Some listener-oriented accounts of hache aspiré in French

Paul Boersma, University of Amsterdam

April 13, 2005

Abstract

This article shows that the usual speaker-based account of *h*-aspiré in French can explain at most three of the four phonological processes in which it is involved, whereas a listener-oriented account can explain all of them. On a descriptive level, the behaviour of *h*-aspiré is accounted for with a grammar model that involves a *control loop*, whose crucial ingredient is *listener-oriented faithfulness constraints*. These constraints evaluate phonological recoverability, which is the extent to which the speaker thinks the listener will be able to recover the phonological message. On a more reductionist level, however, the pronunciation of *h*-aspiré and its variation is accounted for with a new, very simple, grammar model that can be called *phonology and phonetics in parallel*. This model uses a single constraint set for all four processes involved in bidirectional three-level phonology, namely perception, recognition, phonological production, and phonetic implementation. In this model, the phenomenon of phonological recoverability is not built in, as in control-loop grammars, but emerges from the interaction of four equally simple learning algorithms.

This paper is about the pronunciation of the French form *une hausse*. It has often been observed that the first *e* tends to be pronounced, but no satisfactory explanation has yet been given. Section 1 gives the French facts, points out the descriptive inadequacy of derivational approaches, and gives a new informal listener-oriented account for why the pronunciation [ynəos] is better than both [ynos] and [ynʔos]. Section 2 points out the descriptive inadequacy of speaker-based approaches in Optimality Theory (OT), and shows that the simplest form of listener-oriented OT already accounts for the preference of [ynəos] over [ynos]. Section 3 then formalizes the listener's behaviour, so that section 4 can formalize the speaker's listener-orientedness in more detail, finally leading to an accurate account for the preference of [ynəos] over [ynʔos] and for the occasional occurrence of a phonetically enhanced fourth form, namely [ynəʔos].

1. Hache aspiré: facts and explanations

This section presents the facts about *hache aspiré*, shows that a derivational analysis suffers from conspiracies, and proposes an informal explanation in listener-oriented terms.

1.1. Three types of word-initial segments in French

French words can start with any of three segment types: consonants, vowels, and a third type called *hache aspiré*. The types are listed in (1), each time with two examples, namely a masculine and a feminine noun.

(1) *Three types of word starts in French*

Consonant-initial:
|gaʁsɔ̃| ‘boy’, |fam| ‘woman’
Hache-aspiré-initial:
|ʔazaʁ| ‘coincidence’, |ʔos| ‘rise’
Vowel-initial:
|ɔm| ‘man’, |ide| ‘idea’

In (1) I have provisionally denoted *h*-aspiré with the IPA symbol for a glottal stop (ʔ); in §1.2 and §2.2 I discuss a variety of alternative representations. I use pipes (|) to enclose underlying representations.

The triple distinction in (1) is made on the ground that *h*-aspiré words sometimes act as if they start with a consonant, sometimes as if they start with a vowel, and sometimes in a way different from both consonant-initial and vowel-initial words. This third case will be seen to be the case for which a listener-oriented account becomes inevitable.

The case in which *h*-aspiré acts in the same way as a vowel is the case of phrase-initial neutralization. At the beginning of a phrase, both vowel-initial and *h*-aspiré-initial words are pronounced with a weak attack or with a glottal stop (Dell 1973:79). In isolation, therefore, we have the pronunciations [(ʔ)azaʁ], [(ʔ)os], [(ʔ)ɔm], [(ʔ)ide], and the underlying contrast between *h*-aspiré and a vowel is completely neutralized. There are four phonological processes, however, in which *h*-aspiré-initial words act differently from vowel-initial words. Each of the following four sections discusses one of these cases.

1.2. Process 1: elision

The first process to consider is *elision*, the case of a final vowel that deletes if the next word starts with a vowel. This happens to the final vowels |ə̲| and |a̲| in the singular of the definite article, whose underlying forms can be written as |lə̲| ‘the-MascSg’ and |la̲| ‘the-FemSg’. Thus, |lə̲#ɔm| ‘the man’ is pronounced [lɔm], and |la̲#ide| ‘the idea’ is pronounced [lide]. To indicate that these vowels have to be marked as deletable in the lexicon, I simply underline them in underlying forms (the “#” denotes a phrase-internal word boundary).

Elision does not apply before consonant-initial words, so |lə̲#gaʁsɔ̃| ‘the boy’ is pronounced [ləgaʁsɔ̃], and |la̲#fam| ‘the woman’ is pronounced [lafam]. Crucially, *h*-aspiré-initial words act like consonant-initial words: |lə̲#ʔazaʁ| ‘the coincidence’ is pronounced [ləazaʁ], and |la̲#ʔos| ‘the rise’ is pronounced [laos].

The fact that *h*-aspiré acts like a consonant when it comes to elision has led many researchers to represent it as a consonant underlyingly. Chao (1934 [1958:46]) proposes |h̲| (including the underlining to indicate its deletability); for Schane (1968:8) and Selkirk & Vergnaud (1973) it is |h|. For reasons discussed below in §1.3, Dell (1973:256) argues explicitly for |ʔ|, and Meisenburg & Gabriel (2004) for |ʔ̲|. Some proposed representations are severely underspecified: for Bally (1944:164) *h*-aspiré is a ‘zero consonant’, for Hyman (1985) it is a consonant specified for the feature [+consonantal] alone, and for Prunet (1986) it is a consonant without any features. For denoting *h*-aspiré underlyingly I follow Dell in using the rather arbitrary |ʔ|, at the same time noting that the precise underlying representation does not matter for most of my purposes.

Accounts of the behaviour of *h*-aspiré words in terms of serial rule ordering have profited from the consonantal representation. A typical derivational analysis is shown in (2).

(2) *Derivation of elision*

lə̲#gaʁsɔ̃	elision	ləgaʁsɔ̃	*ʔ/V_V	ləgaʁsɔ̃
lə̲#ʔazaʁ	→	ləʔazaʁ	→	ləazaʁ
lə̲#ɔm		lɔm		lɔm

In this scenario, elision is formalized as "delete V̲ (i.e. any expendable vowel) before another vowel". The reason why *h*-aspiré can block elision is that it is still a consonant at the stage where elision tries to apply. After this, the underlying consonant has to be deleted; this rule is abbreviated here as "*ʔ/V_V", making explicit that it only applies intervocalically (the reason for this restriction will become clear in §1.3). The rule order in (2) was proposed by Schane (1968:7), Selkirk & Vergnaud (1973), and Dell (1973:253,257).

Beside the segmental solutions just discussed several authors have tried to represent *h*-aspiré within nonlinear phonology, most often with reference to syllable structure. In order to keep §1 concise, I defer a discussion of these analyses to §2.2; for the points made in §1, it hardly matters whether *h*-aspiré is accounted for in terms of segmental or syllable structure.

1.3. Process 2: enchainment

The second phonological process to consider, *enchainment* (or *enchaînement*), is the phenomenon that a word-final consonant syllabifies into the onset of a following vowel-initial word. This happens, for instance, to the final |l| of the masculine interrogative or exclamatory pronoun |kɛl| and to many masculine singular adjectives. If we enrich phonetic representations with syllable boundaries, we can roughly say that |kɛl#ɔm| 'which man, what a man' is pronounced [.kɛ.lɔm.], while |kɛl#gaʁsɔ̃| 'which boy' becomes [.kɛl.gaʁ.sɔ̃.]. Crucially, *h*-aspiré words again act like consonant-initial words: |kɛl#ʔazaʁ| 'what a coincidence' becomes [.kɛl.ʔa.zaʁ.]. As with the elision case of §1.2, we can give a derivational account, as in (3).[1]

(3) *Derivation of enchainment*

kɛl#gaʁsɔ̃	enchain	.kɛl.gaʁ.sɔ̃.	*ʔ/V_V	.kɛl.gaʁ.sɔ̃.
kɛl#ʔazaʁ	→	.kɛl.ʔa.zaʁ.	→	.kɛl.ʔa.zaʁ.
kɛl#ɔm		.kɛ.lɔm.		.kɛ.lɔm.

In this scenario, enchainment is formalized as "move any final consonant into the onset of the following syllable, if the following word starts with a vowel". H-aspiré will then block enchainment because it is still a consonant at the stage where enchainment tries to apply. The reason for the condition "V_V" in the formulation of glottal stop deletion now becomes clear: the glottal stop is actually pronounced if the environment is C_V, as here.

Published descriptions about the pronunciation of *h*-aspiré vary. Bally (1944:164), Fouché (1959:251), and Grammont (1948:124) simply state that *h*-aspiré is generally inaudible, although Bally and Fouché limit themselves to discussing the elision and liaison cases of §1.2 and §1.4 and Grammont limits himself to the schwa drop case of §1.5, so it is not clear that these authors had the present postconsonantal case in mind when expressing their

[1] The order of the two rules is immaterial. However, enchainment does not always stand by itself as here: the processes of elision (§1.2) and liaison (§1.4), both of which have to be ordered before glottal stop deletion, usually entail enchainment as well. By ordering enchainment first in (3), we can regard it as a single rule that also plays a role in the elision and liaison processes.

generalization. According to Hall (1948:10), “The phoneme /ʻ/ involves slight faucal constriction, with renewed syllable onset and optional glottal stop.” Since a faucal constriction can be associated with glottal creak, Hall’s description suggests the phonetic transcriptions [kɛl˷azaʁ] and [kɛl˷ˀazaʁ]. According to Meisenburg & Gabriel’s (2004) descriptions based on an acoustic analysis of a corpus of indirectly elicited speech, there can indeed be a plain pause ([kɛl_azaʁ]), a pause with creak ([kɛl˷azaʁ]), or either of these with a glottal stop ([kɛl_ˀazaʁ], [kɛl˷ˀazaʁ]), or there could just be creak in the following vowel ([kɛla̰zaʁ]). Following Meisenburg & Gabriel, I will summarize these five types of phonetic realizations (and all their mixtures) under the umbrella transcription [kɛlʔ̰azaʀ].[2] For Dell (1973:256) as well, the best phonetic transcription is [ʔ].

For the purposes of §1 and the speaker-based view of §2.1, I will simply write the surface forms as “kɛlɔm” versus “kɛlʔazaʁ”, ignoring syllable structure and equating the phonetic detail ([ʔ̰]) with the abstract underlying representation (|ʔ|).

1.4. Process 3: liaison

The third phonological process to consider is *liaison*, the phenomenon that final consonants that are lexically marked as expendable (or ‘latent’) only show up at the surface if the next word starts with a vowel, which allows the consonant to become the onset of the next syllable. This happens to final consonants in inflected verb forms, in preposed adjectives, in plurals of nouns and adjectives, and in several adverbs. For the purposes of this paper I will consider the plural form of the definite article, which is underlyingly |lez̲|, where (following Chao 1934 [1958:46]) the underlining again marks expendability. Thus, |lez̲#ɔmz̲| ‘the men’ and |lez̲#idez̲| ‘the ideas’ are pronounced [lezɔm] and [lezide], while |lez̲#gaʁsɔ̃z̲| ‘the boys’ and |lez̲#famz̲| ‘the women’ are pronounced [legaʁsɔ̃] and [lefam]. Crucially again, *h*-aspiré words act like consonant-intial words: |lez̲#ʔazaʁz̲| ‘the perils’ and |lez̲#ʔosz̲| ‘the rises’ are pronounced [leazaʁ] and [leos] (for references, see all the works cited above in §1.2). As with elision, we can give a derivational account by ordering liaison before glottal stop deletion, as in (4).

(4) *Derivation of liaison*

lez̲#gaʁsɔ̃		legaʁsɔ̃		legaʁsɔ̃
lez̲#ʔazaʁ	liaison →	leʔazaʁ	*ʔ/V_V →	leazaʁ
lez̲#ɔm		lezɔm		lezɔm

In this scenario, liaison is formalized as “delete C̲ (i.e. any expendable consonant) if followed by a consonant”. H-aspiré will then block liaison because it is still a consonant at the stage where liaison tries to apply. Within the sequential rule framework, the order of application in (4) was proposed by Schane (1968:7) and Dell (1973:253,257).

As is the case for *h*-aspiré, the expendable liaison consonant has enjoyed several representations. For Clements & Keyser (1983) it is a consonant lexically marked as extrasyllabic, and for Hyman (1985) it is underlyingly a consonant without a skeletal slot. These matters are hardly relevant for the purposes of the present paper, so I will stay with Chao’s underlining, analogously to the representation for expendable vowels.

2 For enchained realizations, see §4.6.

1.5. Process 4: schwa drop

The final, and most intriguing, phonological process to consider is the drop of schwa in many positions. The position that is relevant to the *h*-aspiré phenomenon is the word-final schwa of feminine function words and adjectives, which is deleted before a consonant as well as before a vowel. For the current paper I will consider the final schwa in the feminine indefinite article |ynə̲|. Thus, |ynə̲#fam| 'a woman' is usually pronounced [ynfam] (sometimes [ynəfam]), and |ynə̲#ide| 'an idea' is always pronounced [ynide]. The degree of obligatoriness is different in the two positions. The prevocalic case is handled by the rule of elision (§1.2), which is fully obligatory within a phrase (Dell 1973:203). The preconsonantal case is subject to a complicated array of conditions involving the number of preceding and following consonants, the nature of those consonants, the presence of word and phrase boundaries, and variation between regions, speakers, and styles (Grammont 1948, chapter "L'*e* caduc"; Dell 1973:221–260). The pronunciation [ynəfam] can probably be considered marginal, but not out of the question. This suggests that we have in effect two different rules, the obligatory elision, or "*ə/_V", and the less-than-obligatory (preconsonantal) schwa drop, or "*ə/VC_C".

It is crucial now that with respect to the schwa drop rule *h*-aspiré words act neither like consonant-initial nor like vowel-initial words: |ynə̲#ʔos| 'a rise' retains its schwa at the surface and is pronounced [ynəos]. Not all authors mention this phenomenon: Bally (1944:323) only mentions the lack of enchainment, thus suggesting a pronunciation like [ynʔ̲os]. Other authors do mention the phenomenon: "tandis que l'on dit « un(e) tache » sans *e*, on prononce l'*e* de *une* dans « une hache »" (Grammont 1948:124). Dell (1973) addresses it repeatedly (pp. 186, 189, 224, 253, 257), even suggesting that the emergence of schwa before *h*-aspiré could be used as a test for discovering whether schwa is underlyingly absent or present. An example of a minimal pair will illustrate this: since 'seven rises' is pronounced [sɛtʔ̲os], its underlying form must be |sɛt#ʔos| (cf. §1.3), and since 'that rise' is pronounced [sɛtəos], its underlying form must be |sɛtə̲#ʔos|.[3]

It is possible that for some speakers [ynʔ̲os] is the only option, perhaps because for them the underlying form of the article is simply |yn|. For these speakers, the triple distinction noted in this section will still exist for words with an uncontroversial underlying schwa, such as the article |lə̲|, if a vowel-final word precedes it within the phrase. Thus |vwasi#lə#ɔm| 'here is the man' will be obligatorily pronounced [vwasilɔm] as a result of elision, |vwasi#lə#gaʁsɔ̃| 'here is the boy' will be usually pronounced [vwasilgaʁsɔ̃] as a result of preconsonantal schwa drop, and, crucially, |vwasi#lə#ʔazaʁ| 'here is the coincidence' will be obligatorily pronounced [vwasiləazaʁ], never *[vwasilʔ̲azaʁ]. Perhaps because of their lower degree of variation, forms like these were discussed instead of [ynəos]-like forms by Tranel (1996) and Tranel & Del Gobbo (2002).

Several formal derivational accounts of French phonology have noticed the failure of schwa drop before *h*-aspiré but not tried to solve it: "there must be some unique initial segment if one is to account for this phenomenon" (Schane 1968:8); "In the absence of reliable phonetic data, we will not propose an account of this phenomenon here" (Clements & Keyser 1983:113). And indeed, if one orders preconsonantal schwa drop (as elision and liaison) before intervocalic glottal stop deletion, |ynə̲#ʔos| will be pronounced [ynʔ̲os],

[3] Tranel (1974:110) did not consider this particular contrast very regular. In general, he argued that underlying final schwa does not exist in French. He changed his mind in later work, perhaps because of the non-occurrence of forms like *[kɛləazaʁ], which Meisenburg & Gabriel (2004) indeed did not observe.

analogously to [kɛlʔ̰azaʁ]. While the pronunciation [ynʔ̰os] does occur (Meisenburg & Gabriel 2004; see also §4.6), the form [ynəos] is usual (it is in fact the only form mentioned by Dell 1973 and Tranel 1995) and has to be explained, especially in the light of the fact that an analogous form like *[kɛləazaʁ] is illegal.

But the derivation can be saved if we realize that the obligatory prevocalic schwa drop is already taken care of by the early rule of elision, so that the less obligatory preconsonantal schwa drop is free to be ordered after glottal stop deletion, as in (5).

(5) *Derivation of schwa drop*

ynə̲#fam ynə̲#ʔos ynə̲#ide	elision →	ynəfam ynəʔos ynide	*ʔ/V_V →	ynəfam ynəos ynide	*ə/VC_C →	ynfam ynəos ynide

The process of schwa drop in (5) can be formalized as "delete schwa before a consonant". The reason why *h*-aspiré does not act like a consonant is that it is no longer a consonant when schwa drop tries to apply. The conditioning of schwa drop by the environment of a following consonant is crucial; without this restriction the outcome of |ynə̲#ʔos| would be [ynos]. The conditioning by a preceding VC has to prevent application to forms like [ləgaʁsɔ̃]. As far as I know, the only analysis that proposed the rules in (5) and their order has been that by Dell (1973:253,257). Analyses within OT have been attempted by Tranel & Del Gobbo (2002) and Meisenburg & Gabriel (2004) and are discussed in §2.2 and §2.3, respectively.

1.6. A derivational account that works but is conspirational

The 12 derivations are summarized in (6).

(6) *Correct derivation of all forms without syllabification*

UF	elision *V̲/_V	liaison *C̲/_C	*ʔ/V_V	*ə/VC_C	SF (see §2.4)
|lə̲#gaʁsɔ̃|	ləgaʁsɔ̃	ləgaʁsɔ̃	ləgaʁsɔ̃	[ləgaʁsɔ̃]	/ləgaʁsɔ̃/
|lə̲#ʔazaʁ|	ləʔazaʁ	ləʔazaʁ	**ləazaʁ**	[ləazaʁ]	/ləʔazaʁ/
|lə̲#ɔm|	lɔm	lɔm	lɔm	[lɔm]	/lɔm/
|kɛl#gaʁsɔ̃|	kɛlgaʁsɔ̃	kɛlgaʁsɔ̃	kɛlgaʁsɔ̃	[kɛlgaʁsɔ̃]	/kɛlgaʁsɔ̃/
|kɛl#ʔazaʁ|	kɛlʔazaʁ	kɛlʔazaʁ	**kɛlʔazaʁ**	[kɛlʔazaʁ]	/kɛlʔazaʁ/
|kɛl#ɔm|	kɛlɔm	kɛlɔm	kɛlɔm	[kɛlɔm]	/kɛlɔm/
|lez̲#gaʁsɔ̃|	lez̲gaʁsɔ̃	legaʁsɔ̃	legaʁsɔ̃	[legaʁsɔ̃]	/legaʁsɔ̃/
|lez̲#ʔazaʁ|	lez̲ʔazaʁ	leʔazaʁ	**leazaʁ**	[leazaʁ]	/leʔazaʁ/
|lez̲#ɔm|	lez̲ɔm	lezɔm	lezɔm	[lezɔm]	/lezɔm/
|ynə̲#fam|	ynəfam	ynəfam	ynəfam	[**ynfam**]	/ynfam/
|ynə̲#ʔos|	ynəʔos	ynəʔos	**ynəos**	[ynəos]	/ynəʔos/
|ynə̲#ide|	ynide	ynide	ynide	[ynide]	/ynide/

This table is a rather severe simplification, ignoring the intricacies of several French schwa drop rules and ignoring the dependence of all the rules in (6) on word and phrase boundaries. The table does account for all phrase-internal phenomena of *h*-aspiré. Going from the underlying form (UF) towards the phonetic form in the fifth column, the early obligatory rules *V̲/_V and *C̲/_C amount to vowel coalescence and consonant coalescence, respectively, and need not be ordered with respect to each other (Schane 1968:4 even uses alpha-notation to collapse them into a single 'truncation' rule), while the later slightly optional rules *ʔ/V_V and *ə/VC_C are mutual bleeders and have to be ordered as here. The forms in the sixth column ('perceived phonological surface forms') are discussed in §2.4.

There are three kinds of problems with the derivational account in (6). First, there is a general problem with the explanatory adequacy of derivational accounts, namely that such accounts do not usually manage to propose a learning algorithm for the detection of rules and their ordering. The second problem concerns observational adequacy and occurs when the attested variation is taken into account. Meisenburg & Gabriel (2004) report the pronunciations [ynəos], [ynəʔ̰os], and [ynʔ̰os], and forms like [ləʔ̰azaʁ] and [leʔ̰azaʁ] can be predicted on the basis of Meisenburg & Gabriel's observed pronunciations [tuʔ̰ɔ̃gʁwa] for |tut̠#ʔɔ̃gʁwa| 'all Hungarian' and [tʁwaʔ̰ɑ̃bœʁgœʁ] for |tʁwaz̠#ʔɑ̃bœʁgœʁ| 'three hamburgers'. This means that the glottal stop deletion rule in (6) is optional. However, this optionality cannot predict the observed form [ynəʔ̰os]: if glottal stop deletion does not apply, the resulting intermediate form "ynəʔos" would be eligible for preconsonantal schwa drop and surface as [ynʔ̰os]. The form [ynəʔ̰os] could only surface if schwa drop were optional as well, but that would predict the occurrence of forms like [ynəfam]. If Meisenburg & Gabriel are correct in suggesting that [ynəʔ̰os] is a more likely event than [ynəfam], the rule set in (6) cannot be correct.

The third and most salient problem with the account in (6) is that it does not ascribe the rules to their causes, thereby failing the criterion of descriptive adequacy if such causes are linguistic. Whereas the elision and liaison rules have some claim to universal naturalness, as would have the unconditional rules *ʔ ("delete glottal stop") and *ə ("delete schwa"), the environmental conditions in the rules *ʔ/V_V and *ə/VC_C change this picture. Why does glottal stop delete only intervocalically, not in a C_V environment? Why does schwa delete only before a consonant, not before a vowel? Especially this latter question is worrisome, since the preconsonantal condition seems to be unnatural crosslinguistically as well as French-internally. Crosslinguistically, a preconsonantal position is a more natural position for a vowel than a prevocalic position is, so one would think that if schwa deletes before a consonant, it will certainly have to delete before a vowel as well (and so it does obligatorily in French in the earlier elision stage). French-internally, the preconsonantal condition is specific to the case of schwa followed by *h*-aspiré; the condition does not apply non-vacuously to any other consonants in French. Together, the two rules *ʔ/V_V and *ə/VC_C seem to *conspire* to keep an unexpected part of the underlying form alive: [ʔ̰] in the case of [kɛlʔ̰azaʁ], [ə] in the case of [ynəos]. Conspirations like these formed the original criticism that led Prince & Smolensky (1993) to propose the constraint-based framework of Optimality Theory. Although this is an old point, therefore, it is relevant to point it out again here, since it affirms that any accounts of *h*-aspiré within an Optimality-Theoretic framework should make exclusive use of universally defensible constraints. In §2.1 to §2.3 I will show that speaker-based Optimality-Theoretic accounts fail to meet this requirement (they contain conspiracies themselves), and in §2.4 and §2.5 I will show that a listener-oriented Optimality-Theoretic account does meet it. But first, the purpose behind the conspiracy has to be identified.

1.7. An informal listener-oriented account: explaining the conspiracy

The goal of the conspiracy between the rules *ʔ/V_V and *ə/VC_C in (6) seems to be the avoidance of neutralization. Without the V_ condition, the *h*-aspiré-initial form |kɛl#ʔazaʁ|

would end up as [kɛlazaʁ], neutralizing with a hypothetical vowel-initial form |kɛl#azaʁ|. Without the _C condition, the *h*-aspiré-initial form |ynə̲#ʔos| would end up as [ynos], neutralizing with a hypothetical vowel-initial form |ynə̲#os|.[4] All forms in (6), then, can be understood in informal terms if we suppose that the speaker takes into account the degree to which the listener will be able to recover the underlying form from the phonetic form. To see this more explicitly, consider the *recoverability* of *h*-aspiré in the seven forms listed in (7).

(7) *Recoverability of h-aspiré*

h-aspiré-initial words:	*vowel-initial words:*	*recoverability (auditory cue):*
[ləazaʁ]	[lɔm]	good (vowel)
?[kɛlazaʁ]	[kɛlɔm]	bad (none)
[kɛlʔ̰azaʁ]	[kɛlɔm]	okayish (creaky pause)
[leazaʁ]	[lezɔm]	good (consonant)
*[ynos]	[ynide]	bad (none)
?[ynʔ̰os]	[ynide]	okayish (creaky pause)
[ynəos]	[ynide]	good (vowel)

Every row in (7) lists the *h*-aspiré-initial word and the vowel-initial word and judges the quality of the recoverability of the contrast between the two on the basis of their auditory difference. The *h*-aspiré in |lə̲#ʔazaʁ|, for instance, can be recovered by the listener on the basis of the schwa in [ləazaʁ]. The hiatus (sequence of two vowels) in [ləazaʁ] is a clear sign of an underlying *h*-aspiré: had there been an underlying initial vowel (|lə̲#azaʁ|), there would have been no schwa ([lazaʁ]), and had there been an underlying initial full consonant (e.g. |lə̲#gazaʁ|), this consonant would have surfaced ([ləgazaʁ]). If we ignore for simplicity the consonant-initial words (which contrast well with both *h*-aspiré-initial and vowel-initial words in all cases discussed here), then the degree of recoverability of *h*-aspiré amounts to the perceptual distincitivity of [ləazaʁ] and [lazaʁ], which table (7) lists as *good* since the difference is a complete segment (a vowel). Similarly, the recoverability of *h*-aspiré in [leazaʁ] is again *good*, since the difference with [lezɔm] is a full consonant, so that listeners can easily infer the underlying *h*-aspiré from the hiatus.

Next consider the two candidates for |kɛl#ʔazaʁ| in (7). If the phonetic form were [kɛlazaʁ], *h*-aspiré would be totally unrecoverable, since that phonetic form is also how a hypothetical |kɛl#azaʁ| would surface. Hence the listing of the [kɛlazaʁ]-[kɛlɔm] contrast as *bad*. The recoverability of *h*-aspiré from [kɛlʔ̰azaʁ], as compared to [kɛlɔm], is better than that from [kɛlazaʁ], since some auditory cues for *h*-aspiré are available. Nevertheless, the recoverability is less good than in the [ləazaʁ] and [leazaʁ] cases, since the auditory difference between *h*-aspiré-initial and vowel-initial words now has to be carried by the creaky pause [ʔ̰], which, as mentioned before, may only consist of a pause, creak, or a glottal stop, or of any combination of these, and these cues are often a subset of those available in a real consonant (see §3.3); if a vowel is as audible as an average consonant, [ʔ̰] must be less audible than a vowel. Hence the listing of the [kɛlʔ̰azaʁ] – [kɛlɔm] contrast as *okayish*.

Finally we arrive at the three candidates for |ynə̲#ʔos| in (7). Had the phonetic form been [ynos], *h*-aspiré would have been totally unrecoverable, since that phonetic form is also how a

[4] The fact that there are not many minimal pairs does not make the neutralization much less bad. In noisy environments, every phonologically contrastive feature must contribute to disambiguation.

hypothetical |ynə̲#os| would surface. Hence the listing of the [ynos]-[ynide] contrast as *bad*. The candidate [ynʔ̰os] does as well as [kɛlʔ̰azaʁ], hence the listing of the recoverability of *h*-aspiré as *okayish*. Much better, though, is the candidate [ynəos], which has a vowel (schwa) before hiatus, from which the listener will have no trouble inferring the underlying *h*-aspiré, entirely analogously to the cases of [ləazaʁ] and [leazaʁ].

In the listener-oriented account, therefore, the surfacing of the underlying glottal stop in [kɛlʔ̰azaʁ] and that of the underlying schwa in [ynəos] are strategies that the speaker can use to establish the recoverability of *h*-aspiré. The reason why speakers do not attempt to further improve [kɛlʔ̰azaʁ] by pronouncing it as [kɛləazaʁ] must lie in the fact that schwa is not underlyingly present in |kɛl#ʔazaʁ|. The generalization seems to be, then, that speakers prefer to pronounce an underlying *h*-aspiré with the auditorily best cues, namely as a hiatus; if there is no underlying material to fill the two vowel positions involved, speakers fall back on the auditorily second-best cues, namely those of a glottal stop.

The two conditions "V_" and "_C" in (6) can thus be explained directly by the speaker's desire to make an underlying contrast emerge at the surface. Within the sequential-rule formalism of (6), however, the two conditions are accidental and the whole explanation for the conspiracy must remain extragrammatical. My use of the terms *candidates*, *second-best*, and even *fill*, strongly suggests that instead an account in terms of Optimality Theory should be pursued. This will be the subject of §2, where we will nevertheless see that the usual speaker-based two-level version of OT encounters many of the same problems as the speaker-based derivational account of (6). Only a listener-oriented three-level version of OT will be seen to do the trick.

2. Formalizations in OT

There are two reasons why the *h*-aspiré case should be handled in OT: the possibility of expressing the phonology in a typologically correct way by solely using universally defensible elements, which the derivational account of §1.6 failed to achieve, and the possibility of comparing multiple candidates, which even the informal listener-oriented account of §1.7 already had to do. In §2.1 to § 2.3 I show that the universal defensibility does not go through in the usual speaker-based version of OT. In §2.4 and §2.5 I show that the cause of this problem is that no constraints in that version of OT are able to evaluate recoverability, and that the problem is solved by using *listener-oriented faithfulness* constraints, which do evaluate recoverability.

2.1. Speaker-based OT account

An OT account requires a number of universally defensible constraints. Since McCarthy & Prince (1995), these constraints should be divided in structural constraints, which evaluate output (surface) structures, and faithfulness constraints, which evaluate the similarity between input (underlying) and output (surface) forms.

For the *h*-aspiré case, I will regard one constraint as so high-ranked that candidates that violate it will not turn up in tableaus. It is the constraint that has to rule out forms like *[lgaʁsɔ̃] and *[lʔ̰azaʁ] phrase-initially. While the rule set by Schane (1968) allows only [lə-] before any consonant, the rules formulated by Grammont (1948:117) and Dell (1973:227) allow for schwa drop after a single phrase-initial consonant, with some restrictions that nevertheless, if taken literally, do not seem to apply to [lgaʁsɔ̃]. However, Tranel (1996) does not consider the (syllabified) form /.lgaʁ.sɔ̃./ worth including in his tableaus, perhaps for reasons of violation of sonority sequencing, and rules out the form /.l.gaʁ.sɔ̃./ by a constraint against non-vocalic syllable nuclei. Since the present paper cannot take into account the many intricate phenomena of French schwa deletion (Dell 1973:221–260), I will ignore forms like these. Phrase-internally, the matter is quite different, since schwa can be deleted more easily there (§1.5); the surfacing of schwa in [vwasiləazaʁ] 'here is the coincidence' will have to be related to the same cause that will have to explain the surfacing of schwa in [ynəos].

The first three constraints that will appear in tableaus handle the surfacing of schwa. An undominated faithfulness constraint has to rule out *[kɛləazaʁ] (cf. §1.3, §1.5).

(8) *Constraint handling the surfacing of schwa*

DEP(ə): "do not pronounce any schwas that are not underlyingly present."

The two cases of schwa deletion in (6), namely prevocalic elision and preconsonantal deletion, are provisionally handled by the constraint ranking *ə >> MAX(V̲), which directly expresses the facts that [lɔm] is much better than *[ləɔm] and [ynfam] is somewhat better than ?[ynəfam].[5]

(9) *Constraints handling the surfacing of schwa*

*ə: "do not pronounce any schwas."
MAX(V̲): "underlying expendable vowels (i.e. schwa and the vowel in |la̲|) must surface."[6]

Next, we need several MAX constraints for underlying consonants. MAX(C) promotes the surfacing of underlying consonants, but must not do so for the final liaison consonants (e.g. the |z̲| in |lez̲|), nor for underlying glottal stop, which, as we have seen, does not always surface. We must therefore distinguish at least three consonantal MAX constraints.

(10) *Consonantal MAX constraints*

MAX(C): "underlyingly non-expendable non-glottal-stop consonants must surface."
MAX(ʔ): "underlying glottal stops must surface."
MAX(C̲): "underlyingly expendable consonants must surface."

To force the deletion of latent consonants before other consonants, i.e. to rule out *[lezgaʁsɔ̃] and *[lezʔazaʁ], there must be a ranking like { MAX(C), MAX(ʔ) } >> *CC >> MAX(C̲), which ensures that consonant clusters will surface faithfully if both consonants are underlyingly non-expendable, but that one of the consonants (in practice always the first) will be deleted if it is expendable.[7]

---

5 The difference in grammaticality between *[ynəide] and ?[ynəfam] is formalized in §2.5.

6 The only non-schwa expendable vowel is that in |la̲|. Tranel (1996) proposes that this vowel is not actually expendable, but that the whole allomorph |la| can be replaced by its masculine counterpart |lə̲|. An underlying form is then |{lə̲$_{\text{MASC}}$,la$_{\text{FEM}}$}#ide$_{\text{FEM}}$| 'the idea', and the /l/ in its surface form /.li.de./ then represents an elided version of |lə̲$_{\text{MASC}}$|. This is plausible, since the same gender changing mechanism is required for handling allomorphies like {mɔ̃n̲$_{\text{MASC}}$, ma$_{\text{FEM}}$} 'my' and {bo$_{\text{MASC}}$, bɛlə̲$_{\text{FEM}}$} 'beautiful', where the selected allomorph before vowel-initial words is always the one that is capable of providing an onset (/.mɔ̃.n-/ with liaison, /.bɛ.l-/ with elision). In this theory, the constraint MAX(V̲) could be replaced with MAX(ə).

7 Other formalizations of liaison are possible. In Tranel's (1996) view, latent consonants lack an underlying prosodic node, perhaps an X-slot (De Jong 1990), that other consonants do possess underlyingly. In order to be able to pronounce the latent consonant, the speaker will have to add such a prosodic node into the surface

(11) *Constraint handling liaison*

*CC: “no consonant clusters.”[8]

In order to force the appearance of forms like [ynfam] rather than [ynəfam], this constraint must be ranked below *ə.

Finally, glottal stop deletion must be handled by *ʔ̰, but since [kɛlʔ̰azaʁ] is better than [kɛlazaʁ], this constraint must be outranked by MAX(ʔ).

(12) *Constraints handling the surfacing of glottal stop*

*ʔ̰: “do not pronounce any creaky pauses.”

The relative ranking of the constraints in (11) and (12) could be determined by the relative harmony of [ynʔ̰os] and [ynəos]. The preference for [ynəos] suggests the ranking *ʔ̰ >> *ə. A tentative complete ranking is given in (13).

(13) *A ranking for handling the French hache-aspiré effects*

{ MAX(C), DEP(ə) } >> MAX(ʔ) >> *ʔ̰ >> *ə >> { MAX(V̲), *CC } >> MAX(C̲)

This ranking consists of constraints that are universally defensible to some extent, i.e. of constraints that simply penalize uncommon structures or militate against deletion of underlying material or against insertion of non-underlying material. Although the ranking will be refined in several ways later, my main point about the empirical difference between the speaker-based and listener-oriented accounts can be entirely illustrated with the ranking in (13). Specifically, I will show that ranking (13) does not suffice for a successful speaker-based account but that it does suffice for a successful listener-oriented account.

The next 12 tableaus show that in the usual OT with two levels (underlying and surface) ranking (13) predicts 9 of the 12 forms in (6) correctly, thus (somewhat unfairly) claiming that a speaker-based account has trouble accounting for 3 of the 12 forms, at least with constraints as simple as those in (13). In all tableaus, the pointing finger (“☞”) indicates the winner according to the grammar, and the check mark (“√”) indicates the correct French form. If the two point at different candidates, this indicates a failure of the speaker-based account.

(14a) *Speaker-based elision, consonant case*

|lə̲#gaʁsɔ̃|	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ̰	*ə	MAX (V̲)	*CC	MAX (C̲)
√☞ ləgaʁsɔ̃					*		*	
ləaʁsɔ̃	*!				*		*	
laʁsɔ̃	*!					*	*	

structure, thus violating a faithfulness constraint that Tranel calls “avoid integrating floaters” (AIF) but could be called DEP(X) (if the lacking node is an X-slot) in the terminology of Correspondence Theory. The pronunciation of latent consonants can then be forced by a higher-ranked ONSET. The formulation in terms of *CC is meant to not preclude the phenomenon of *liaison without enchainment* (Encrevé 1988), in which a latent consonant shows up in coda before a vowel-initial word.

[8] Like the liaison rule “*C̲/_C” in (6), this constraint does not handle the phrase-final drop of latent consonants.

(14b) *Speaker-based elision, h-aspiré case*

|lə̲#ʔazaʁ|	Max (C)	Dep (ə)	Max (ʔ)	*ʔ̰	*ə	Max (V̲)	*CC	Max (C̲)
☞ ləʔ̰azaʁ				*	*			
√ ləazaʁ			*!		*			
lazaʁ			*!			*		

(14c) *Speaker-based elision, vowel case*

|lə̲#ɔm|	Max (C)	Dep (ə)	Max (ʔ)	*ʔ̰	*ə	Max (V̲)	*CC	Max (C̲)
ləɔm					*!			
√☞ lɔm						*		
ləʔ̰ɔm				*!	*			

(14d) *Speaker-based enchainment, consonant case*

|kɛl#gaʁsɔ̃|	Max (C)	Dep (ə)	Max (ʔ)	*ʔ̰	*ə	Max (V̲)	*CC	Max (C̲)
√☞ kɛlgaʁsɔ̃							**	
kɛgaʁsɔ̃	*!						*	
kɛlaʁsɔ̃	*!						*	
kɛləgaʁsɔ̃		*!			*		*	

(14e) *Speaker-based enchainment, h-aspiré case*

|kɛl#ʔazaʁ|	Max (C)	Dep (ə)	Max (ʔ)	*ʔ̰	*ə	Max (V̲)	*CC	Max (C̲)
√☞ kɛlʔ̰azaʁ				*			*	
kɛʔ̰azaʁ	*!			*				
kɛləazaʁ		*!	*		*			
kɛlazaʁ			*!					

(14f) *Speaker-based enchainment, vowel case*

|kɛl#ɔm|	Max (C)	Dep (ə)	Max (ʔ)	*ʔ̰	*ə	Max (V̲)	*CC	Max (C̲)
√☞ kɛlɔm								
kɛɔm	*!							
kɛlʔ̰ɔm				*!			*	

(14g) *Speaker-based liaison, consonant case*

|lez̲#gaʁsɔ̃|	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ̰	*ə	MAX (V̲)	*CC	MAX (C̲)
lezgaʁsɔ̃							**!	
√☞ legaʁsɔ̃							*	*
lezaʁsɔ̃	*!						*	
lezəgaʁsɔ̃		*!			*		*	

(14h) *Speaker-based liaison, h-aspiré case*

|lez̲#ʔazaʁ|	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ̰	*ə	MAX (V̲)	*CC	MAX (C̲)
lezʔ̰azaʁ				*			*!	
☞ leʔ̰azaʁ				*				*
lezazaʁ			*!					
√ leazaʁ			*!					*

(14i) *Speaker-based liaison, vowel case*

|lez̲#ɔm|	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ̰	*ə	MAX (V̲)	*CC	MAX (C̲)
√☞ lezɔm								
leɔm								*!
leʔ̰ɔm				*!				*

(14j) *Speaker-based schwa drop, consonant case*

|ynə̲#fam|	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ̰	*ə	MAX (V̲)	*CC	MAX (C̲)
ynəfam					*!			
√☞ ynfam						*	*	

(14k) *Speaker-based schwa drop, h-aspiré case*

|ynə̲#ʔos|	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ̰	*ə	MAX (V̲)	*CC	MAX (C̲)
ynəʔ̰os				*	*!			
☞ ynʔ̰os				*		*	*	
√ ynəos			*!		*			
ynos			*!			*		

(14l) *Speaker-based schwa drop, vowel case*

|ynə̲#ide|	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ̰	*ə	MAX (V̲)	*CC	MAX (C̲)
ynəide					*!			
√☞ ynide						*		
ynʔ̰ide				*!		*	*	

From these tableaus we see that all proposed constraints are active. In (14i) we see that even the lowest-ranked constraint, that for the surfacing of an underlying latent consonant, can be active.[9] But there remain three failures, enumerated in (15). There is no ranking of the constraints that can give the correct forms in all twelve cases.

(15) *The three failures of the speaker-based account*

a. [ləʔ̰azaʁ] instead of [ləazaʁ].
b. [leʔ̰azaʁ] instead of [leazaʁ].
c. [ynʔ̰os] instead of [ynəos].

The three failures have something in common. All three have hiatus in the correct form. It looks, therefore, as if the constraint ranking could easily be fixed by a constraint that favours hiatus, but there is a problem. The constraint set in (13), which corresponds to the rules in (6), suffers from *opacity* when applied to the French data. In (6) we already saw a counterfeeding relationship between the rules: glottal stop deletion re-creates hiatus structures that elision had just removed. Such a situation of opacity is notoriously difficult for OT to handle in a principled way. An OT account would either require the trick of *local conjunction* or the brute-force method of proposing a specific high-ranked constraint for the exceptional case. Both tricks have been attempted in the literature, and I will discuss them in §2.2 and §2.3, respectively.

2.2 Patch by Tranel & Del Gobbo (2002)

Beside the segmental representations mentioned in §1.2, *h*-aspiré has often been accounted for in terms of prosodic elements. Clements & Keyser (1983) proposed that *h*-aspiré words start with an empty onset whereas vowel-initial words start with no onset. Conversely, De Jong (1990) proposed that *h*-aspiré words start with no onset whereas vowel-initial words start with an empty onset. A simpler and less hierarchical syllable-based solution was proposed by Tranel (1995), who argued that *h*-aspiré constitutes a *syllable island*, i.e. that our example words are represented underlyingly as |.azaʁ|, |.os|, |ɔm|, and |ide|, where |.| stands for a syllable boundary. The phonological surface structures then look like /.kɛ.lɔm./ and /.kɛl.a.zaʁ./. Tranel did not consider at all the phonetic realization of these abstract structures; presumably a separate process of phonetic implementation has to turn the prevocalic syllable boundary in /.kɛl.a.zaʁ./ into a pause, a glottal stop, and/or creak: [kɛl_azaʁ], [kɛl_̰azaʁ], [kɛl_ˀazaʁ], [kɛl_̰ˀazaʁ], or [kɛla̰zaʁ] (§1.3). But deciding that overt phonetic forms need not be discussed fails to capture the observation that the phonetic implementation is linguistically relevant: since there are languages in which creak is a

[9] An equally low ranked ONSET could perform the same function.

contrastive feature of vowels, the realization of a syllable boundary as creak in the following vowel (as in [kɛla̰zaʁ]) cannot be universal and must be specific to French.

Freed of the responsibility to account for phonetic detail, the syllable island approach can be expected to work better than tableaus (14). And indeed, the failures in (15ab) vanish, since both [ləʔazaʁ] and [ləazaʁ] must be considered as reflexes of the same phonological structure /.lə.a.zaʁ./, and the question whether a glottal stop will be pronounced or not remains merely a case of phonetic implementation, apparently outside the scope of Tranel's article. While removing a language-specific phonetic implementation phenomenon can be regarded as evading the problem, the remaining case, that of [ynəos], turns out to be problematic for Tranel's account anyway. In Tranel's view, this form is /.y.nə.os./, whereas the competing candidate [ynʔ̰os] is just /.yn.os./. What, now, makes *h*-aspiré in this form act neither like a consonant nor like a vowel? Tranel (1995: 812) tries to answer this question by proposing that the retention of schwa in /.y.nə.os./ is "a possible strategy for resolving the conflict caused on the one hand by the phonological pressure exerted by forward syllabification in VCV sequences and on the other hand by the syllable-island constraint characteristic of h-aspiré words."[10] In the OT implementation of Tranel (1996), the 'syllable-island constraint' becomes an alignment constraint that militates against /.y.nos./, in which the left edge of /os/ does not coincide with the left edge of a syllable. The 'phonological pressure' that Tranel mentioned in 1995 is the preference of /.y.nos./ over /.yn.os./; in his OT implementation Tranel did not yet dare to formalize it: a form like /.yn.os./ or /.vwa.sil.a.zaʁ./ "is not optimal because it incurs both a NoCoda violation and an Onset violation, which results in the worst possible transsyllabic contact (Clements 1988). But constraints in OT are taken to act independently, rather than in synergy. I leave open here the resolution of this problem." (1996 [1994:19]). The 'synergy' Tranel refers to here is the problematic device of constraint conjunction. Having no choice, Tranel & Del Gobbo (2002) ultimately did formalize the phenomenon with the conjoined constraint Onset&NoCoda, as shown in (16).[11]

(16) *The patch by Tranel and Del Gobbo (2002)*

|ynə̲#os|	Align-L(/os/, σ)	Onset&NoCoda	*ə	Max (V̲)
√☞ .y.nə.os.			*	
.yn.os.		*!		*
.y.nos.	*!			*

The tableau shows that what remains is /.y.nə.os./. Meisenburg & Gabriel (2004) criticize Tranel & Del Gobbo for the constraint Align-L(/os/, σ), because it refers to a specific morpheme. However, I feel that this is not a major problem, since the underlying form can be specified with an underlying syllable boundary, i.e. as |.os|, so that the alignment constraint can be replaced with faithfulness constraints like McCarthy & Prince's (1995) Contiguity and Linearity, i.e. with alignment-as-faithfulness (Boersma 1998:196–199; Horwood 2002). Instead, the real problem in (16) is the constraint Onset&NoCoda. First, proposing

[10] An earlier proposal (Tranel 1987:95) that a form like /.vwa.sil.a.zaʁ./ is ruled out because the /l/ 'belongs' to the morpheme |azaʁ| rather than to |vwasi| cannot be correct, given that /.vwa.sil.gaʁ.sɔ̃./ is grammatical.

[11] The order of the first two constraints is determined by the form /.kɛl.a.zaʁ./, which satisfies alignment but violates Onset&NoCoda.

constraints that are conjoined from two heterogenous simpler constraints already constitutes a complication. Secondly and more crucially, there is the problem of the two surface representations (phonological and phonetic): how can we say that a form that is pronounced [ynʔ̰os] violates ONSET? Hasn't glottal stop insertion been regarded as one of the many ways to *satisfy* ONSET (e.g. McCarthy 2003:100)?

To sum up, two specific problems can be identified in Tranel & Del Gobbo's approach: the introduction of a conjoined constraint, and the failure to address the 'mere' phonetics. To obtain the coverage of the French phenomena that the present paper aims to achieve, Tranel & Del Gobbo would have to extend their model with a serious formalization of the phonetics. The present paper implements precisely such a formalization (§2.4 to §5), and achieves the desired coverage with less constraint conjunction and more explanatory force.

2.3 Patch by Meisenburg & Gabriel (2004)

The account by Meisenburg & Gabriel (2004) has the advantage that it does try to account for the pronunciation of the glottal stop, i.e. it does not try to move linguistically relevant phenomena to an extralinguistic phonetic implementation module as Tranel & Del Gobbo did. H-aspiré is represented underlyingly as a (creaky) glottal stop, and the constraint set is similar to the one I used in §2.1, but with some additional constraints to get rid of the failures in (15).

The constraint that Meisenburg & Gabriel propose to rule out [ləʔ̰azaʁ] and [leʔ̰azaʁ] is reminiscent of the postvocalic glottal stop deletion rule in (6).

(17) *Meisenburg & Gabriel's constraint for glottal stop deletion*

*VʔV: "do not pronounce any intervocalic glottal stops."

While this constraint rules out [ləʔ̰azaʁ] and [leʔ̰azaʁ], it does not yet solve (15a) and (15b). We can see this by noting that if *VʔV is inserted with a high ranking in tableau (14b), the winning candidate will become [lazaʁ]. In Meisenburg & Gabriel's view, the forms *[lazaʁ] and *[lezazaʁ] violate an alignment constraint.

(18) *Meisenburg & Gabriel's alignment constraint (slightly reworded)*

ALIGN-L (|ʔ|, σ): "the left edge of a lexical morpheme starting with an underlying |ʔ| is aligned on the surface with the onset of its first syllable."

The form [lazaʁ] violates this constraint because the left edge of the morpheme |ʔazaʁ|, which starts with a glottal stop, ends up on the surface between [l] and [a], whereas its first syllable is [la]. With the two new constraints, the correct form [ləazaʁ] becomes the winner in (14b). In (14h), [leazaʁ] becomes better than [lezazaʁ], as is shown in (19).

(19) *Meisenburg & Gabriel's speaker-based liaison, h-aspiré case*

|lez̲#ʔazaʁ|	MAX (C)	ALIGN (ʔ)	*VʔV	DEP (ə)	MAX (ʔ)	*ʔ̰	*ə	MAX (V̲)	*CC	MAX (C̲)
☞ lezʔ̰azaʁ						*			*!	
leʔ̰azaʁ			*!			*				*
lezazaʁ		*!			*					
√ leazaʁ					*!					*

There remains a problem, as (19) shows. The winning candidate becomes [lezʔ̰azaʁ], a form that Meisenburg & Gabriel fail to take into account.[12] To rule out this candidate, Meisenburg & Gabriel would need an additional undominated constraint like "*C̲.", i.e. "expendable consonants must not surface in coda".

The addition of three constraints, then, handles problems (15a) and (15b). To handle (15c), additional machinery is called for, since the ranking of (19) would still make [ynʔ̰os] the winner in (14k). Meisenburg & Gabriel propose that the latter candidate is ruled out by the conditional faithfulness constraint in (20).

(20) *Meisenburg & Gabriel's conditional faithfulness constraint (reworded)*

MAX(ə/_ʔ): "an underlying schwa before an underlying glottal stop must surface."

The reason for requiring that the environment of this constraint (the following glottal stop) be considered as 'underlying' rather than at the surface is that this constraint must be regarded as satisfied in [ynəos] but violated in [ynos]. As (21) shows, this final patch saves the speaker-based OT formulation.

(21) *Meisenburg & Gabriel's speaker-based schwa drop, h-aspiré case*

|ynə̲#ʔos|	MAX (C)	ALIGN (ʔ)	MAX (ə/_ʔ)	*VʔV	DEP (ə)	MAX (ʔ)	*ʔ̰	*ə	MAX (V̲)	*CC	MAX (C̲)
ynəʔ̰os				*!			*	*			
ynʔ̰os			*!				*		*	*	
√☞ ynəos						*		*			
ynos		*!	*			*			*		

But saving the OT formulation in this way comes at a great cost, comparable to that of the derivational account in (6). The cost is that several of the proposed constraints are little or not at all universally defensible, like some of the derivational rules in (6).

The constraint *VʔV seems to be little universally defensible, just like the rule *ʔ/V_V in (6). If this constraint punishes the articulation of [ʔ], i.e. if it can be formulated as *[VʔV], it does not seem to be universally defensible: what makes a glottal stop more difficult between vowels than somewhere else? If anywhere, glottal stop would seem to be articulated *most* easily in intervocalic position. If the constraint *VʔV is a structural constraint, i.e. */VʔV/, it does not seem to be universally defensible either, because I know of no reasons why the structure /VʔV/ would be more difficult to maintain as a mental representation than /CʔV/.

The constraint MAX(ə/_ʔ) seems even less universally defensible. It just states the offending exception. Even if all the other constraints were perfect, the need for this constraint would have to lead to a rejection of the whole speaker-based OT account. The point is that the hidden objective of this constraint is to save a schwa even (or especially) if the glottal stop is deleted, i.e. to save the language-specific contrast between *h*-aspiré-initial and vowel-initial

[12] Meisenburg & Gabriel's account differs from (19) on two minor points. They handle liaison with Tranel's (1996) constraint "avoid integrating floaters" (see footnote 7), and they do not discuss anything like *ə >> MAX(V̲) because they do not consider candidates like *[ləɔm] and ?[ynəfam]. When included as an additional candidate in their tableaus, the form [lezʔ̰azaʁ] would win, as in (19).

words: schwa must be saved in [ynəos] because that improves the recoverability of *h*-aspiré (§1.7). In fact, all three constraints proposed by Meisenburg & Gabriel can be understood in the light of this recoverability, as summarized in (22).

(22) *Informal listener-oriented reformulation of Meisenburg & Gabriel's three constraints*

*VʔV: "you don't have to pronounce a glottal stop between vowels, because the listener will recover it from the resulting hiatus anyway."

ALIGN-L(ʔ,σ): "a syllable onset (realized as hiatus, as a glottal stop, or as creak in the following vowel) is needed to recover *h*-aspiré."

MAX(ə/_ʔ): "*h*-aspiré is easier to recover from hiatus (or from a postvocalic glottal stop) than from a postconsonantal glottal stop."

To improve the descriptive adequacy of the account one would like to build the recoverability of the underlying glottal stop into the candidate set and into the constraint evaluation. The next section shows that this can be done and that none of the three constraints in (22), nor "*C.", is necessary.

2.4. The simplest listener-oriented OT account

The deeper cause of the problems identified in §1.6 through §2.3 is that speaker-based theories model the speaker only. The simplest solution to these problems, then, is to model the listener in a trivial way by trying to include recoverability into the formulation of existing Optimality-Theoretic constraints, and that is the procedure followed in this section. The locus of recoverability must be faithfulness, i.e. it is the faithfulness constraints that will have to evaluate the degree to which the listener will be able to recover the phonological structure. And indeed, very little has to be changed in comparison to McCarthy & Prince's (1995) two-level formulation of faithfulness constraints. In (23) I compare McCarthy & Prince's version of MAX to Boersma's (1998) three-level listener-oriented version.

(23) *Listener-oriented faithfulness as recoverability*

MAX(ʔ) (speaker-based): "pronounce an underlying |ʔ| as /ʔ/ (or [ʔ̰])."

MAX(ʔ) (listener-oriented): "pronounce an underlying |ʔ| as a phonetic form from which the listener will be able to recover /ʔ/."

I will now make explicit what the recovery procedure is. Following Boersma (1998) I assume that this recovery does not involve lexical access, i.e. the recovery is the mapping from a detailed auditory phonetic form to a discrete phonological surface structure without information from the lexicon. Phoneticians call this *perception*, and psycholinguists call it *prelexical processing*. If the lexicon is not involved, the listener has to be able to compute the phonological surface structure directly from the auditory form, regardless of what the lexicon contains.

The non-involvement of the lexicon can be illustrated with *h*-aspiré. In §1.7, I have argued that the listener will be able to recover *h*-aspiré from hiatus, i.e. an auditory phonetic form [leazaʁ] contains a vowel sequence from which the listener will be able to recover the underlying glottal stop. But if she cannot access the underlying form, the listener will *have to* recover a glottal stop, whether the lexicon contains one or not. Because of the automaticity and obligatoriness of this recovery I will simply call the recovery procedure *perception* from now on. The auditory forms [ləazaʁ], [leazaʁ] and [ynəos], then, are *perceived* as the phonological structures /ləʔazaʁ/, /leʔazaʁ/ and /ynəʔos/.[13] This yields the last column in table (6). Note that the perceived phonological structures in this column are not identical to those in any of the other columns. Apparently, this surface structure cannot be an intermediate stage in production, but can only be derived in (6) by undoing the third step (glottal stop deletion) after applying the fourth (which neutralizes and can therefore not be undone).[14]

The perception of hiatus has consequences that go beyond cases of underlying *h*-aspiré. After all, if the vowel sequence [ea] is perceived as /eʔa/ in [leazaʁ], it must be perceived as /eʔa/ in *any* auditory form, regardless of what the lexicon contains, so an auditory form like [agʁeabl] must be perceived as /agʁeʔabl/ despite the fact that this form reflects a single word ('agreeable') whose word-internal hiatus will never be in a position that could show any consonant-like *h*-aspiré effects.[15] The same automaticity of perception predicts that the candidate [ləɔm] in (14c) is perceived as /ləʔɔm/, with a glottal stop that is not present in the underlying form. Likewise, [leɔm] and [ynəide] would be perceived as /leʔɔm/ and /ynəʔide/. The definitions in (24) of the listener-oriented faithfulness constraints in terms of perception show that the perception of hiatus as containing a hidden glottal stop must have an influence on the violation of these constraints.

(24) *Listener-oriented faithfulness as perceptibility (Boersma 1998)*

MAX(ʔ) (speaker-based): "pronounce an underlying |ʔ| as /ʔ/ (or [ʔ̰])."
MAX(ʔ) (listener-oriented): "pronounce an underlying |ʔ| as a phonetic form that the listener will perceive as /ʔ/."
DEP(ʔ) (speaker-based): "do not pronounce a /ʔ/ (or [ʔ̰]) if there is no corresponding underlying |ʔ|."
DEP(ʔ) (listener-oriented): "do not pronounce a phonetic form that the listener will perceive as /ʔ/ if there is no corresponding underlying |ʔ|."

With the perceptions described above (24), the forms [ləazaʁ], [leazaʁ] and [ynəos] satisfy the listener-oriented version of MAX(ʔ), and the forms [ləɔm], [leɔm] and [ynəide] violate the listener-oriented version of DEP(ʔ). This will have repercussions for the tableaus in (14). If we ignore DEP(ʔ) for a while (until §4.1), we see that with a listener-oriented MAX(ʔ) eight tableaus stay the same, but the remaining four tableaus, namely those with an underlying *h*-aspiré, change their violation patterns. For (14e) the winner does not change, as shown in (25).

[13] Or they are perceived with syllable structure as /.lə.ʔa.zaʁ./, /.le.ʔa.zaʁ./ and /.y.nə.ʔos./, or as the Tranel-type surface forms /.lə.a.zaʁ./. /.le.a.zaʁ./ and /.y.nə.os./. Which type of surface structure is perceived depends on one's theory of French phonological structures.

[14] I am aware of the fact that glottal stop deletion neutralizes as well, as is apparent from the next paragraph. Deleted schwas may be recoverable in some positions, such as inside [tl].

[15] Native speakers often object to me that they do not generally hear a glottal stop in hiatus. Although native-speaker intuitions can hardly be decisive about hidden forms that speakers should have no conscious access to, the generalization proposed here is consistent with a remark by Bally (1944:164), according to whom all instances of hiatus, including forms like [ʒeuvɛʁ] from |ʒe#uvɛʁ| 'I have opened' contain a 'zero consonant' (which is Bally's way to denote *h*-aspiré): "tout hiatus contient une consonne zéro, car tout se passe comme si une consonne réelle séparait deux voyelles."

(25) *Listener-oriented enchainment, h-aspiré case*

|kɛl#ʔazaʁ|	Max (C)	Dep (ə)	Max (ʔ)	*[ʔ̰]	*[ə]	Max (V̲)	*CC	Max (C̲)
√☞ [kɛlʔ̰azaʁ] → /kɛlʔazaʁ/				*			*	
[kɛʔ̰azaʁ] → /kɛʔazaʁ/	*!			*				
[kɛləazaʁ] → /kɛləʔazaʁ/		*!			*			
[kɛlazaʁ] → /kɛlazaʁ/			*!					

The candidates now consist of two forms: an overt phonetic form (between square brackets) that itself consists of articulatory and auditory representations related by the speaker's sensorimotor system, and an abstract phonological structure (between slashes) that the listener constructs (with the arrow that denotes perception) from the overt form without knowledge of the underlying form. Faithfulness constraints thus evaluate the similarity between the underlying form and this phonological structure. The only difference in this respect between (14e) and (25) is that the third candidate, given its hiatus, does not violate Max(ʔ) any longer, but this difference does not lead to a different winning candidate (it does point out the importance of Dep(ə), which (14e) did not really do).

Now that the tableau contains three types of representations, it becomes important to state explicitly what forms are evaluated by the starred constraints: the overt phonetic forms or the perceived phonological forms. A constraint like *[ʔ̰] would militate against the articulatory effort of producing a creaky pause, so it would not be violated in the third candidate, since the overt phonetic (articulatory and auditory) form is [kɛləazaʁ]. A constraint like */ʔ/ would militate against maintaining a glottal stop as a mental structure, so it would be violated in the third candidate, since the perceived phonological surface structure is /kɛləʔazaʁ/. In (25) I entertain the articulatory interpretation of *ʔ̰, mainly for an empirical reason discussed below (27). I also choose *[ə] and */CC/, although these choices will not become crucial until §4.

While listener-oriented faithfulness does not change much for |kɛl#ʔazaʁ|, the situation is different for the remaining three *h*-aspiré forms, which are handled incorrectly in the speaker-based tableaus (14b), (14h) and (14k). A listener-oriented Max(ʔ) solves the elision case, as shown in (26).

(26) *Listener-oriented elision, h-aspiré case*

|lə̲#ʔazaʁ|	Max (C)	Dep (ə)	Max (ʔ)	*[ʔ̰]	*[ə]	Max (V̲)	*CC	Max (C̲)
[ləʔ̰azaʁ] → /ləʔazaʁ/				*!	*			
√☞ [ləazaʁ] → /ləʔazaʁ/					*			
[lazaʁ] → /lazaʁ/			*!			*		

In (14b), candidate "ləazaʁ" violated Max(ʔ) because the underlying glottal stop was not pronounced. In (26), candidate [ləazaʁ] does not violate Max(ʔ), because the underlying glottal stop is recovered from the hiatus, giving the perceived form /ləʔazaʁ/. This candidate, which is the correct form in French, thereby becomes the winner. The liaison case is solved entirely analogously, as (27) shows.

(27) *Listener-oriented liaison, h-aspiré case*

|lez̲#ʔazaʁ|	MAX (C)	DEP (ə)	MAX (ʔ)	*[ʔ̰]	*[ə]	MAX (V̲)	*CC	MAX (C̲)
[lezʔ̰azaʁ] → /lezʔazaʁ/				*!			*	
[leʔ̰azaʁ] → /leʔazaʁ/				*!				*
[lezazaʁ] → /lezazaʁ/			*!					
√☞ [leazaʁ] → /leʔazaʁ/								*

In (27) we see that it is crucial that *ʔ̰ is regarded as an articulatory constraint that evaluates the overt phonetic form. Had it evaluated the phonological surface structure instead, there would have been no difference in the violation patterns of the second and the fourth candidate. Understanding *ʔ̰ as an articulatory constraint ensures that of the pronunciations [leʔ̰azaʁ] and [leazaʁ], which are perceived identically, the articulatorily simpler one wins.

In order to prevent misunderstandings it seems appropriate here to point out what it means to say that [leʔ̰azaʁ] and [leazaʁ] are 'perceived identically'. This statement does not mean that the listener cannot hear the difference between these two forms. In a discrimination experiment in the laboratory, the listener can probably hear the two forms apart, and the difference can influence constraint violations in a tableau (see §3.4). The only thing 'perceived identically' means is that the listener maps the two overt phonetic forms to the same phonological surface structure during the language-specific process that in the laboratory can be measured in an *identification* experiment. Other than an discrimination experiment, in which subjects have to perform a slightly unnatural task, an identification experiment reflects the task that a listener has to perform in natural communicative situations, namely preprocessing the auditory input into discrete phonological elements ready for accessing meaning in the lexicon.

The remaining and most crucial form is |ynə̲#ʔos|. Tableau (28) shows that even this form is handled correctly by a listener-oriented MAX(ʔ).

(28) *Listener-oriented schwa drop, h-aspiré case*

|ynə̲#ʔos|	MAX (C)	DEP (ə)	MAX (ʔ)	*[ʔ̰]	*[ə]	MAX (V̲)	*/CC/	MAX (C̲)
[ynəʔ̰os] → /ynəʔos/				*!	*			
[ynʔ̰os] → /ynʔos/				*!		*	*	
√☞ [ynəos] → /ynəʔos/					*			
[ynos] → /ynos/			*!			*		

We see that with listener-oriented faithfulness, which takes into account the recoverability of /ʔ/ from hiatus, the three failures mentioned in (15) are corrected, without the addition of any more constraints. A small remaining problem that could not be handled with the speaker-based model in §2.1, can now also be addressed, as is done in the next section.

2.5. Improving the constraint ranking: the need for DEP(ʔ)

While tableaus (25) through (28) have illustrated a listener-oriented interpretation of MAX(ʔ), a listener-oriented interpretation of its counterpart DEP(ʔ), as defined in (24), also turns out to be relevant for the phonology of French. This is because something is slightly wrong with the analysis so far.

As is apparent from tableaus (14c), (14j) and (14l), I have attributed the ungrammaticality of [ləɔm], [ynəfam] and [ynəide] to the ranking of *[ə] over MAX(V̲). But as noted in §1.5, these three forms are not equally ungrammatical. The forms *[ləɔm], *[ynəide] are much worse than ?[ynəfam], which can occur in practice in a non-negligible minority of cases. The variation between [ynfam] and ?[ynəfam] can be explained by ranking *[ə] just a bit above MAX(V̲). The stochastic form of evaluation in Optimality Theory (Boersma 1997, Boersma & Hayes 2001) then predicts that at evaluation time MAX(V̲) will outrank *[ə] in a large minority of cases, so that forms like ?[ynəfam] will occur, although less often than [ynfam].[16] But a close ranking of *[ə] and MAX(V̲) would also predict the existence of forms like *[ləɔm] and *[ynəide], which are fully ungrammatical. The ungrammaticality of these forms can be explained if we realize that the hiatuses in these forms lead to the perception of a glottal stop that is not underlyingly present, i.e. these forms violate the listener-oriented faithfulness constraint DEP(ʔ). This is made explicit in tableaus (29) through (31).

(29) *Listener-oriented elision, vowel case*

|lə̲#ɔm|	MAX (C)	DEP (ə)	DEP (ʔ)	MAX (ʔ)	*[ʔ̰]	*[ə]	MAX (V̲)	*/CC/	MAX (C̲)
[ləɔm] → /ləʔɔm/			*!			*			
√☞ [lɔm] → /lɔm/							*		
[ləʔ̰ɔm] → /ləʔɔm/			*!		*	*			

(30) *Listener-oriented 'schwa drop', vowel case (= elision)*

|ynə̲#ide|	MAX (C)	DEP (ə)	DEP (ʔ)	MAX (ʔ)	*[ʔ̰]	*[ə]	MAX (V̲)	*/CC/	MAX (C̲)
[ynəide] → /ynəʔide/			*!			*			
√☞ [ynide] → /ynide/							*		
[ynʔ̰ide] → /ynʔide/			*!		*		*	*	

(31) *Listener-oriented 'schwa drop' (elision), consonant case*

|ynə̲#fam|	MAX (C)	DEP (ə)	DEP (ʔ)	MAX (ʔ)	*[ʔ̰]	*[ə]	MAX (V̲)	*/CC/	MAX (C̲)
[ynəfam] → /ynəfam/						*!			
√☞ [ynfam] → /ynfam/							*	*	

[16] I ignore here the possibility that it is */CC/ rather than MAX(V̲) that is variably ranked with respect *[ə]. The constraint */CC/ was specifically introduced as a slightly ad-hoc constraint to account for liaison, and might not survive in a more accurate analysis of that process, e.g. the one below (33).

In (31) the relative grammaticality of ?[ynəfam] can be attributed to a close ranking of *[ə] and MAX(V̲), while in (29) and (30) the high ranking of DEP(ʔ) ensures the ungrammaticality of *[ləɔm] and *[ynəide].

The height of DEP(ʔ) can be determined more precisely than just as DEP(ʔ) >> *[ə]. This constraint cannot be undominated, as can be seen by considering underlying forms like |ʃɑ̃bʁə̲#a#ɛʁ| ‘inner bicycle tube’ or |ʒe#uvɛʁ| ‘I opened’. We know that the morpheme |ɛʁ| ‘air’ in the first form does not have an underlying glottal stop, because |lə̲#ɛʁ| ‘the air’ is pronounced [lɛːʁ], not *[ləɛːʁ]. Nevertheless, the form is pronounced as [ʃɑ̃bʁaɛːʁ], which as a result of its hiatus has to be perceived as /ʃɑ̃bʁaʔɛʁ/. This attested correct French form thus violates DEP(ʔ). The cause must be that all its competitors violate a constraint that is ranked higher than DEP(ʔ). Most notably, the competitor [ʃɑ̃bʁɛːʁ], which is perceived as /ʃɑ̃bʁɛʁ/ and therefore satisfies DEP(ʔ) by deleting one of the vowels that caused hiatus, violates a faithfulness constraint that disallows the deletion of a non-expendable vowel, i.e. the constraint MAX(V), which was still missing from our hierarchy. Tableau (32) shows how the winning form is determined.

(32) *The insertion of a perceived glottal stop*

|ʃɑ̃bʁə̲#a#ɛʁ|	MAX (C)	MAX (V)	DEP (ə)	DEP (ʔ)	MAX (ʔ)	*[ʔ̰]	*[ə]	MAX (V̲)	*/CC/	MAX (C̲)
[ʃɑ̃bʁəaɛːʁ] → /ʃɑ̃bʁəʔaʔɛʁ/				**!			*		*	
√☞ [ʃɑ̃bʁaɛːʁ] → /ʃɑ̃bʁaʔɛʁ/				*				*	*	
[ʃɑ̃bʁɛːʁ] → /ʃɑ̃bʁɛʁ/		*!						*	*	

We thus have evidence for the categorical ranking MAX(V) >> DEP(ʔ) >> *[ə].

The ungrammaticality of *[leɔm] in (14i) can now be attributed to the high-ranked DEP(ʔ) as well, rather than to the bottom-ranked *MAX(C̲), which has now therefore become superfluous.

(33) *Listener-oriented liaison*

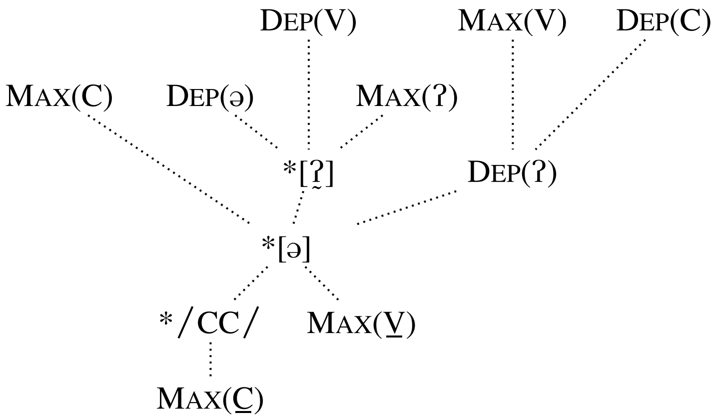
|lez̲#ɔm|	MAX (C)	MAX (V)	DEP (ə)	DEP (ʔ)	MAX (ʔ)	*[ʔ̰]	*[ə]	MAX (V̲)	*/CC/	MAX (C̲)
√☞ [lezɔm] → /lezɔm/										
[leɔm] → /leʔɔm/				*!						*
[leʔ̰ɔm] → /leʔɔm/				*!		*				*

More simplifications could be performed. The constraint */CC/ could be replaced with the simpler */C/. To see that this is possible, remember that the only cases in which */CC/ was active in previous tableaus, are the cases in which it had to rule out *[lezgaʁsɔ̃] and *[lʔ̰azaʁ]. In these tableaus */C/ would work as well. What prevented us from using */C/ from the start was tableau (14i), in which *C would disfavour the correct form [lezɔm]. With DEP(ʔ) high-ranked, this problem is out of the way, and liaison could be handled by MAX(C) >> */C/ >> MAX(C̲), which is a normal positional faithfulness ranking (Beckman 1998). In

(33), the constraint */C/ would be violated three times in each candidate. As we see in tableau (25), */C/ would have to be ranked below MAX(ʔ).[17]

The full set of crucial rankings is shown in (34). Two unviolated constraints have been added in this graph: DEP(V), which has to rule out the insertion of a non-expendable vowel in a form like *[kɛluazaʁ] in tableau (25), and DEP(C), which has to rule out the insertion of a consonant in a form like [ʃɑ̃bʁatɛːʁ] in tableau (32).

(34) *A ranking for hache aspiré without optionality*



2.6. Evaluative comparison of the speaker-based and listener-oriented accounts

When compared to the two-level speaker-based OT accounts of Tranel & Del Gobbo (2002) and Meisenburg & Gabriel (2004), which vary in their degrees of abstractness of surface forms, the three-level listener-oriented account proposed in §2.4 and §2.5 makes a principled distinction between the overt phonetic form and the abstract hidden phonological surface structure. But a comparison in terms of grammar evaluation is also possible. When compared to Meisenburg & Gabriel's proposal, the listener-oriented proposal economizes on four constraints: *VʔV, ALIGN(ʔ,σ), "*C̲." and MAX(ə/_ʔ). The cause of this economy lies in the fact that these four constraints express the same idea, namely connecting *h*-aspiré to auditory cues such as hiatus or a glottal stop, an idea that the listener-oriented account can replace with a single rule, namely that hiatus leads to the perception of a glottal stop.

The listener-oriented account comes with its own cost, namely that of the rule that the phonetic form [VV] is perceived as the phonological form /VʔV/. The cost of the listener's rule [VV] → /VʔV/ seems to be comparable to the cost of the speaker's rule in the fifth column of (6), which can be written /VʔV/ → [VV]: the two are perfect mirror images, as we can expect from their opposite viewpoints (production versus comprehension). When compared to Meisenburg & Gabriel's constraint set, the cost of having the glottal stop deletion rule in (6), in turn, seems to be comparable to the cost of having the constraint *VʔV.

After this cost-gain analysis, the listener-oriented account can roughly be said to economize on three grammar elements, namely the constraints ALIGN(ʔ,σ), "*C." and MAX(ə/_ʔ), of which especially the last one is problematic as far as universal defensibility is concerned. Three large questions remain, however. The first is how to assess the rule [VV] →

[17] Yet another analysis of liaison that has become possible is Tranel's DEP(X), see footnote 7. In §2.2, Tranel's analysis required a ranking like ONSET >> DEP(X), which would be problematic in the present syllable-free framework. Now that DEP(ʔ) is available, however, the same effect can be achieved with the ranking DEP(ʔ) >> DEP(X), i.e. [lezɔm] /lezɔm/ is better than [leɔm] /leʔɔm/ because the former violates DEP(X), the latter DEP(ʔ).

/VʔV/. Is it extragrammatical, or can it be incorporated into the OT grammar? Is it universally defensible, or is it a quirk specific to French? These questions are answered in §3. The second question concerns the fact that although §2.4 has shown that a simple listener-oriented OT can handle the preference of [ynəos] over [ynos], by proposing that the underlying glottal stop is recoverable from [ynəos] but not from [ynos], the same section §2.4 has attributed the preference of [ynəos] over [ynʔ̰os] to the speaker-based ranking *[ʔ̰] >> *[ə], contrary to the listener-oriented explanation of §1.7 in terms of the quality of auditory cues. Can the idea that [ynəos] has better auditory cues for recovering *h*-aspiré than [ynʔ̰os] be formalized? This question is answered affirmatively in §4, where some explicit listener-oriented models of the interplay between phonology and phonetics are proposed. The third question concerns the form [ynəʔ̰os], with both schwa and a creaky pause. In (28), this form is harmonically bounded, but it has been observed in reality. Can it ever win? This question is answered in the affirmative in §4.

3. Modelling the recovery process

Although the account presented in §2.4 to §2.6 was called 'listener-oriented', only the speaker was modelled explicitly: the production process starts from an underlying form and chooses the optimal pronunciation partly on the basis of the phonological structure that the listener will reconstruct (or that the speaker thinks the listener will reconstruct). A more descriptively adequate account will require a bidirectional model of phonology, which formalizes not only the speaker's behaviour, but the listener's behaviour (and the speaker's view of the listener's behaviour) as well. This section first discusses the representations and mappings between them that are needed for modelling the listener, then proposes the constraints and rankings that are needed for handling the French perception of *h*-aspiré. Finally, some explanatory adequacy is achieved when it is shown that the ranking can be explained as a result of lexicon-driven learning. We will see that the learning algorithm leads to a ranking that can be seen as the result of auditory distinctivity, warped by language-specific frequency effects.

3.1. Four representations

As we can see from §2.4, a model for production will have to posit more than one representation outside the underlying lexical form (UF). I propose that there are three: the *auditory form* (AudF), the *articulatory form* (ArtF), and the *surface form* (SF). The first two are phonetic forms, continuous in time and extent, written in two universal alphabets (auditory spectrum, noise, pitch; articulatory gestures), and reasonably accessible to the scientist who investigates the inner ear or the speech tract. The two phonological forms (surface and underlying), on the other hand, are abstract mental structures, written in discrete phonological elements that either emerge in a language-specific way during acquisition or are innately given by Universal Grammar, and only very indirectly accessible to the scientist in perception or recognition experiments. This fourfold distinction, introduced to phonology by Boersma (1998), is a union of the representations proposed by Chomsky & Halle (1968: UF and ArtF), Boersma (1989: ArtF and SF), and Flemming (1995: AudF and ArtF). Outside phonological theory proper, it can be seen as the union of the representations proposed by the psycholinguists McQueen & Cutler (1997: AudF, SF, UF) and Levelt (1989: UF, SF, ArtF).

In the tableaus of §2.4 and §2.5, all four representations are relevant: the top left cell is UF; the candidate cells contain a phonetic form (ArtF/AudF) and SF, as well as arrows that represent the speaker's view of the listener's AudF→SF mapping; the constraints *[ʔ̰] and *[ə] evaluate ArtF; the constraint */CC/ evaluates SF; and the faithfulness constraints evaluate the similarity between SF and UF.

3.2. Representations and mappings for the listener

The speaker's view of the listener's perception has already been discussed: it is the AudF→SF mapping in the production model. But the full comprehension process involves more. At least it has to involve lexical access, i.e., the listener has to construct an UF. There are theories of comprehension in which the listener starts by creating a mental image of the speaker's ArtF, which presumably must involve an AudF→ArtF mapping (Fowler 1986, Best 1995). But since 12-month-old infants already have a lexicon with extensive phonological representations while at the same time being unable to speak, I bluntly assume that ArtF is not passed through in adult comprehension either. This restricts the listener's task to mapping AudF to SF and UF.

The mapping from AudF to SF and UF could be done in parallel, i.e. as AudF → { SF, UF }, in which case the listener's perception would be influenced by lexical access. In §2.4, I assumed that the listener's perception does not involve lexical access. This would still allow several models of how UF is accessed: directly from AudF, indirectly from the perceived SF, or from AudF and SF at the same time. The simplest option is the one defended by McQueen & Cutler (1997), in which UF is computed from SF alone, and it is this view that was proposed for three-level OT by Boersma (1998: 143, 269). The comprehension model in that view is therefore AudF → SF → UF: it consists of two sequential modules, where the output of the first (perception) is the input to the second (recognition).

Whatever the form of UF access, the listener's AudF → SF mapping is a language-specific process, which means that phonological theory cannot evade modelling it with explicit linguistic means. Boersma (1998) calls the mapping *perception*, following the use of this term in speech research, and models it with Optimality-Theoretical constraint ranking; examples include the mapping from auditory cues such as the first formant to discrete phonological categories such as vowel heights (Boersma 1997), the mapping from a continuous sequence of auditory cues to single or multiple discrete phonological elements (Boersma 1998:ch.18, 2000), or the integration of multiple auditory cues into a bivalued or multivalued phonological contrast (Escudero & Boersma 2003, 2004; Boersma & Escudero 2004). A very similar OT mapping was proposed by Tesar (1997, 1998, 1999) and Tesar & Smolensky (1998, 2000), namely *(robust) interpretive parsing*; their example is the mapping from a concrete overt string of stressed and unstressed syllables, e.g. [σ σ́ σ], to an abstract hidden prosodic structure with feet and head syllables, e.g. /σ (σ́ σ)/. Since all of the examples are about the construction of a more abstract from a less abstract representation, the terms *interpretive parsing* and *perception* can refer to the same thing and have been treated as synonyms in work on metrical phonology (Apoussidou & Boersma 2004), although Smolensky (p.c.) maintains that the two work on a different level.

3.3. A grammar for perceiving *h*-aspiré: one example

In the tableaus of §2.4 and §2.5, perception was modelled as a simple arrow in every candidate cell. As an example of perception in phonology I will now make explicit what is behind the third arrow in tableau (28), namely the French listener's mapping from an overt [VV] to a hidden /VʔV/ for the auditory form [ynəos]. The Optimality-Theoretic perception tableau is shown in (35).

(35) *The perception of hiatus*

[ynəos]$_{\text{Aud}}$	*/VV/	*/CV/ ╳ [VC]	*/VCV/ \ / [VV]	*/V/ \\ [VV]	*/VʔV/ \ / [VV]
/ynəos/	*!				
/ynos/				*!	
/ynəso/		*!			
☞ /ynəʔos/					*
/ynətos/			*!		

In (35) the input to the grammar is an auditory phonetic form (or at least a rather concrete structure), and the output of the grammar is a perceived phonological form, in this case /ynəʔos/, a sequence of six French phonemes. I will now discuss the meaning of the five constraints and for four of them I will also explain how the acquisition process has ranked them in the order they are ranked in (35).

The candidate with hiatus, /ynəos/, is ruled out by a high-ranked constraint against hiatus at the perceptual level. This constraint has to be formulated with slashes, i.e. as */VV/, in order to make sure that it evaluates the SF. It is a structural constraint that evaluates a structure (SF) that occurs in the output of perception as well as production, so it could be capable of influencing perception as well as production. This bidirectional use of constraints at SF was stressed by Tesar & Smolensky (2000) in their robust interpretive parsing, as well as by Pater (2004) for perception. In Boersma's (1998) control-loop model of production, structural constraints directly evaluate only the output of perception, whereas their influence on production is indirect; in §2.4 and §2.5, for instance, */VV/ does not make its appearance in the tableaus, nor does it have to, since its workings are already expressed in the arrows that relate AudF to SF in the candidate cells. Tableau (35) formalizes precisely one of those arrows, namely the third arrow in (28). The contribution of */VV/ to the complexity of the grammar is comparable to that of the rule *ʔ/V_V in (6) or Meisenburg & Gabriel's constraint *[VʔV] in §2.3. The constraint is top-ranked in (35); I defer a discussion of how this high ranking might come about until §4.8 and §5.

If [ynəos] cannot be perceived as /ynəos/ because of an overriding constraint, it will be perceived as something else: this is *robust perception*, it will not fail. The perceived phonological structure will be the one that minimally violates the remaining four constraints in (35), all of which are negatively formulated *auditory-phonological mapping constraints* (Escudero & Boersma 2003, 2004) that militate against mapping any auditory cue to any phonological element.[18] It is important to realize that these constraints relate two representations that are written in different alphabets, i.e. they are continuous-to-discrete perceptual mapping constraints, different from the continuous-to-continuous perceptual faithfulness constraints proposed by Boersma (1997) and from the discrete-to-discrete perceptual faithfulness constraints by Pater (2004).

[18] One of the reasons for the negative formulation is explained in Boersma & Escudero (2004), the other here in §4.6.

The second constraint in (35) is about order. The auditory form [ynəos] contains information on the relative timing of sibilance and back vocality. Perceiving it as /ynəso/ would discard this information, violating a constraint loosely statable as *[V_1C_2]/C_2V_1/, which is short for "do not perceive auditory vowel cues followed by auditory consonant cues as the consonant that corresponds to those consonant cues followed by the vowel that corresponds to those vowel cues." In the horizontal formulation *[V_1C_2]/C_2V_1/, the correspondence between auditory and phonological elements is expressed by coindexation; in the vertical formulation in (35), the same correspondence is expressed in a more visually appealing way, namely by association lines.

The question concerning the descriptive adequacy of the present model of phonology is now: what is the linguistic goal of having *[V_1C_2]/C_2V_1/ ranked so high? And the question concerning the explanatory adequacy is: how has *[V_1C_2]/C_2V_1/ become ranked so high? The deep answer to the first question is the idea that pre-lexical perception has to convert the acoustic-phonetic signal into something maximally prepared for lexical access (Lahiri & Marslen-Wilson 1991, McQueen & Cutler 1997). An optimal perception, then, is one that constructs an SF that is as close as possible to the UF under the restriction that it cannot access the lexicon. For French, which does not use metathesis in alternations (i.e. it never violates McCarthy & Prince's LINEARITY in production), an optimal SF always has its elements in the same order as the intended UF and in the same order as the cues in AudF. This means that *[V_1C_2]/C_2V_1/ has to be ranked high in perception. The answer to the second question is that this high ranking is an automatic result of lexicon-driven learning of perception (Boersma 1997, Escudero & Boersma 2004). Tableau (36) illustrates what happens if at a certain point during acquisition the learner perceives [gaʁsɔ̃] incorrectly as /.gʁa.sɔ̃./, because of a ranking of *[V_1C_2]/C_2V_1/ below the structural constraint *CODA (which I write as */C./ in order to make explicit that it works at SF).

(36) *Learning the high ranking of auditory-to-phonological 'linearity'*

<table>
<tr><th>[gaʁsɔ̃]_{Aud}
UF = |gaʁsɔ̃|</th><th>*/C./</th><th>*/CV/
╳
[VC]</th></tr>
<tr><td>√ /.gaʁ.sɔ̃./</td><td>*!→</td><td></td></tr>
<tr><td>☞ /.gʁa.sɔ̃./</td><td></td><td>←*</td></tr>
</table>

The auditory form [gaʁsɔ̃] is perceived as /.gʁa.sɔ̃./, but the subsequent process of word recognition maps this SF to the lexical item |gaʁsɔ̃| 'boy' at UF (thus violating LINEARITY, which is a kind of UF-SF faithfulness). The learner uses this lexical information to mark the UF-identical form in the tableau with a check mark (√). Now that the winning candidate /.gʁa.sɔ̃./ is different from the one that the learner considers correct, namely /.gaʁ.sɔ̃./, the learner will take action by raising the ranking of the constraints that prefer the correct form over the winning form, and lowering the ranking of the constraints that have the opposite preference. The rankings move along a continuous scale, assuming Stochastic OT (Boersma 1998); the learning algorithm is the Gradual Learning Algorithm (Boersma 1997, Boersma & Hayes 2001). In tableau (36), *CODA will fall and *[V_1C_2]/C_2V_1/ will rise, thus making it more likely that the learner will perceive /.gaʁ.sɔ̃./ at future occurrences of [gaʁsɔ̃]. Since no cases of metathesis occur in French, $*[V_1C_2]/C_2V_1/$ will ultimately end up ranked far above *CODA.

The third and fifth constraints in (35) militate against the perception of a phonological structure for which there is no direct auditory evidence. Such 'hallucinatory' behaviour must be part and parcel of human speech perception, since background noise must be considered capable of erasing auditory cues all the time. If the listener wants to heed all the positive auditory information in [ynəos] and perceiving it as /ynəos/ is out of the question, she will introduce a consonant, perhaps perceiving it as /ynətos/ or /ynəʔos/. But these two are not equally viable candidates. Even in an over-faithful version of French in which underlying |ynətos| and |ynəʔos| are pronounced as the articulations $[\text{ynətos}]_{Art}$ and $[\text{ynəʔos}]_{Art}$, respectively, the first one has a much smaller chance of being heard as $[\text{ynəos}]_{Aud}$ in a noisy environment than the second, simply because $[t]_{Art}$ produces a superset of the auditory cues that $[ʔ]_{Art}$ produces (the shared cue is the silence; the cues that $[t]_{Art}$ does and $[ʔ]_{Art}$ does not produce are formant contours and release burst). If $[t]_{Art}$ is not much more common than $[ʔ]_{Art}$ in this over-faithful French, $[\text{ynəos}]_{Aud}$ will more often derive from |ynəʔos| than from |ynətos|, so that lexicon-driven learning along the lines of tableau (36) will lead to a ranking of *[VV]/VtV/ >> *[VV]/VʔV/.[19] In a bidirectional situation, such a ranking is self-reinforcing: the speaker, knowing that the listener is likely to perceive $[\text{ynəos}]_{Aud}$ as /ynəʔos/, may simply stop pronouncing the glottal stop and get away with it (Boersma 1998:182),[20] thus making it even less likely that positive glottal stop cues make it to the listener, who will as a result lower the ranking of *[VV]/VʔV/ even further.

The final constraint to discuss is the one against discarding the auditory cues for the presence of /ə/. Perceiving $[\text{ynəos}]_{Aud}$ as /ynos/ would discard the auditory information in a pronounced $[ə]_{Art}$, violating $*[V_1V_2]/V_2/$. Background noise is much more likely to conceil cues that were actually pronounced than to create cues that were not pronounced; the random creation of cues that would fit in the speech stream must be a relatively improbable event.[21] Given an auditory cue X for a feature Y, then, the constraint against discarding it in perception, i.e. *[X]//, will usually be ranked higher than the constraint against hallucinating it in perception, i.e. *[]/Y/.[22] If schwa cues have a comparable quality as glottal stop cues, $*[V_1V_2]/V_2/$ will be ranked above *[VV]/VʔV/, as it is in (35). If schwa has better cues than a glottal stop (as proposed in §1.7), the ranking difference will be even larger. In (35), I have ranked $*[V_1V_2]/V_2/$ below *[VV]/VtV/, because I think $[\text{ynəos}]_{Aud}$ is more likely to be perceived as /ynos/ than as /ynətos/. A perception experiment may shed light on issues like these but is outside the scope of the present paper (see footnote 23).

[19] The relative commonness of |t| and |ʔ| also plays a role. The more common an underlying element is, the lower constraints against perceiving it will be ranked.

[20] This is an informal explanation of the underlying mechanism behind what Steriade (1995) called *licensing by cue*, the phenomenon that faithfulness constraints are ranked higher for phonological elements with more auditory cues or in positions where they have more auditory cues. For an explicit version of this informal explanation in terms of minimization of the probability of confusion, see Boersma (1998:182,370,374); for a formal explanation in terms of blind mechanisms, see §4.3.

[21] This is the explanation for the phenomenon that loanword adaptation more often inserts than deletes something, e.g. Russian [tak] is borrowed in Japanese as /taku/, not /ta/. See also the next footnote.

[22] *Usually*. Namely, if all else is equal. Language-specific frequencies may change these rankings as an automatic result of the lexicon-driven learning mechanism. By the way, the loanword adaptation phenomenon mentioned in the previous footnote is due to the universal and learnable tendency to rank *[X]// >> *[]/Y/ in perception, not by an unexplained universal tendency to rank MAX(Y) >> DEP(Y) in production as proposed by unidirectional two-level theories of loanword adaptation.

3.4. Ranking by cue

This section gives all the constraints responsible for the order of auditory cue quality proposed in §1.7. I ignore high-ranked constraints like the second and third constraints in tableau (35) because I restrict the candidate set to the most plausible phonological structures with and without a glottal stop. Beside the structural constraint */VV/ we need eight mapping constraints rather than the two present in tableau (35).

The first four mapping constraints to consider are those against perceiving a glottal stop. In §1.7, the cues for the glottal stop were proposed to be better in [ynəos] than in [ynʔ̰os], since this difference was able to account for the speaker's preference of [ynəos] over [ynʔ̰os]. Thus, I propose to derive auditory cue qualities from language data rather than from a perception experiment; in this respect, doing phonology by modelling perception can be much like doing phonology by modelling production, which is immeasurably more common in the history of phonological theory.[23] The (perhaps language-specific) quality of the auditory cues in [ynəos] and [ynʔ̰os] will lead to the ranking *[VV]/VʔV/ >> *[Cʔ̰V]/CʔV/. But there are another two relevant auditory forms. The form with the best cues for glottal stop must be the form that includes both schwa and the creaky pause, i.e. [ynəʔ̰os]; the constraint *[Vʔ̰V]/VʔV/ must therefore be very low ranked. Finally, the form with neither schwa nor the creaky pause, i.e. [ynos], must have the poorest cues for glottal stop; the constraint *[CV]/CʔV/ must therefore be very high ranked. The ranking of the four constraints, then, is that in (37).

(37) *Ranking of constraints against perceiving a glottal stop in French*

```
*/CʔV/        */CʔV/        */VʔV/        */VʔV/
  \ /    >>     | | |   >>    \ /     >>    | | |
  [CV]         [Cʔ̰V]         [VV]         [Vʔ̰V]
```

This ranking must already be a simplification. For instance, the formulation of the second constraint, the one against perceiving the underlying *h*-aspiré in [kɛlʔ̰azaʁ], is a summary for five more precise formulations, as shown in (38).

(38) *Simplifying a perceptual mapping constraint*

```
*/CʔV/      */CʔV/      */CʔV/      */CʔV/      */CʔV/      */CʔV/
 | | |   =   | | |   or   | | |   or  / /\ \   or  / /\ \   or   \ \/
[Cʔ̰V]       [C_V]       [C ̰V]       [C_ˀV]      [C ̰ˀV]       [CV̰]
```

[23] This is not to say that perception experiments cannot be useful. They are, and they should ideally give the same results as can be derived from language data. In the present case a perception experiment would involve having French listeners classify auditory forms like [ynəotmɔ̃taɲ] and [ynʔ̰otmɔ̃taɲ] as either "une haute montagne" (i.e. |ynə̲#ʔotə̲#mɔ̃taɲə̲| 'a high mountain') or "une autre montagne" (i.e. |ynə̲#otʁə̲#mɔ̃taɲə̲| 'another mountain'). Large methodological challenges include (but seem not to be limited to): varying realistically the degree of reduction of schwa in [ynəot], varying realistically the multitude of realizations of [ynʔ̰ot], varying realistically the amount and type of background noise, and varying realistically the listener's attention level. For want of experimental set-ups that realistically reflect phonology in use, the most direct source of evidence for researchers interested in modelling perception must be the discrete speaker's choices reported in the literature (such as the choice for or against pronouncing a schwa), just as they are for the much more numerous researchers interested in modelling production.

These examples may be instructive because they show that two consecutive cues can contribute to the perception of a single segment and that two simultaneous cues can be perceived as a sequence of two segments.

The other four constraints to consider are those against perceiving a structure *without* a glottal stop, i.e. the structures /CV/ and /V/ (the structure /VV/ is ruled out independently). One of them occurred in (35). Their natural order must be the reverse of that of their counterparts in (37).

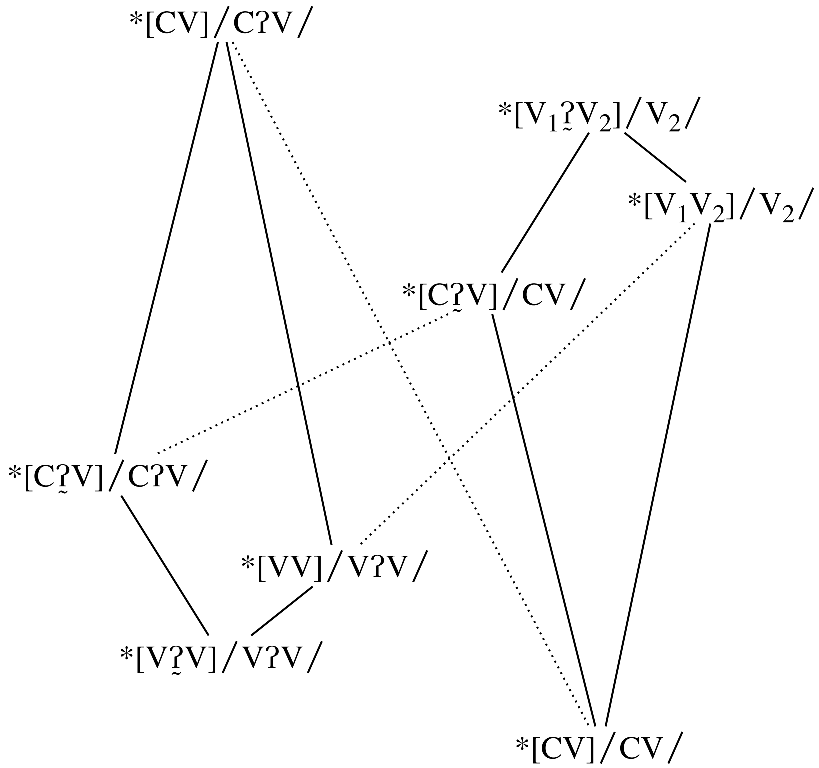
(39) *Ranking of constraints against not perceiving a glottal stop in French*

* /V/		* /V/		*/CV/		*/CV/
\	>>	\	>>	/ \	>>	| |
[Vʔ̰V]		[VV]		[Cʔ̰V]		[CV]

Again, these constraints are mainly universally ranked by their degree of discarding French-specific auditory cues to underlying |ʔ| (or underlying syllable boundary). The form [Vʔ̰V] has a superset of the cues in [VV] and [Cʔ̰V], which again have more cues than [CV], which has none. The idea that [VV] has better cues for French listeners than [Cʔ̰V] is again based on the linguistic observation (§1.5) that [ynəos] is preferred over [ynʔ̰os] (or at least [vwasiləazaʁ] over [vwasilʔ̰azaʁ]).

To obtain the relative recoverabilities listed in (7), the eight constraints must be ranked along a continuous ranking scale in the vein of Stochastic OT (Boersma 1997, 1998; Boersma & Hayes 2001), i.e. the higher the ranking of constraint A above constraint B along this scale, the more often A will outrank B at evaluation time. I propose the ranking in (40).

(40) *Constraint ranking for the recovery of a glottal stop in French*



In the figure, the rankings-by-cue are depicted by solid lines. The three dotted lines depict the rankings that make sure that [ynos] is most often perceived as /ynos/, that [ynʔ̰os] is most often perceived as /ynʔos/, and that [ynəos] is most often perceived as /ynəʔos/. The usual perception of [ynəʔ̰os] as /ynəʔos/ is guaranteed by transitivity.

If we assign numerical values to the rankings of perception constraints, we can compute the probability that an auditory form X is perceived as the phonological structure Y. The four tableaus in (41) show this for the present case.

(41) *Variation in perception*

	98.0	97.0	96.0	95.0	93.0	92.0	91.0	90.0	
[ynəʔ̰os]	*/CʔV/ \ / [CV]	* /V/ \\ [Vʔ̰V]	*/V/ \\ [VV]	*/CV/ / \\ [Cʔ̰V]	*/CʔV/ ||| [Cʔ̰V]	*/VʔV/ \ / [VV]	*/VʔV/ ||| [Vʔ̰V]	*/CV/ | | [CV]	
/ynəʔos/							*		98%
/ynos/		*							2%

[ynʔ̰os]									
/ynʔos/					*				76%
/ynos/				*					24%

[ynəos]									
/ynəʔos/						*			92%
/ynos/			*						8%

[ynos]									
/ynʔos/	*								0.2%
/ynos/								*	99.8%

The ranking value of each constraint has been written above it in the tableau. These are fake values guessed by the author, but lead to plausible percentages of perception. These percentages, written to the right of each candidate row, were computed by filtering each auditory form through the tableaus 100,000 times with an *evaluation noise* (standard deviation of the ranking at evaluation time) of 2.0. The auditory form [ynos] is nearly always perceived as /ynos/, which accounts for the label 'bad' for the recoverability in (7) if [ynos] is used to implement |ynə̠+ʔos|. The form [ynʔ̰os] is perceived as /ynʔos/ 76% of the time, which makes the postconsonantal creaky pause an 'okayish' implementation of an underlying glottal stop as far as recoverability is concerned. The form [ynəos] fares better: it is perceived as /ynəʔos/ 92 percent of the time, so that hiatus can be considered a 'good' implementation of an underlying glottal stop. Although the label 'excellent' does not occur in (7), it would be a suitable verdict on the form [ynəʔ̰os]: if used for implementing a glottal stop, the listener will make an error in no more than 2 percent of the cases.

Now that the listener has been modelled, it is time for modelling the speaker in more detail.

4. Modelling recoverability

As noted in §2.6, the simple listener-oriented grammar in (34) is not listener-oriented enough. In §1.7, I attributed the preference for [ynəos] over [ynʔ̰os] to the idea that [ə] has better auditory cues than [ʔ̰], i.e. that [ynəos] is more clearly distinguishable from [ynos] than [ynʔ̰os] is. But in tableau (28), the preference for [ynəos] over [ynʔ̰os] seems to be due to the ranking *[ʔ̰] >> *[ə], i.e. to the idea that a creaky pause is more difficult to pronounce than a schwa. If this were correct, the form that led me to propose listener-orientation to begin with would suddenly obtain a speaker-based explanation.

A speaker-based explanation for the preference for [ynəos] over [ynʔ̰os] is unlikely. The existence of the variants [ynəʔ̰os], [ləʔ̰azaʁ] and [leʔ̰azaʁ] (§1.6) cannot have a speaker-based explanation in terms of *[ʔ̰] >> *[ə], since these forms violate *both* articulatory constraints, and are in fact harmonically bounded in (26) to (28). The explanation for the occurrence of these three forms must lie in perceptual enhancement, i.e. the speaker's wish to choose a variant with a minimum degree of confusion (in this case, only 2 percent). I will therefore assume that the preference of [ynəos] over [ynʔ̰os] is wholly or partly due to the same drive that makes [ynəʔ̰os] an option. The theory to be developed has to allow a ranking of *[ʔ̰] below *[ə] and still predict that [ynəos] is the preferred form. It is reassuring to see that in the other 11 tableaus, namely (14acdfgijl), (25), (26), and (27), the constraint *[ʔ̰] could just as well have been bottom-ranked, and the same winners would have resulted.

In the following sections I discuss how the ranking-by-cue proposed in §3.4 can be integrated into an Optimality-Theoretic model of production. Several grammar models have been proposed for production. I will only discuss models that include phonetic detail, i.e. those enumerated in (42).

(42) *Five production models that integrate phonology and phonetics*

Sequential:	UF → SF → phonetic form
Two-level:	UF → phonetic form
Control loop:	UF → { ArtF → AudF → SF }
Stochastic control loop:	UF → { ArtF → AudF → SF }
Parallel:	UF → { ArtF, AudF, SF }

I will argue that only the last three are capable of handling the *h*-aspiré case, but that the parallel model is probably the best.

4.1. The sequential production model

The sequential model in (42a) is probably the current mainstream view. The first arrow represents *phonology* and the second arrow *phonetic implementation*. Even researchers who have worked on the phonetics-phonology interface are able to share this view, e.g. Hayes (1999 [1996:7]). It does not work, though, for the case of *h*-aspiré. In the form [ynəos], phonetic considerations of perceptibility have led to a discrete phonological decision, namely to include /ə/ in the surface form. A modular production model, in which the output of phonology is the input of phonetic implementation, can never account for such a phenomenon.

4.2. The two-level production model

Although strong evidence for a non-lexical discrete level of representation is provided by the existence of phenomena like language games and slips of the tongue (Fromkin 1971, Bagemihl 1995), several phonologists have rejected the role of such a form in production. A model that did not accept a privileged intermediate form in a derivation, i.e. the model in (42b), was proposed by Halle (1959) and Chomsky & Halle (1968).

Within Optimality Theory, several models that take UF as their input and a phonetic form as their output have been proposed. Jun (1995) accounts for place assimilation by using constraints that favour the 'preservation' of auditory cues in the phonetic output. Since his model lacks a SF to which these cues can relate in a language-specific and position-dependent way (as here in §3.4), Jun has to propose that these cues are given as invariants for each of the phonological elements that make up the UF. Other proposals with detailed output representations in two-level phonology are given by Kirchner (1998) and Flemming (2003). Steriade (1995) has a more discrete view of the output, something in between the SF and phonetic form as interpreted in the present paper: for her, the output consists of discrete but low-level features such as [short VOT], which must as a result of their discreteness be regarded as invariable. Steriade admits that the assumption of invariance could be a weak point in her theory (ch.1:p.25), and so it is in the case of *h*-aspiré, where one and the same discrete element (/ʔ/) has such varying discrete and continuous cues (hiatus, creak, pause, glottal stop). While these two-level accounts have been able to include the influence of perceptibility on the ranking of faithfulness, they have done so indirectly, by proposing that perception is extralinguistic and universal and that an underlying phonological element /X/ is invariably specified for the auditory cues [Y] and [Z]. Steriade's unrealistic simplification is understandable, since a formal model of ranking-by-cue that would account for an interplay between auditory distinctivity (i.e. robustness against background noise) and language-specific perceptual bias would require an SF level, something not available in two-level OT.

4.3. The control-loop model with probabilistic faithfulness

Since the interplay between auditory distinctivity and language-specific perceptual bias exists, and the invariance between phonological features and auditory cues does not exist, a correct formal account of the influence of ranking-by-cue on production requires a three-level model of production, i.e. a model that includes an SF level (§4.2). Since phonetic considerations can influence discrete phonological decisions, this SF level cannot, however, be an intermediate representation in a sequence of two modules (§4.1). In the remaining three models in (42), SF exists but is not an intermediate representation. All three are listener-oriented to some extent.

The listener-oriented production model of §2.4 and §2.5 can be seen as a *control grammar* (Boersma 2003b), in the sense that by choosing an ArtF the speaker controls the listener's perception of the SF, which the speaker's faithfulness constraints compare with the UF. In this production model, the idea that [ynəos] has better auditory cues for *h*-aspiré than [ynʔ̰os] can be expressed with *probabilistic faithfulness constraints* (Boersma 2000; Boersma 2003a: 42–44; Boersma & Hamann 2005). These are listener-oriented faithfulness constraints that take into account the probability (between 0 and 100 percent) that a listener will fail to recover the given feature. For *h*-aspiré, the probabilistic faithfulness constraint would be MAX (ʔ, *p*%): "pronounce an underlying |ʔ| as a phonetic form from which the listener has at most *p* percent probability of failing to recover /ʔ/." We can see in (41) that the form [ynəos] violates MAX (ʔ,3%) but not MAX (ʔ,10%), whereas [ynʔ̰os], which has poorer *h*-aspiré cues than [ynəos], violates both of these constraints and can therefore be ruled out, as shown in (43).

(43) *Listener-oriented schwa drop, h-aspiré case*

|ynə̲#ʔos|	MAX (C)	DEP (ə)	MAX (ʔ,80%)	MAX (ʔ,10%)	*[ə]	*[ʔ̰]	MAX (ʔ,3%)	MAX (V̲)	*/CC/	MAX (C̲)
[ynəʔ̰os] → /ynəʔos/					*	*!				
[ynʔ̰os] → /ynʔos/				*!		*	*	*	*	
☞ [ynəos] → /ynəʔos/					*		*			
[ynos] → /ynos/			*!	*			*	*		

We see that although *[ə] outranks *[ʔ̰] in this example, the form [ynʔ̰os] can still be ruled out as a result of its lower perceptibility. This is therefore a more listener-oriented tableau than (28).

Variation can also be accounted for. In (43) we can see that the occasional occurrence of the enhanced form [ynəʔ̰os] could be explained by a variable ranking of *[ʔ̰] and MAX (ʔ,3%), i.e. by a variable weighing of the enhanced perceptibility brought about by producing a creaky pause and the increased effort required for articulating it. Probabilistic faithfulness constraints directly take into account percentages of perception such as those computed in §3.4, and therefore provide a close link between constraint ranking in production and universal and language-specific cue quality. They also provide a link between two seemingly disparate phenomena: for high *p* values, these constraints express Steriade's (1995) *licensing by cue*; for low *p* values (as in the *h*-aspiré example), they express enhancement.

However, there is no known on-line learning algorithm for probabilistic faithfulness constraints, which is a serious hindrance if we want to achieve explanatory adequacy. According to Boersma (1998:269), a learning algorithm for ranking the constraints in a control-loop model of production is based on a comparison between two phonological surface structures. The first of these is the form that the learner perceives when somebody else produces an auditory form. The second is based on the learner's own production: from the perceived SF the learner will reconstruct an UF, and from this UF she will compute the ArtF, AudF and SF that she herself would have produced, a procedure that Apoussidou & Boersma (2004) call *virtual production*. This second SF, therefore, is the form that the learner imagines to be able to evoke in the listener. If the two surface forms are different, the learner will use her Gradual Learning Algorithm to rerank the constraints.[24] But if only one piece of data can arrive at a time, the algorithm has no way to tell which of the FAITH(*p*%) constraints to rerank.

Other problems with probabilistic faithfulness may be the large number of constraints (one for 'every" value of *p*), the analogous need for probabilistic structural constraints, and the duplication of the ranking of the perceptual mapping constraints, i.e. (41), in a set of heterogenous constraints in the production grammar.

[24] The comparison strategy is similar to that in Tesar & Smolensky's (1998, 2000) Robust Interpretive Parsing with Constraint Demotion.

4.4. The stochastic control-loop model

The problems with probabilistic faithfulness can be solved by allowing a stochastic evaluation of the perception mapping in the candidate cells. If we look at tableau (28), we see that the choice for the winning candidate is based on the mapping [ynəos] → /ynəʔos/. But if we look at tableau (41), we see that [ynəos] is not always perceived as /ynəʔos/. In fact, it is perceived as /ynos/ in 8 percent of the cases. It can now be argued that in those 8% of the production cases, tableau (28) has to be changed to (44).

(44) *Une hausse, 8 percent of the time*

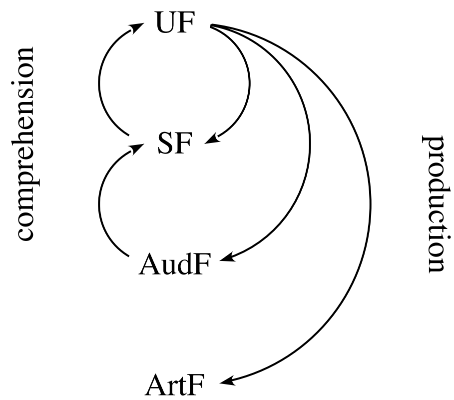
|ynə̲#ʔos|	MAX (C)	DEP (ə)	MAX (ʔ)	*[ʔ̰]	*[ə]	MAX (V̲)	*/CC/	MAX (C̲)
[ynəʔ̰os] → /ynəʔos/				*!	*			
☞ [ynʔ̰os] → /ynʔos/				*!		*	*	
[ynəos] → /ynos/			*!		*	*		
[ynos] → /ynos/			*!			*		

We see that if the perception of [ynʔ̰os] does not change, this form becomes the winner in these 8 percent of the cases. In some percentage of these cases, however, [ynʔ̰os] will also be perceived as /ynos/, so that [ynəʔ̰os] becomes the winner. In this 'stochastic control loop' model, the variation in production reflects the variation in perception. While learnability is no longer a problem, it is also no longer possible to base an obligatory decision on gradient perceptibility, and the ranking *[ə] >> *[ʔ̰] is no longer compatible with a preference for [ynəos] over [ynʔ̰os] (as a result of the independent ranking of the hiatus and creaky pause cues). The degree of listener-orientedness of this model is therefore smaller than that of the model in §4.3.

4.5. The parallel phonological-phonetic production model

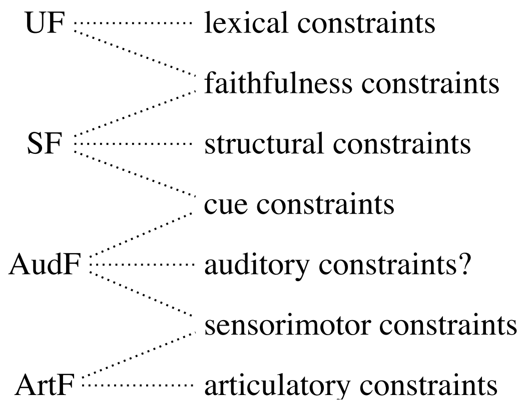
Given the disadvantages of the first four production models in (42), I propose that the remaining model, *phonology and phonetics in parallel*, is closer to the truth. As far as the degree of listener-orientedness is concerned, this model falls in between those of §4.3 and §4.4. A summary is shown in (45).

(45) *A bidirectional model of phonology and phonetics with parallel production*



The picture combines the comprehension and production models. Comprehension consists of two sequential modules, perception and recognition, whereas production consists of a single parallel mapping. The input of production is the usual underlying form, but its output consists of freely combining triplets of ArtF, AudF and SF. The representations are related by constraints that work in two directions, as shown in (46).

(46) *Constraints in bidirectional three-level phonology*



The constraints in (46) are shared with the control-loop models. Four of the constraint types have been discussed before and are used in production: faithfulness constraints evaluate the similarity between UF and SF, structural constraints evaluate SF, articulatory constraints evaluate ArtF, and the same mapping constraints as used in perception evaluate the relation between AudF and SF (this is possible thanks to their negative formulation, see §4.6). This model is thus truly bidirectional, as will be stressed repeatedly below.[25] The relation between ArtF and AudF, finally, is the usual universal sensorimotor mapping and could be described with constraints as well; for simplicity, however, I will asume in the present paper that this relation has been learned perfectly, and I will use a single notation for the two phonetic forms ArtF and AudF. I will ignore any 'auditory constraints', which perhaps militate against loud or unpleasant noises. The 'lexical constraints' are included in (46) for completeness; they play a role in comprehension by influencing the competition between items at UF in lexical access (Boersma 2001), but are irrelevant for production, where the UF is given and fixed.

The application of the parallel production model to the case of *h*-aspiré is handled in the next section.

---

[25] Another bidirectional parallel constraint satisfaction model is Lakoff's (1993) Cognitive Phonology. It works with three levels of representation (M, W and P) that could loosely be identified with UF, SF and a discrete phonetic form. However, the only direction in which this model works deterministically is the production direction. For instance, Lakoff's constraints can handle obligatory vowel insertion and deletion in the UF→SF mapping as well as in the SF→PF mapping, but in the comprehension direction these constraints must express *optional* vowel deletion and insertion, respectively. For instance, Lakoff proposes for Mohawk a constraint that could be written as $|V_1V_2|/V_2/$. In the production direction it obligatorily deletes the first of a sequence of two underlying vowels, e.g. |je+ʌ̃k+hrek+ʔ| → /jʌ̃́khrekʔ/. In the comprehension direction, the same constraint can only express *optional* vowel insertion, i.e. a perceived /jʌ̃k/ must be compatible with |jʌ̃k| as well as with |jeʌ̃k|. The 'constraint' $|V_1V_2|/V_2/$ therefore behaves very much like the *rule* $|V_1V_2| \rightarrow /V_2/$. The deeper cause of the unidirectionality of this constraint is a combination of its positive formulation and its inviolability. The constraint should probably be replaced with a structural constraint */VV/; that it is the first, not the second consonant that deletes, could be handled with a faithfulness ranking $*|V_1V_2|/V_1/ >> *|V_1V_2|/V_2/$, but this would require OT, which was not available to Lakoff.

4.6. A parallel phonological-phonetic account of the variable production of *h*-aspiré

There is variation not only in the perception of *h*-aspiré (§3.4), but as mentioned in §1.6, there is a high degree of variation in its pronunciation as well. I will ignore here the phonetically detailed variability in what the symbol [ʔ̰] stands for, and discuss only the variability with respect to what Meisenburg & Gabriel (2004) regarded as categorical variation, namely the variability in whether something transcribable as [ʔ̰] is present or not, and the variability in whether [ə] is present or not.

The first case of variation to discuss is that between [kɛlʔ̰azaʁ] and [kɛlazaʁ]. According to Dell (1973:256) the pronunciation with the glottal stop is obligatory for some speakers, optional for others.[26] I will discuss here the grammar for those speakers for whom it is optional. Meisenburg & Gabriel propose that Tranel & Del Gobbo would account for this variation (which they do not discuss) with a variable ranking of ALIGN-L(/azaʁ/,σ) and ONSET&NOCODA, which would cause a variation between the forms /.kɛl.a.zaʁ./ and /.kɛ.la.zaʁ./, as can be inferred from our tableau (16). Meisenburg & Gabriel themselves do discuss the variation (by noticing in their data 6 occurrences of [kɛleʁo], from |kɛl#ʔeʁo| 'what a hero'), but do not provide an analysis.[27] In the present framework, the variation can be attributed to a variable ranking of MAX(ʔ) among the phonetic mapping constraints of §3.4. This is shown in tableau (47).

[Tableau (47) is printed on its side on a following page]

In tableau (47) the mapping constraints are ranked at the same height as in §3.4. Their negative formulation (e.g. *[Cʔ̰V]/CʔV/ means "a postconsonantal creaky pause at AudF is not perceived as a glottal stop at SF") allows a bidirectional interpretation (e.g. *[Cʔ̰V]/CʔV/ also means "a glottal stop at SF is not implemented as a postconsonantal creaky pause at ArtF"). Three of the constraints that we know from the production model in §2.4 and §2.5 have been interspersed among the eight mapping constraints. The structural constraint */VV/ has been put at the top, as in §3.4. The rankings of all the constraints are shown along the top of the tableau.

Along the right edge of the tableau we see the resulting frequencies of occurrence in production, if the evaluation noise is 2.0. The most common winning candidate is [kɛlʔ̰azaʁ] /kɛlʔazaʁ/. This is a form that is articulated and heard with a creaky pause and has an SF with a glottal stop. It is important to realize that the two forms are related bidirectionally. From the speaker's standpoint, one can say that the SF /kɛlʔazaʁ/ is phonetically implemented as the ArtF [kɛlʔ̰azaʁ]. From a listener-oriented standpoint, one can say that the speaker thinks that the AudF [kɛlʔ̰azaʁ] will be perceived as /kɛlʔazaʁ/. These two statements are equivalent. A parallel model of phonology and phonetics necessarily comes with this speaker-listener duality. The second grammatical form is [kɛlazaʁ] /kɛlazaʁ/. This form will win if at evaluation time the importance of producing the glottal stop, i.e. the ranking of MAX(ʔ), is less than the articulatory effort of the creaky pause, i.e. the ranking of

[26] According to Tranel (1995), enchainment is forbidden for some lexical items, e.g. in */.kɛ.le.ʁo./ for |kɛl#ʔeʁo| 'what a hero'. However, Meisenburg & Gabriel did observe this form in their corpus, so I will ignore this possible lexical determination of enchainment.

[27] They do analyse the variation between [nɛtʔ̰ɔ̃gʁwa] and [nɛtɔ̃gʁwa] 'clear Hungarian', but attribute it to variation in the lexicon, i.e. |nɛt#(ʔ)ɔ̃gʁwa|, on the basis of the variation between [tu(ʔ̰)ɔ̃gʁwa] and [tutɔ̃gʁwa] for |tu̲t#(ʔ)ɔ̃gʁwa| 'all Hungarian', a phenomenon apparently not found with |tu̲t#ʔazaʁ| (*[tutazaʁ]), which must therefore have an invariable underlying glottal stop.

*[ʔ̰], or less than the perceptibility of a postconsonantal creaky pause, i.e. the ranking of *[Cʔ̰V]/CʔV/. The third possible candidate, which is predicted to occur in 4 percent of the cases, is [kɛlazaʁ] /kɛlʔazaʁ/. This form is interesting, since it did not occur as a candidate in tableau (25). Both of the candidates [kɛlʔ̰azaʁ] /kɛlʔazaʁ/ and [kɛlazaʁ] /kɛlazaʁ/ are *listener-optimal* (the term is by Jäger 2002), i.e. their SF part is the winner (or perhaps most common winner) in the perception tableau where the AudF part is the input. The form [kɛlazaʁ] /kɛlʔazaʁ/ is not listener-optimal. Instead, this looks very much like a speaker who decides to produce /kɛlʔazaʁ/ but manages only to articulate the impoverished [kɛlazaʁ]. If we assume that /kɛlazaʁ/ is also implemented as [kɛlazaʁ], we therefore have a case of neutralization in phonetic implementation. But if we remember the speaker-listener duality, we realize that this phenomenon may be looked upon differently. From the listener-oriented viewpoint, what we have here is a case in which the speaker, although actually saying [kɛlazaʁ], has hallucinated that she has produced the faithful /kɛlʔazaʁ/. The parallel model is thus less listener-oriented than the control-loop model of §2.4 and §2.5. See §4.9 for more discussion.

The second case of variation is that between hiatus and creaky pause. The variation between [leazaʁ] /leʔazaʁ/ and [leʔ̰azaʁ] /leʔazaʁ/ corresponds to that between Meisenburg & Gabriel's forms [tuɔ̃gʁwa], which they observed 3 times in their data, and [tuʔ̰ɔ̃gʁwa], which they observed 5 times. For Meisenburg & Gabriel, forms like [leazaʁ] violate MAX(ʔ), and forms like [leʔ̰azaʁ] violate *VʔV. Gabriel & Meisenburg (2005) model this variation with Stochastic OT by ranking MAX(ʔ) just above *VʔV. In the present framework, the account for the variation must be different, since both [leazaʁ] and [leʔ̰azaʁ] tend to be perceived with a glottal stop. According to §3.4, however, [leʔ̰azaʁ] has even better auditory cues for the underlying glottal stop than [leazaʁ] has, which according to (41) should lead to a glottal stop recoverability of 98 percent for [leʔ̰azaʁ]. The form [leʔ̰azaʁ] now becomes a possible realization if its gain in auditory cues by satisfying *[Vʔ̰V]/VʔV/ approaches the articulatory cost of pronouncing the creaky pause, i.e. *[ʔ̰]. A similar variation is that between [ləazaʁ] and [ləʔ̰azaʁ]. Tableau (48) formalizes this.

[Tableau (48) is printed on its side on a following page]

The choice between [ləazaʁ] and [ləʔ̰azaʁ] is determined by the relative ranking of *[ʔ̰] and *[VV]/VʔV/. Since the former is ranked just above the latter, the hiatus form is slightly more frequent than the form with the creaky pause.

The largest variation is found in the form |ynə#os|, which Meisenburg & Gabriel (2004) report can be pronounced as [ynəos] (twice), [ynʔ̰os] (6 times), and [ynəʔ̰os] (3 times). It is not clear how much of this variation can exist within a single speaker. Since there is a suspicion that some speakers have |yn| as an underlying form (§1.5), I will assume that some of the 6 tokens of [ynʔ̰os] must be ascribed to those speakers, and therefore model a speaker for whom [ynəos] is the preferred pronunciation and [ynʔ̰os] is just a minority pronuncation. Tableau (49) gives the analysis.

[Tableau (49) is printed on its side on a following page]

(47) *Variation between creaky pause and enchainment*

|kɛl#ʔazaʁ|	110.0 */VV/	98.0 */CʔV/ \ / [CV]	97.0 * /V/ \ [Vʔ̰V]	96.5 Max (ʔ)	96.0 */V/ \ [VV]	95.0 */CV/ / \ [Cʔ̰V]	93.0 */CʔV/ | | | [Cʔ̰V]	92.5 *[ə]	92.5 *[ʔ̰]	92.0 */VʔV/ \ / [VV]	91.0 */VʔV/ | | | [Vʔ̰V]	90.0 */CV/ | | [CV]	
[kɛlʔ̰azaʁ] /kɛlazaʁ/				*		*			*				0.2%
☞ [kɛlʔ̰azaʁ] /kɛlʔazaʁ/							*		*				81%
[kɛlazaʁ] /kɛlʔazaʁ/		*											4%
☞ [kɛlazaʁ] /kɛlazaʁ/				*								*	15%

(48) *Variation between creaky pause and hiatus*

|lə̠#ʔazaʁ|	110.0 */VV/	98.0 */CʔV/ \ / [CV]	97.0 * /V/ \ [Vʔ̰V]	96.5 Max (ʔ)	96.0 */V/ \ [VV]	95.0 */CV/ / \ [Cʔ̰V]	93.0 */CʔV/ | | | [Cʔ̰V]	92.5 *[ə]	92.5 *[ʔ̰]	92.0 */VʔV/ \ / [VV]	91.0 */VʔV/ | | | [Vʔ̰V]	90.0 */CV/ | | [CV]	
☞ [ləʔ̰azaʁ] /ləʔazaʁ/								*	*		*		32%
[ləʔ̰azaʁ] /ləazaʁ/	*!			*				*	*				
[ləʔ̰azaʁ] /lazaʁ/			*	*				*	*				
☞ [ləazaʁ] /ləʔazaʁ/								*		*			59%
[ləazaʁ] /ləazaʁ/	*!			*				*					
[ləazaʁ] /lazaʁ/				*	*			*					
[lazaʁ] /lazaʁ/				*								*	9%

(49) *Triple variation*

	110.0	98.0	97.0	96.5	96.0	95.0	93.0	92.5	92.5	92.0	91.0	90.0	
|ynə̠#ʔos|	*/VV/	*/CʔV/ \ / [CV]	* /V/ \ [Vʔ̰V]	MAX (ʔ)	*/V/ \ [VV]	*/CV/ / \ [Cʔ̰V]	*/CʔV/ | | | [Cʔ̰V]	*[ə]	*[ʔ̰]	*/VʔV/ \ / [VV]	*/VʔV/ | | | [Vʔ̰V]	*/CV/ | | [CV]	
☞ [ynəʔ̰os] /ynəʔos/								*	*		*		18%
[ynəʔ̰os] /ynəos/	*!			*				*	*				
[ynəʔ̰os] /ynos/			*	*				*	*				
☞ [ynʔ̰os] /ynʔos/							*		*				32%
[ynʔ̰os] /ynos/				*		*			*				
☞ [ynəos] /ynəʔos/								*		*			45%
[ynəos] /ynəos/	*!			*				*					
[ynəos] /ynos/				*	*			*					
[ynos] /ynʔos/		*											1%
[ynos] /ynos/				*								*	5%

Although *[ə] and *[ʔ̰] are ranked equally high, [ynəos] is preferred over [ynʔ̰os] because the former is favoured by the perceptual mapping constraints. The phonetically enhanced form [ynəʔ̰os] is now possible as well. In (28) it was harmonically bounded by [ynəos], so it could never win, but since it has better auditory cues for the underlying glottal stop than [ynəos] has, the cue constraints make sure that it can win in (49), which therefore ends up showing a three-way variation. The ungrammatical candidate [ynos] scores a very low frequency of occurrence, as required.

All three winning candidates in (49), plus [ynos] /ynos/, are listener-optimal, i.e. they contain an SF part that is the most probable candidate in a perception tableau given the AudF part as an input. These four forms have therefore appeared before in the simple listener-oriented tableaus of §2.4 and §2.5. Six more candidates appear in tableau (49), since AudF and SF can be combined freely. Nevertheless, five of these forms have a frequency below 0.1 percent, although three of them are not harmonically bounded within the set of 11 variably ranked constraints. The only form that occasionally makes it to the surface is [ynos] /ynʔos/, which is a case of a hallucinated glottal stop or of neutralizing phonetic implementation, depending on the direction of your view.

Gabriel & Meisenburg (2005) modelled the variation within Stochastic OT as well, and seemed to find rankings that led to a perfect match between the observed and predicted frequencies of occurrence, i.e. a better match than found in the present section. However, they allowed different rankings for each UF, whereas the presence section assumes the more commonly held view that all forms in the language should derive from a single constraint ranking, so Gabriel & Meisenburg's better match does not indicate that their constraint set is better. Of course, if a grammar model is to be correct, the match between observed and predicted frequencies of occurrence should be within reasonable statistical limits; given the interspeaker variation for [kɛlʔ̰azaʁ] ~ [kɛlazaʁ] and [ynəos] ~ [ynʔ̰os], it probably is in the present case. The thing not explained in this section, if it exists, is any grammaticality difference between [kɛlazaʁ] (14%) and *[lazaʁ] (9%). This is addressed in the next section.

4.7. Improving the ranking: low *[ə]

Now that the point has been proven that a preference of [ynəos] over [ynʔ̰os] need not be due to a ranking of *[ʔ̰] over *[ə], i.e. that it can be due to a difference in auditory cue quality, we are now free to rank these two constraints in such a way that the differential grammaticality of [kɛlazaʁ] and *[lazaʁ] can be accounted for. The trick is to move *[ə] to the bottom of the hierarchy. The frequencies of the resulting forms are listed in (50).

(50) *Predicted frequencies of occurrence when *[ə] is ranked low*

UF	{ ArtF/AudF, SF }	Frequency
|kɛl#ʔazaʁ|	[kɛlʔ̰azaʁ] /kɛlazaʁ/	0.2%
	[kɛlʔ̰azaʁ] /kɛlʔazaʁ/	81%
	[kɛlazaʁ] /kɛlʔazaʁ/	4%
	[kɛlazaʁ] /kɛlazaʁ/	15%
|lə̱#ʔazaʁ|	[ləʔ̰azaʁ] /ləʔazaʁ/	34%
	[ləazaʁ] /ləʔazaʁ/	64%
	[ləazaʁ] /lazaʁ/	0.02%
	[lazaʁ] /lazaʁ/	2%
|ynə̱#ʔos|	[ynəʔ̰os] /ynəʔos/	28%
	[ynʔ̰os] /ynʔos/	8%
	[ynəos] /ynəʔos/	61%
	[ynos] /ynʔos/	0.4%
	[ynos] /ynos/	2%

The illegal forms have indeed shrunk to no more than 2 percent occurrence, while [kɛlazaʁ] still works as before. The preference of [ynəos] over [ynʔ̰os] has become greater than it was in §4.6. In the end, this preference turns out to be partly due to a difference in perceptibility, partly to a difference in articulatory effort. The occasional occurrence of [ynəʔos] remains solely due to the desire to improve perceptibility.

The success of (50) in showing plausible frequencies does not mean that the constraint *[ə] can be discarded with. It still has to outrank MAX (V) and */CC/ to account for the form [ynfam] /ynfam/ in (31). A real improvement on the ranking can be made if we realize that [ə] and [ʔ̰] may not be the articulatorily most effortful vowel and consonant, respectively. In fact, [ə] is likely to be the *easiest* vowel. In all tableaus from §2.5 on, *[ə] could be replaced with the more general *[V], and *[ʔ̰] could be replaced with *[C], thanks to high-ranked faithfulness for non-expendable vowels and consonants. Some inspection teaches us that in that case the constraint */CC/ would become superfluous. This would leave only the two simplest articulatory constraints at ArtF, and if */VV/ is replaced with *[V_1V_2]/V_1V_2/ there would be no structural constraints left at SF (which according to Wheeler & Touretzky 1993 would be advantageous for connectionist modelling). Future theorizing will tell us whether such simplifications are viable.

4.8. Stability by four learning algorithms

It is of some concern how a ranking like that in (49) can be learned. I propose that four learning algorithms help to keep it stable.

The first algorithm to consider is lexicon-driven learning of perception (§3.3; Boersma 1997). If a child's ranking of */VV/ in (35) and (47) is too low, she will perceive an incoming $[\text{ynəos}]_{\text{Aud}}$ as /ynəos/. If she subsequently accesses the correct underlying form |ynə̱#ʔos|, she will consider the UF-faithful candidate at SF, namely /ynəʔos/, to be the correct form that she should have perceived. The gradual learning algorithm will now shift the constraints in such a way that [ynəos] /ynəʔos/ becomes more likely in the future. In the perception tableau (35) we can see that the constraint that prefers [ynəos] /ynəʔos/ to [ynəos] /ynəos/, namely */VV/, will have to rise, and that the constraint with the opposite preference, namely *[VV]/VʔV/, will have to fall. The same can be seen in (47), where we can also see that if comprehension were just as parallel as production, the constraint Max(ʔ) would have to rise as well.

The second algorithm is meaning-driven learning of recognition (Boersma 2001). A higher level than UF, perhaps a representation of the meaning of the sentence such as Logical Form, can correct errors in the SF-to-UF mapping. This is of little relevance to the |ynə̱#ʔos| case and is not elaborated on here.

The third algorithm is learning by virtual production (§4.3; Boersma 1998). If the perception system works well, [ynəos]$_{\text{Aud}}$ will be perceived as /ynəʔos/ and access |ynə̱#ʔos|. If the child's ranking of Max(ʔ) is too low, tableau (47) makes her compute the virtual production [ynos] /ynos/, independently of where *[ə] and *[ʔ̰] are ranked.[1] The comparison of a 'correct' [ynəos] /ynəʔos/ and an 'incorrect' [ynos] /ynos/ will lead to a rise of Max(ʔ) and *[CV]/CV/ and to a fall of *[ə] and *[VV]/VʔV/, as can be seen from (47).

The fourth algorithm has to be learning by self-perception, the mirror-image of learning by virtual production. If Max(ʔ), *[ə] and *[ʔ̰] are ranked too high, the winning form in production, given the underlying form |ynə̱#ʔos|, will be [ynos] /ynʔos/, as can be seen from (47). If the child's ranking of the cue constraints is adultlike, she will subsequently perceive her own [ynos] as /ynʔos/. The comparison of a 'correct' [ynos] /ynos/ and an 'incorrect' [ynos] /ynʔos/ will lead to a rise of *[ə] and *[CV]/CʔV/ and to a fall of Max(ʔ) and *[CV]/CV/, as can be seen from (47).

Together, these learning algorithms seem to lead to a situation in which the cue constraints are ranked optimally for perception and Max(ʔ) is ranked neither too low nor too high. A solid proof that this optimization really works would involve a full-language computer simulation that is far beyond the scope of the present paper.

4.9. Assessment

The parallel model seems to be more listener-oriented than the stochastic control-loop model. Like the non-stochastic control-loop model with probabilistic faithfulness constraints, the parallel model could, for instance, account for an obligatory [ynəos] /ynəʔos/, even if the ranking is *[ə] >> *[ʔ̰]. If one multiplies all the ranking values in (49) by 100, the form [ynəos] /ynəʔos/ will always win.[2] Such an obligatoriness is also possible with the probabilistic-faithfulness model, as tableau (43) shows, but not with the stochastic control-loop model, which cannot do better than the speaker-based ranking *[ʔ̰] >> *[ə] in (28).

Since the parallel model does not have the learnability problems of the probabilistic-faithfulness model and does not have the listener-orientation problems of the stochastic control-loop model, the parallel model seems to be preferable.

---

[1] Even if the two articulatory constraints are ranked low, [ynos] /ynos/ is the perceptually least ambiguous candidate, i.e. it has the least violating relation between auditory and surface form. Such considerations might explain infants' preferences for 'unmarked' forms, such as those observed by Davidson, Smolensky & Jusczyk (2004).

[2] I may look problematic to propose that the rankings in (49) could be further apart; after all, the relative rankings of the cue constraints are based on variability in perception. However, the variability in perception is the result of two sources: noise in the communication channel and variability in the ranking of the cue constraints. The latter variability may therefore be smaller than that proposed in §3.4. An alternative possibility worth pursuing is that the evaluation noise could be smaller in production than in perception.

5. Conclusion

All speaker-based accounts of *h*-aspiré (§1.6, §2.2, §2.3) are observationally more or less adequate, but require a special trick to account for [ynəos]. Descriptive adequacy has been improved by introducing an Optimality-Theoretic listener-oriented account, in which the [ə] in [ynəos] ultimately falls out naturally as a means to improve the perceptibility of an underlying |ʔ| (§2.4, §2.5, §3.4, §4.3, §4.4, §4.6). Explanatory adequacy has partly been achieved by discussing the acquisition of the constraint ranking in perception (§3.3) and in production (§4.8).

The present paper does not provide the final answer to the question of *h*-aspiré. Questions remaining to be answered include the following.

Is *h*-aspiré better represented as an underlying glottal stop or as an underlying syllable boundary (§2.2)? With Tranel's view of *h*-aspiré as an underlying syllable boundary, the explanation for the surfacing of schwa in *une hausse* would be equally listener-oriented as the one presented here: in order to get the underlying syllable boundary across to the listener, the speaker can implement hiatus, a glottal stop, or creak on the vowel. The constraints would only change slightly; instead of MAX(ʔ), for instance, we would have the constraint CONTIG(.V); the constraint */VV/ would be formulated naturally as "no multiple vowels within a syllable" (this may contribute to its high ranking). The empirical advantage of the syllable-island approach would be that it could immediately explain the case of initial neutralization (§1.1), since every phrase boundary must necessarily be a syllable boundary. It would also account for other phenomena in which *h*-aspiré words act as if they start with a vowel, namely the fact that their first syllable cannot contain a schwa (this point is not very strong, since most Germanic languages do have an uncontroversial /h/ but still do not allow it to coexist with schwa in a syllable) and the fact that the first syllable is skipped for purposes of reduplication in hypocoristics (for an overview, see Tranel 1995). A disadvantage of the syllable-island approach would be that it cannot explain cases where *h*-aspiré is realized as a glottal stop but not as a syllable boundary; as a case in point, Meisenburg & Gabriel (2004) mention ".tʁwaˀɑ.bœʁ.gœʁ." for |tʁwaz̲#ʔɑ̃bœʁgœʁz̲| 'three hamburgers', in which the glottal stop ends up in the middle of a monosyllabic diphthong. In the end, a whole-language computer simulation of the acquisition of the phonology of French may provide an answer to this question.

Can the high ranking of */VV/ be explained if the high frequency of underlying |V#V| is taken into account? Will a perception experiment such as that described in footnote 23 confirm the predictions based on discrete language data? Will a production experiment confirm the prediction that the creaky pause in intervocalic position is limited to cases of underlying *h*-aspiré?

The last question is which of the grammar models is correct. All three listener-oriented models can handle the observed facts of *h*-aspiré. The stochastic control-loop model, however, seems to have problems with its degree of listener orientation, and the model with probabilistic faithfulness seems to have problems with learnability. The model of *phonology and phonetics in parallel* looks most promising. Apart from enhancement, it seems to be able to handle production phenomena such as incomplete neutralization, licensing by cue, and counterbleeding opacity, in ways that require fewer stipulations than former grammar models. These matters must be left for future research.

References

Apoussidou, Diana, & Paul Boersma (2004). Comparing two Optimality-Theoretic learning algoithms for Latin stress. *WCCFL* **23**: 29–42.

Bagemihl, Bruce (1995). Language games and related areas. In John A. Goldsmith (ed.) *The handbook of phonological theory*. Cambridge, Mass. & Oxford: Blackwell. 697–712.

Bally, Charles (1944). *Linguistique générale et linguistique française*. Second edition. Berne: Francke.

Beckman, Jill N. (1998). *Positional faithfulness.* Doctoral thesis, University of Massachusetts, Amherst.

Best, Catherine T. (1995). A direct realist view of cross-language speech perception. In Winifred Strange (ed.) *Speech perception and linguistic experience: theoretical and methodological issues*. Baltimore: York Press. 171–203.

Boersma, Paul (1989). Modelling the distribution of consonant inventories by taking a functionalist approach to sound change. *Proceedings Institute of Phonetic Sciences Amsterdam* **13**: 107–123.

Boersma, Paul (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences* **21**: 43–58. University of Amsterdam.

Boersma, Paul (1998). *Functional phonology*. PhD thesis, University of Amsterdam.

Boersma, Paul (2000). The OCP in the perception grammar. *Rutgers Optimality Archive* **435**.

Boersma, Paul (2001). Phonology-semantics interaction in OT, and its acquisition. In Robert Kirchner, Wolf Wikeley, & Joe Pater (eds.) *Papers in Experimental and Theoretical Linguistics*. Vol. 6. Edmonton: University of Alberta. 24–35. [*Rutgers Optimality Archive* **369**, 1999]

Boersma, Paul (2003a). The odds of eternal optimization in Optimality Theory. In D. Eric Holt (ed.) *Optimality Theory and Language Change*. Dordrecht: Kluwer. 31–65.

Boersma, Paul (2003b). Overt forms and the control of comprehension. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. Department of Linguistics, Stockholm University. 47–56. [*Rutgers Optimality Archive* **597**]

Boersma, Paul, & Silke Hamann (2005). The violability of backness in retroflex consonants. *Rutgers Optimality Archive* **713**.

Boersma, Paul & Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* **32**: 45–86.

Chao, Yuen-ren (1934). The non-uniqueness of phonemic solutions to phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica*, Vol. IV, Part 4, 363–397. [Reprinted in Martin Joos (ed., 1958), *Readings in linguistics: the development of descriptive linguistics in America since 1925*, Second edition. New York: American Council of Learned Societies. 38–54]

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

Clements, George N. (1988). The role of the sonority cycle in core syllabification. *Working Papers of the Cornell Phonetics Laboratory* **2**: 1–68. [not seen]

Clements, George N., & Samuel J. Keyser (1983). *CV Phonology: a generative theory of the syllable*. Cambridge, Mass.: MIT Press.

Davidson, Lisa, Paul Smolensky, & Peter Jusczyk (2004). The initial and final states: theoretical implications and experimental explorations of Richness of the Base. In René Kager, Joe Pater & Wim Zonneveld (eds.) *Fixing priorities: constraints in phonological acquisition*. Cambridge: Cambridge University Press.

Dell, François (1973). *Les règles et les sons*. Paris: Hermann.

Encrevé, Pierre (1988). *La liaison avec et sans enchaînement*. Paris: Seuil.

Escudero, Paola, & Paul Boersma (2003). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In Sudha Arunachalam, Elsi Kaiser & Alexander Williams (eds.): *Proceedings of the 25th Annual Penn Linguistics Colloquium. Penn Working Papers in Linguistics* **8.1**: 71–85. [Rutgers Optimality Archive **439**, 2001]

Escudero, Paola, & Paul Boersma (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* **26**: 551–585.

Flemming, Edward (1995). *Auditory representations in phonology*. PhD thesis, UCLA.

Flemming, Edward (2003). The relationship between coronal place and vowel backness. *Phonology* **20**: 335–373.
Fouché, Pierre (1959). *Traité de prononciation française*. Paris: Klincksieck. [2nd edition, 1959; printing 1969]
Fowler, Carol A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* **14**: 3–28.
Fromkin, Victoria (1971). The non-anomalous nature of anomalous utterances. *Language* **47**: 27–52.
Gabriel, Christoph, & Trudel Meisenburg (2005). Silent onsets? An optimality-theoretic approach to French h aspiré words. Poster presented at OCP 2, Tromsø, January 20–22. [*Rutgers Optimality Archive* **709**]
Grammont, Maurice (1948). *Traité pratique de prononciation française*. Paris: Delagrave.
Hall, Robert (1948). *French*. *Language* monograph **24** (*Structural skech* **1**).
Halle, Morris (1959). *The sound pattern of Russian*. The Hague: Mouton.
Hayes, Bruce (1999). Phonetically-driven phonology: the role of Optimality Theory and Inductive Grounding. In Michael Darnell, Edith Moravcsik, Michael Noonan, Frederick Newmeyer & Kathleen Wheatley (eds.) *Functionalism and Formalism in Linguistics*, Vol. I: *General Papers*. Amsterdam: John Benjamins. 243–285. [*Rutgers Optimality Archive* **158**, 1996]
Horwood, Graham (2002). Precedence faithfulness governs morpheme position. In Line Mikkelsen & Christopher Potts (eds.) *Proceedings WCCFL* **21**: 166–179. Somerville: Cascadilla.
Hyman, Larry (1985). *A theory of phonological weight*. Dordrecht: Foris.
Jäger, Gerhard (2003). Learning constraint sub-hierarchies: The Bidirectional Gradual Learning Algorithm. In Henk Zeevat & Reinhard Blutner (eds.) *Optimality Theory and pragmatics*. Basingstoke: Palgrave Macmillan. 251–287.
Jong, Daan de (1990). On floating consonants in French. *Western Conference on Linguistics* 21.
Jun, Jongho (1995). Place assimilation as the result of conflicting perceptual and articulatory constraints. *Proceedings WCCFL* **14**: 221–237.
Kirchner, Robert (1998). *Lenition in phonetically-based Optimality Theory*. PhD dissertation, UCLA.
Lahiri, Aditi & William Marslen-Wilson (1991). The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition* **38**: 245–294.
Lakoff, George (1993). Cognitive phonology. In John Goldsmith (ed.) *The last phonological rule: reflections on constraints and derivations*. Chicago & London: The University of Chicago Press. 117–145.
Levelt, Willem (1989). *Speaking: from intention to articulation*. Cambridge, Mass.: MIT Press.
McCarthy, John J. (2003). OT constraints are categorical. *Phonology* **20**: 75–138.
McCarthy, John, & Alan Prince (1995). Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*. University of Massachusetts Occasional Papers **18**. Amherst, Mass.: Graduate Linguistic Student Association. pp. 249–384.
McQueen, James M., & Anne Cutler (1997). Cognitive processes in speech perception. In William J. Hardcastle & John Laver (eds.) *The handbook of phonetic sciences*. Oxford: Blackwell. 566–585.
Meisenburg, Trudel, & Christoph Gabriel (2004). Silent onsets? The case of French h aspiré words. Talk presented at workshop *Phonetik und phonologie 1*, Potsdam, June 19.
Pater, Joe (2004). Bridging the gap between receptive and productive development with minimally violable constraints. In René Kager, Joe Pater & Wim Zonneveld (eds.) *Constraints in phonological acquisition*. Cambridge: Cambridge University Press. 219–244.
Prince, Alan, & Paul Smolensky (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Technical Report TR-**2**, Rutgers University Center for Cognitive Science.
Prunet, Jean-François (1986). *Spreading and locality domains in phonology*. PhD thesis, McGill.
Schane, Sanford (1968). *French phonology and morphology*. Cambridge, Mass.: MIT Press.
Selkirk, Elisabeth, & Jean-Roger Vergnaud (1973). How abstract is French phonology? *Foundations of Language* **10**: 249–254.
Steriade, Donca (1995). Positional neutralization. Two chapters of an unfinished manuscript, UCLA.

Tesar, Bruce (1997). An iterative strategy for learning metrical stress in Optimality Theory. In Elizabeth Hughes, Mary Hughes, & Annabel Greenhill (eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development*, 615–626. Somerville, Mass.: Cascadilla.

Tesar, Bruce (1998). An iterative strategy for language learning. *Lingua* **104**: 131–145.

Tesar, Bruce (1999). Robust interpretive parsing in metrical stress theory. In Kimary Shahin, Susan Blake & Eun-Sook Kim (eds.) *Proceedings WCCFL* **17**: 625–639. Stanford, Calif.: CSLI.

Tesar, Bruce, & Paul Smolensky (1998). Learnability in Optimality Theory. *Linguistic Inquiry* **29**: 229–268.

Tesar, Bruce, & Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, Mass.: MIT Press.

Tranel, Bernard (1974). *The phonology of nasal vowels in Modern French*. Doctoral thesis, UC San Diego.

Tranel, Bernard (1987). *The sound of French: an introduction*. Cambridge: Cambridge University Press.

Tranel, Bernard (1995). Current issues in French phonology. In John A. Goldsmith (ed.) *The handbook of phonological theory*. Cambridge, Mass. & Oxford: Blackwell. 798–816.

Tranel, Bernard (1996). French liaison and elision revisited: a unified account within Optimality Theory. In T. Parodi, C. Quicoli, M. Saltarelli & M.L. Zubizarreta (eds.) *Aspects of Romance linguistics*. Washington: Georgetown University Press. 433–455. [*Rutgers Optimality Archive* **15**, 1994]

Tranel, Bernard, & Francesca del Gobbo (2002). Local conjunction in Italian and French phonology. In Caroline R. Wiltshire & Joaquim Camps (eds.). *Romance phonology and variation*. Amsterdam: John Benjamins. 191–218.

Wheeler, Deirdre W., & David S. Touretzky (1993). A connectionist implementation of cognitive phonology. In John Goldsmith (ed.) *The last phonological rule: reflections on constraints and derivations*. Chicago & London: The University of Chicago Press. 146–172.