much so that for a large number of words (nouns in

²Final [f] probably also became [v], but other voiceless obstruents were not affected.

particular) the final consonant disappeared entirely in all contexts (8a). In other cases, conservative normative influences succeeded in reversing the phonological evolution, and the final consonant was restored in all positions (8b).

- (8) a. *clef* [kle] ‘key’, *outil* [uti] ‘tool’, *coup* [ku] ‘blow’
 b. *chef* [ʃɛf] ‘leader’, *net* [nɛt] ‘net’, *dormir* [dɔʁmir] ‘sleep’

In all such cases, the result is a maximal simplification of the MidFr lexical entry in (7), with the neutralization of all contextual distinctions.

For the words that continue to exhibit a liaison alternation in ModFr, the evolution of (7) can be described as follows (leaving aside the voicing alternations encoded by the function S-Z):

- (9) a.
$$\left[\begin{array}{c} \text{PHON} \\ \left[\begin{array}{c} \text{SEGS } \boxed{1} \vee \boxed{2} \\ \text{P-CTXT } \textit{nil} \end{array} \right] \vee \left[\begin{array}{c} \text{SEGS } \boxed{1} \\ \text{P-CTXT } \left[\text{SEGS } \langle \textit{vow}, \dots \rangle \right] \end{array} \right] \vee \left[\begin{array}{c} \text{SEGS } \boxed{2} \text{ trunc}(\boxed{1}) \\ \text{P-CTXT } \left[\text{SEGS } \langle \textit{cons}, \dots \rangle \right] \end{array} \right] \end{array} \right]$$

 b.
$$\rightsquigarrow \left[\begin{array}{c} \text{PHON} \\ \left[\begin{array}{c} \text{SEGS } \boxed{2} \\ \text{P-CTXT } \textit{nil} \vee \left[\text{SEGS } \langle \textit{cons}, \dots \rangle \right] \end{array} \right] \vee \left[\begin{array}{c} \text{SEGS } \textit{liaison}(\boxed{2}) \\ \text{P-CTXT } \left[\text{SEGS } \langle \textit{vow}, \dots \rangle \right] \end{array} \right] \end{array} \right]$$

Variation is introduced in the pre-pausal context—the leftmost disjunct in (9a). This form eventually aligns with the pre-C form 2, and this historically truncated form acquires the status of basic form, as explained above. The pre-V form $\boxed{1}$ is consequently reinterpreted as a derived form, as indicated in (9b).

The function *liaison* cannot represent a simple phonological process. The relation between liaison forms and non-liaison forms is grammaticalized in the form of a two-slot paradigm, which is used in the analysis of all manifestations of liaison in ModFr, including those that have historical origins other than the final consonant deletion described in the previous section. The slots of the paradigm can be filled in in several different ways. In all of the examples considered up to now, the liaison form is derived from the non-liaison form by the addition of an extra final consonant. This “latent” consonant can correspond to an unpredictable (historical) root consonant (10a), or it can be systematically associated with the grammatical features of the word (b). The liaison form can be suppletive (10c,d), or it can be defective (e). And finally, words that show no liaison alternation in ModFr, for example as in (8) above, simply have identical forms in both slots of their paradigm (f).

(10)	non-liaison form	liaison form	
a.	boku	bokup	<i>beaucoup</i> ‘a lot’
b.	pəti	pətiz	<i>petits</i> ‘small.pl’
c.	sə	sɛt	<i>ce / cet</i> ‘this’
d.	nuvo	nuvɛl	<i>nouveau / nouvel</i> ‘new’ (prenominal)
e.	fʁã	*	<i>franc</i> ‘frank’ (prenominal)
f.	ku	ku	<i>coup</i> ‘blow’ / <i>cou</i> ‘neck’ / <i>coût</i> ‘cost’

The increasing grammaticalization of the liaison alternation in the transition to ModFr also motivates a move away from the P-CTXT model assumed in (7) and (9). P-CTXT is appropriate for sandhi phenomena that are primarily phonologically conditioned, because it gives a word direct access to the PHON values of its neighbors. The conditions on liaison in ModFr are no longer primarily phonological. We assume that consonant-initial words in MidFr became associated with an abstract feature [–LIAISON-TRIGGER], encoding the fact that they could not license the appearance of a liaison form. The switch to a non-phonological feature is crucial for the class of “aspirated *h*” words, which lost their initial consonant in early ModFr period (e.g. *hache* ‘axe’: MidFr [haʃœ] \rightsquigarrow ModFr [aʃ]), but still fail to trigger liaison, despite being vowel-initial phonologically.

- (11)
$$\left[\begin{array}{c} \text{PHON} \\ \left[\text{SEGS } \langle \textit{h}, \dots \rangle \right] \end{array} \right] \rightsquigarrow \left[\begin{array}{c} \text{PHON} \\ \left[\text{SEGS } \langle \textit{vow}, \dots \rangle \right] \end{array} \right]$$

$$\left[\begin{array}{c} \text{LTRIG} \\ - \end{array} \right] \rightsquigarrow \left[\begin{array}{c} \text{LTRIG} \\ - \end{array} \right]$$

The constraints on liaison in ModFr refer to the value of the lexically-specified feature [\pm LTRIG], instead of directly inspecting the SEGMENTS list of the licensing word.

The strict association between liaison forms and liaison contexts expressed in (7) and (9) must also be relaxed, because in many syntactic environments in ModFr, liaison is optional. The only general constraint is that a liaison form must be immediately followed by a [+LTRIG] word:

- (12) a. *beaucoup aimer* [bokueme] / [bokupeme] ‘like a lot’
b. *beaucoup manger* [bokumãʒe] / *[bokupmãʒe] ‘eat a lot’

It follows that liaison forms cannot appear in isolation or before a pause.

It should be mentioned, finally, that a small group of alternating words in ModFr have a pre-pausal form with a pronounced final consonant:

- (13) a. *huit femmes* [ɥifam] ‘8 women’, *huit hommes* [ɥitɔm] ‘8 men’, *huit* [ɥit] ‘8’
b. *dix femmes* [difam] ‘10 women’, *dix hommes* [dizɔm] ‘10 men’, *dix* [dis] ‘10’

These can be considered to be remnants of the MidFr system and we can assume that their lexical entries incorporate contextual constraints that are exceptional with respect to the dominant system. In other words, the behavior of these words does not motivate the extension of the liaison paradigm to three slots for all ModFr words, or the modification of the general constraint illustrated in (12). These words exhibit a good deal of instability, in part as a result of pressure from the more prevalent pattern, but we cannot conclude that they constitute a completely non-productive class. The adverb *plus* ‘more’, for instance, seems to have re-developed the pre-pausal form [plys] in the latter half of the 20th century. This and other cases of “resurrected” final consonants call for further investigation into the interaction between the competing, partially overlapping analyses in Modern French.

References

- Asudeh, A. and Klein, E. (2002). Shape conditions and phonological context. In F. van Eynde, L. Hellan, and D. Beermann, editors, *Proceedings of the 8th International HPSG Conference*, pages 20–30, Stanford, CA. CSLI Publications.
- Bonami, O. and Boyé, G. (2003). La nature morphologique des allomorphies conditionnées. Les formes de liaison des adjectifs en français. In B. Fradin, G. Dal, F. Kerleroux, N. Hathout, M. Plénat, and M. Roché, editors, *Les Unités morphologiques / Morphological Units*. Silex, Villeneuve d’Ascq.
- Bonami, O., Boyé, G., and Tseng, J. (2004). An integrated approach to French liaison. In G. Jäger, P. Monachesi, G. Penn, and S. Wintner, editors, *Proceedings of Formal Grammar 2004*, pages 29–45. Loria, Nancy.
- Bourciez, É. and Bourciez, J. (1967). *Phonétique française. Étude historique*. Klincksieck, Paris.
- Casagrande, J. (1984). *The Sound System of French*. Georgetown University Press, Washington, DC.
- Harris, A. C. and Campbell, L. (1995). *Historical syntax in cross-linguistic perspective*. Cambridge University Press, Cambridge.
- Thurot, C. (1881–1883). *De la prononciation française depuis le commencement du XVIe siècle d’après les témoignages des grammairiens*. Imprimerie nationale, Paris.
- Zink, G. (1986). *Phonétique historique du français*. Presses Universitaires de France, Paris, 6th (1999) edition.