



# Some listener-oriented accounts of *h*-aspiré in French

Paul Boersma

*University of Amsterdam, Phonetic Sciences, Spuistraat 210, Room 303, 1012VT Amsterdam, Netherlands*

Received 19 April 2005; received in revised form 10 November 2006; accepted 10 November 2006

Available online 1 August 2007

---

## 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 for *bidirectional phonology and phonetics*, which uses a single constraint set for the four processes of perception, recognition, phonological production, and phonetic implementation, and in which phonological and phonetic production are evaluated in parallel. 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.

© 2007 Elsevier B.V. All rights reserved.

*Keywords:* Phonology; Phonetics; French; Acquisition; Learnability

---

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), whereas section 3 shows that the simplest form of listener-oriented OT already accounts for the preference of [ynəos] over [ynos]. Section 4 then formalizes the listener's behaviour, so that section 5 can formalize the speaker's listener-orientedness in more detail, finally leading in section 6 to an accurate account for the preference of [ynəos] over [ynʔos] and for the occasional occurrence of a phonetically enhanced fourth form, namely [ynəʔos]. The remaining

---

*E-mail address:* [paul.boersma@uva.nl](mailto:paul.boersma@uva.nl).

sections address some of the related additional complexities of French phonology. The conclusion will be that a bidirectional model of phonology and phonetics that uses the same elements for modelling production as for modelling comprehension accounts for the observed facts in the most straightforward, least ad-hoc, and least explicitly teleological way.

## 1. H-aspiré: facts and explanations

This section presents the facts about *h*-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 at least three segment types: consonants, vowels, and a third type called *h*-aspiré (this list ignores the two types of glide-initial words, which are discussed in section 6.4). The three 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:

|gɑ̃sɔ̃| ‘boy’, |fam| ‘woman’

*h*-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 sections 1.2, 2.2 and 7, I discuss a variety of alternative representations, several of which do not involve a separate segment for *h*-aspiré. 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 main case in which *h*-aspiré acts in the same way as a vowel is the case of postpausal neutralization. At the beginning of a sentence or after a pause caused by a syntactic boundary or by any other need to separate the word from the preceding words, vowel-initial and *h*-aspiré-initial words are pronounced identically, probably with a weak attack or a glottal stop (Dell, 1973:79; Cornulier, 1981:218; Tranel, 1981:299). After a pause, therefore, we have the pronunciations [ʔazaʁ], [ʔos], [ʔɔm], [ʔide], and the underlying contrast between *h*-aspiré and a vowel is completely neutralized phonetically. 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, for instance, to the final vowels |ə| and |a| in the singular of the definite article, whose underlying forms can be written as |lə| ‘the-MASC SG’ and |la| ‘the-FEM SG’.

Thus, |l̥#ɔm| ‘the man’ is pronounced [lɔm], and |la#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 (“#” denotes a phrase-internal word boundary).

Elision does not apply before consonant-initial words, so |l̥#gaks̥s̥| ‘the boy’ is pronounced [l̥gaks̥s̥], and |la#fam| ‘the woman’ is pronounced [lafam]. Crucially, *h*-aspiré-initial words join with consonant-initial words here: |l̥#ʔazax̥| ‘the coincidence’ is pronounced [l̥azax̥], and |la#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. For [Martinet \(1933\)](#), [Chao \(1934 \[1958:46\]\)](#), [Schane \(1968:8\)](#) and [Selkirk and Vergnaud \(1973\)](#) it is an underlying fricative [h] or even [x] (Chao writes an underlined h to indicate its deletability). For reasons discussed below in section 1.3, [Dell \(1973:256\)](#) argues explicitly for the plosive [ʔ], and [Meisenburg and Gabriel \(2004\)](#) for the creaky plosive [ʔ̥]. 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 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*

l̥#gaks̥s̥		l̥gaks̥s̥		l̥gaks̥s̥
l̥#ʔazax̥	elision	l̥ʔazax̥	*ʔ/V_V	l̥azax̥
l̥#ɔm	→	lɔm	→	lɔm

In this scenario, elision is formalized as “delete V (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 intervocally (the reason for this restriction will become clear in section 1.3). The rule order in (2) was proposed by [Schane \(1968:7\)](#) and [Dell \(1973:253,257\)](#).

Beside the segmental solutions just discussed, *h*-aspiré has often been accounted for in terms of prosodic elements, most often with reference to syllable structure. This derives from the general agreement that [lɔm] constitutes a single syllable, so that the segments of the article and those of the noun share a syllable, whereas [l̥azax̥] consists of three syllables, so that the article and the noun do not share any syllables. All prosodic proposals try to express the idea that a vowel-initial word allows a preceding consonant to intrude into its first syllable and that an *h*-aspiré word tries to keep that consonant out of its first syllable. [Martinon \(1905 \[1962:16\]\)](#) states and [Damourette and Pichon \(1911–1927:I:207\)](#) explicitly argue that *hiatus* is a necessary requirement of *h*-aspiré words, i.e. that these words have to be preceded not only by a syllable boundary but by a vowel as well. Although such a claim is correct for the elision case (and for liaison, section 1.4), [Cornulier \(\[1974\] 1981:210\)](#), for reasons discussed in section 1.3, proposes a somewhat lighter restriction, namely that *h*-aspiré words are immediately preceded by a syllable boundary. Formalizing Cornulier’s idea, [Schane \(1978:141\)](#) proposes that *h*-aspiré words contain an underlying syllable boundary, i.e. that our example words are represented underlyingly as |.azax̥| versus |ɔm|, where |.| stands for a syllable boundary. Cornulier’s and Schane’s proposal can be refined within more hierarchical versions of nonlinear phonology; for

instance, Clements and Keyser (1983:108) propose that *h*-aspiré words start with an empty consonant slot whereas vowel-initial words start with no consonant slot, and Encrevé (1988:197) credits Jean-Roger Vergnaud for proposing that *h*-aspiré words start with an empty onset whereas vowel-initial words start with a null onset.

For the present paper I note that the segmental solutions from Martinet to Dell and the syllable-boundary solutions by Cornulier and Schane (also Tranel (1995, 1996)) are nearly equivalent to each other as far as Optimality-Theoretic solutions are concerned (any differences are discussed in sections 7 and 8.2).<sup>1</sup> For establishing my points about listener-orientedness, it hardly matters whether *h*-aspiré is accounted for in terms of segmental or syllable structure. For simplicity reasons, I will generally write it in terms of an underlying glottal stop, but I will freely refer to statements by proponents of the syllable boundary theory, most notably Cornulier (1981), whose account comes closer than any other previous accounts to the one presented here.

### 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#ɔ̃m| ‘which man, what a man’ is pronounced [.kɛ.lɔ̃m.], while |kɛl#gɑ̃sɔ̃| ‘which boy’ becomes [.kɛ.l.gɑ̃.sɔ̃.]. Crucially, *h*-aspiré words again act like consonant-initial words: |kɛl#ʔazaʁ| ‘what a coincidence’ becomes [.kɛ.l.ʔa.zaʁ.]. As with the elision case of section 1.2, we can give a derivational account, as in (3).<sup>2</sup>

#### (3) Derivation of enchainment

kɛl#gɑ̃sɔ̃		.kɛ.l.gɑ̃.sɔ̃.		.kɛ.l.gɑ̃.sɔ̃.
kɛl#ʔazaʁ	enchain	.kɛ.l.ʔa.zaʁ.	*ʔ/V_V	.kɛ.l.ʔa.zaʁ.
kɛl#ɔ̃m	→	.kɛ.lɔ̃m.	→	.kɛ.lɔ̃m.

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 sections 1.2 and 1.4 and Grammont limits himself to the schwa drop case of section 1.5, so it is not clear that these authors had the present postconsonantal case in mind when expressing their generalization. According to Hall (1948:10), “The phoneme /ʔ/ involves slight faucal constriction, with renewed syllable onset and optional glottal stop.” Since a faucal constriction can be associated with glottal

<sup>1</sup> I also ignore analyses with rule features (Tranel, 1981), which were anticipated and rejected by Selkirk and Vergnaud (1973), although Tranel (1995) again provides some evidence in favour of them.

<sup>2</sup> The order of the two rules is immaterial here. However, enchainment does not always stand by itself as here: the processes of elision (section 1.2) and liaison (section 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.

creak, Hall's description suggests the phonetic transcriptions [kɛl\_azaʁ] and [kɛl\_?azaʁ]. According to Meisenburg and 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\_?azaʁ], [kɛl\_?azaʁ]). Following Meisenburg and Gabriel, I will summarize these five types of phonetic realizations (and all their mixtures) under the umbrella transcription [kɛl\_?azaʁ]. For Dell (1973:256) as well, the best phonetic transcription is [ʔ].

It has to be noted here that Cornulier (1981:218) and Tranel (1981:299,310–311) regarded the postconsonantal *h*-aspiré pause, including the glottal stop variant, as identical to e.g. the citational pause in [lɔ̃vɛʁbʔaʃte] for [lɔ̃#vɛʁb#aʃte] 'the verb *to buy*' or to cases of liaison without enchainment such as [sɛʔemusmɑ̃] 'that stupefaction' (forms described phonetically by Malécot (1975)). Cornulier therefore simply wrote "/kɛl-azaʁ/". Unfortunately, there seems to be no phonetic evidence to corroborate Cornulier's claim that the two kinds of postconsonantal pauses are phonetically identical, and conversely, neither Hall (1948) nor Meisenburg and Gabriel (2004) explicitly say that postconsonantal creaky pauses (other than postvocalic creaky pauses, see section 1.5) are confined to *h*-aspiré cases.

For the purposes of section 1 and the speaker-based view of section 2.1, I will simply write the surface forms as "kɛlɔ̃" versus "kɛl\_?azaʁ", ignoring for a moment both syllable structure and other kinds of pauses (which are taken up in section 7).

Enchainment is often optional. According to Dell (1973:256) and most other authors, the pronunciation with the glottal stop is obligatory for some speakers, optional for others. Enchained realizations such as [kɛlazaʁ] are discussed in section 6.1. A lexical split for some speakers mentioned by Cornulier (1981), namely that some words, such as [ʔɛʁo] 'hero', do not allow enchainment, whereas most words, such as [ʔazaʁ], do, is discussed in section 6.3.

A final possibility is the pronunciation [kɛləazaʁ], with an inserted schwa. Damourette and Pichon's *assurance d'hiatus* explicitly requires the surfacing of this schwa, and these authors in fact mention [lɔ̃jɔ̃] for [lɔ̃#œj#ʔõ] 'the eye high' (1911–1927:I:207) and [ilɛ:l] for [il#ɛl] 'he hails' (1911–1940:VI:279). Damourette and Pichon do note that this inserted schwa occurs only in colloquial French, which is the variety they want to describe. The finding extends to more recent observations. According to Tranel (1981:287), forms like [kɛləazaʁ] were observed by himself and by Y. Morin. However, Cornulier (1981:211) restricts forms like [ʔkɛləazaʁ] to some speakers who diverge from "le bon usage officiel" that prohibits schwa to surface where it is non-underlying. Tranel (1981), not accepting underlying schwas in the first place, predicts global schwa insertion and ascribes non-insertion cases to orthography (although his own example *homme* 'man' on p. 308, which hardly allows its final schwa to surface, contradicts this). Since Tranel (1995) appears to have changed his mind and solely gives forms like /kɛl.a.zaʁ/, without schwa insertion, and Meisenburg and Gabriel (2004) did not observe any forms like [ʔkɛləazaʁ] in their experiments, the present article will stay with the form [kɛl\_?azaʁ], noting at the same time that speakers who insert non-underlying schwas are just as easy to model as speakers who do not (section 6.3).

#### 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 |lez|, where (following Chao, 1934 [1958:46]) the underlining again marks expendability. Thus, |lez#ɔm| ‘the men’ and |lez#ide| ‘the ideas’ are pronounced [lezɔm] and [lezide], while |lez#gɑksɔ̃| ‘the boys’ and |lez#fam| ‘the women’ are pronounced [legɑksɔ̃] and [ləfam]. Crucially again, *h*-aspiré words act like consonant-initial words: |lez#ʔazɑk| ‘the perils’ and |lez#ʔos| ‘the rises’ are pronounced [lezazɑk] and [leos] (as mentioned by all the works cited above in section 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*

lez#gɑksɔ̃		legɑksɔ̃		legɑksɔ̃
lez#ʔazɑk	liaison	leʔazɑk	*ʔ/V_V	lezazɑk
lez#ɔm	→	lezɔm	→	lezɔm

In this scenario, liaison is formalized as “delete C (i.e. any expendable consonant) if followed by a consonant”.<sup>3</sup> *H*-aspiré will then block liaison because it is still a consonant at the stage where liaison tries to apply. Within the serial 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 and 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.

### 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 |ynɔ̃|. Thus, |ynɔ̃#fam| ‘a woman’ is usually pronounced [ynfam] (sometimes [ynɔ̃fam]), and |ynɔ̃#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 (section 1.2), which is fully obligatory within a phrase (Dell, 1973:203).<sup>4</sup> 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 [ynɔ̃fam] 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”.

<sup>3</sup> These expendable consonants are also deleted at the end of a phrase and at several other boundaries. This is largely ignored in the present article.

<sup>4</sup> Not everybody agrees that elision is involved. Cornulier (1981:192–193) calls it the non-application of the ‘right of schwa’, and Tranel (1981:282–294), who starts from an underlying schwa-less [yn], would conversely propose that any schwa in [ynɔ̃fam] has to have been inserted, probably on the basis of orthography.

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: [ynə#ʒos] ‘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 [ynʒos]. 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 [setʒos], its underlying form must be [set#ʒos] (cf. section 1.3), and since ‘that rise’ is pronounced [setəos], its underlying form must be [setə#ʒos].<sup>5</sup>

The schwa in these *h*-aspiré forms is often enhanced by an additional glottal stop (Encrevé, 1988:198). Thus, the recordings by Meisenburg and Gabriel (2004) and Gabriel and Meisenburg (2005) contain two tokens of the form [ynəʒos], with both a schwa and a creaky pause. They interpret such forms as an enhancement of the ‘norm’ pronunciation [ynəos] by additional phonetic material that the speaker introduces in order to avoid a simple hiatus (i.e. the sequence of two vowels in [əo]). The explicit or implicit claim by Encrevé and Meisenburg and Gabriel is that such creaky pauses only occur in *h*-aspiré cases (such as the recorded [ynəʒos] and [tuʒɔŋkwa] ‘all Hungarians’) and do not occur in the more common cases of hiatus, say in words like [teatr] ‘theatre’ or before a non-*h*-aspiré word as in [saika] ‘that will go’. Indeed, Grammont (1948:124) makes a big point of asserting that in sentences like *Cette soirée a eu un succès fou* ‘this evening has been a crazy success’ one hears a sequence [eayœ] uninterrupted by any of the glottal stops that Germans or Russians would insert. The existence of the enhanced intervocalic form could be an argument against a syllable-boundary approach to *h*-aspiré and in favour of a segmental approach. The present paper discusses non-*h*-aspiré hiatus in sections 3.1, 3.2, 7.2, and especially section 8.

It is possible that for some speakers [ynʒos] is the only option, perhaps because for them the underlying form of the article is simply [yn].<sup>6</sup> For these speakers, the triple distinction noted in this section will still exist for words with an uncontroversial underlying schwa, such as the article [lə], 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#gəʁsɔ̃] ‘here is the boy’ will be usually pronounced [vwasilgəʁsɔ̃] as a result of preconsonantal schwa drop, and, crucially, [vwasi#lə#ʒazaʁ] ‘here is the coincidence’ will be obligatorily pronounced [vwasiləzaʁ], never \*[vwasilʒazaʁ]. Perhaps because of their lower degree of variation, forms like these were discussed instead of [ynəos]-like forms by Tranel (1996) and Tranel and 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 and Keyser, 1983:113). And indeed, if one orders preconsonantal schwa drop (as elision and liaison) before intervocalic

<sup>5</sup> Tranel (1974:110; 1981:282–294) did not consider this particular contrast very regular, claiming that a schwa can just as easily be inserted in [kələzaʁ] or [setəos] ‘seven rises’ as in [setəos] ‘that rise’. However, Meisenburg and Gabriel (2004) observed feminine forms like [kələʁ] (with [kələ]) but no masculine forms like [kələzaʁ] (with [kəl]) in their experiments. See section 1.3 for details.

<sup>6</sup> Having the underlying form [yn] is no guarantee that the pronunciation will be [ynʒos]. Tranel (1981:282–294) states that the underlying form is generally [yn], and that a schwa is often inserted, yielding [ynəos].

glottal stop deletion, |ynə#ʔos| will be pronounced [ynʔos], analogously to [kɛʔazar]. While the pronunciation [ynʔos] is attested (e.g. Meisenburg and Gabriel, 2004; see also section 6.1), the form [ynəos] is usual (it is in fact the only form mentioned by Dell (1973) and Tranel (1995)) and has to be explained.

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

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

The process of schwa drop in (5) can be formalized as “delete schwa before a consonant”. The reason why *h*-aspire 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 |ynə#ʔos| would be [ynos]. The conditioning by a preceding VC has to prevent application to forms like [ləgɑʁsɔ̃]. 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 and Del Gobbo (2002) and Meisenburg and Gabriel (2004), and are discussed in sections 2.3 and 2.4, respectively.

1.6. A derivational account that works but is conspiratorial

The 12 derivations are summarized in (6).

(6) Correct derivation of all forms without syllabification

UF	elision *V/_V	liaison *C/_C	*ʔ/V_V	*ə/VC_C	SF (see section 3.1)
lə#gɑʁsɔ̃	ləgɑʁsɔ̃	ləgɑʁsɔ̃	ləgɑʁsɔ̃	[ləgɑʁsɔ̃]	/ləgɑʁsɔ̃/
lə#ʔazar	ləʔazar	ləʔazar	<b>ləazar</b>	[ləazar]	/ləʔazar/
lə#ɔm	ləm	ləm	ləm	[ləm]	/ləm/
kɛl#gɑʁsɔ̃	kɛlgɑʁsɔ̃	kɛlgɑʁsɔ̃	kɛlgɑʁsɔ̃	[kɛlgɑʁsɔ̃]	/kɛlgɑʁsɔ̃/
kɛl#ʔazar	kɛlʔazar	kɛlʔazar	<b>kɛlʔazar</b>	[kɛlʔazar]	/kɛlʔazar/
kɛl#ɔm	kɛləm	kɛləm	kɛləm	[kɛləm]	/kɛləm/
ləz#gɑʁsɔ̃	ləzgɑʁsɔ̃	ləgɑʁsɔ̃	ləgɑʁsɔ̃	[ləgɑʁsɔ̃]	/ləgɑʁsɔ̃/
ləz#ʔazar	ləzʔazar	ləʔazar	<b>ləazar</b>	[ləazar]	/ləʔazar/
ləz#ɔm	ləzɔm	ləzɔm	ləzɔm	[ləzɔm]	/ləzɔm/
ynə#fam	ynəfam	ynəfam	ynəfam	<b>[ynfam]</b>	/ynfam/
ynə#ʔos	ynəʔos	ynəʔos	<b>ynəos</b>	[ynəos]	/ynəʔos/
ynə#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*-aspire. Going from the underlying form (UF) towards the phonetic form in the fifth column, the early obligatory rules \*V/\_V and \*C/\_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  $*\text{ʔ}/V\_V$  and  $*\text{ə}/VC\_C$  are mutual bleeders and have to be ordered as here. The forms in the sixth column (‘perceived phonological surface forms’) are discussed in section 3.1.

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 and Gabriel (2004) report the pronunciations [ynəos], [ynəʔos], and [ynʔos], and forms like [ləʔazaʃ] and [ləʔazaʃ] can be predicted on the basis of their observed pronunciations [tuʔʒgɔwa] for [tuʔʒgɔwa] ‘all Hungarian’ and [tʁwaʔɒbœʒgœʃ] for [tʁwaz#ʔɒbœʒgœʃ] ‘three hamburgers’. This means that the glottal stop deletion rule in (6) is optional. However, this optionality cannot predict the observed form [ynəʔos]: if glottal stop deletion does not apply, the resulting intermediate form “ynəʔos” would be eligible for preconsonantal schwa drop and surface as [ynʔos]. The form [ynəʔos] could only surface if schwa drop were optional as well, but that would predict a frequent occurrence of forms like [ynəfam]. If Meisenburg and Gabriel are correct in suggesting that [ynəʔos] 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  $*\text{ʔ}$  (“delete glottal stop”) and  $*\text{ə}$  (“delete schwa”), the environmental conditions in the rules  $*\text{ʔ}/V\_V$  and  $*\text{ə}/VC\_C$  change this picture. Why does glottal stop delete only intervocally, 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: 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). Together, the  $V\_$  restriction in the rule  $*\text{ʔ}/V\_V$  and the  $\_C$  restriction in the rule  $*\text{ə}/VC\_C$  seem to *conspire* to keep an unexpected part of the underlying form alive: [ʔ] in the case of [kɛʔʔazaʃ], [ə] in the case of [ynəos]. Conspiracies like these formed the original criticism that led Prince and 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*-aspire within an Optimality-Theoretic framework should make exclusive use of universally defensible constraints. In section 2 I show that speaker-based Optimality-Theoretic accounts fail to meet this requirement (they contain conspiracies themselves), and in section 3 I 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  $*\text{ʔ}/V\_V$  and  $*\text{ə}/VC\_C$  in (6) seems to be the avoidance of neutralization. Without the  $V\_$  condition, the *h*-aspire-initial form [kɛʔʔazaʃ] would end up as [kɛlazaʃ], neutralizing with a hypothetical vowel-initial form [kɛl#azaʃ]. Without the  $\_C$  condition, the *h*-aspire-initial form [ynəʔos] would end up as [ynos],

neutralizing with a hypothetical vowel-initial form [ynə#os].<sup>7</sup> 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):*

[ləazax̥]	[lɔm]	good (vowel)
?[kələzax̥]	[kəlɔm]	bad (none)
[kəlʔazax̥]	[kəlɔm]	okayish (creaky pause)
[ləazax̥]	[lezɔm]	good (consonant)
*[ynos]	[ynide]	bad (none)
?[ynʔos]	[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 |lə#azax̥|, for instance, can be recovered by the listener on the basis of the schwa in [ləazax̥]. The hiatus (sequence of two vowels) in [ləazax̥] is a clear sign of an underlying *h-aspiré*: had there been an underlying initial vowel (|lə#azax̥|), there would have been no schwa ([ləzax̥]), and had there been an underlying initial full consonant (e.g. |lə#gazax̥|), this consonant would have surfaced ([ləgazax̥]). 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 distinctivity of [ləazax̥] and [ləzax̥], which table (7) lists as *good* since the difference is a complete segment (a vowel). Similarly, the recoverability of *h-aspiré* in [ləazax̥] 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 |kəl#ʔazax̥| in (7). If the phonetic form were [kələzax̥], *h-aspiré* would be totally unrecoverable, since that phonetic form is also how a hypothetical |kəl#azax̥| would surface. Hence the listing of the [kələzax̥]-[kəlɔm] contrast as *bad*. The recoverability of *h-aspiré* from [kəlʔazax̥], as compared to [kəlɔm], is better than that from [kələzax̥], since some auditory cues for *h-aspiré* are available. Nevertheless, the recoverability is less good than in the [ləazax̥] and [ləzax̥] cases, since the auditory difference between *h-aspiré*-initial and vowel-initial words now has to be carried by the creaky pause [ʔ], 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 section 4.3); if a vowel is as audible as an average consonant, [ʔ] must be less audible than a vowel. Hence, the listing of the [kəlʔazax̥]-[kəlɔm] contrast as *okayish*.

Finally we arrive at the three candidates for |ynə#ʔos| in (7). Had the phonetic form been [ynos], *h-aspiré* would have been totally unrecoverable, since that phonetic form is also how a hypothetical |ynə#os| would surface. Hence, the listing of the [ynos]-[ynide] contrast as *bad*. The

<sup>7</sup> 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.

candidate [ynʔos] does as well as [kɛlʔazax], 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 [ləazax] and [leazax].

In the listener-oriented account, therefore, the surfacing of the underlying glottal stop in [kɛlʔazax] 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 often attempt to further improve [kɛlʔazax] by pronouncing it as [kɛləazax] must lie in the fact that schwa is not underlyingly present in [kɛlʔazax]. 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 phonetic 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 serial-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 is the subject of section 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). In section 3 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 can partly be solved by using *listener-oriented faithfulness* constraints, which do evaluate recoverability. A full solution involves the formalization of the recovery procedure in section 4 and the formalization of fitting the recovery procedure into production in section 5.

## 2. Formalizations in speaker-based 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 section 1.6 failed to achieve, and the possibility of comparing multiple candidates, which even the informal listener-oriented account of section 1.7 already had to do. This section shows that the universal defensibility does not go through in the usual speaker-based version of OT.

### 2.1. A failing speaker-based OT account

An OT account requires a number of universally defensible constraints. Since McCarthy and 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 tableaux. It is the constraint that has to rule out forms like \*[lgaʔs̃] and \*[lʔazax] 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 [lgaʔs̃]. However, Tranel (1996) does not consider the

(syllabified) form /l.ɣaʁ.sɔ̃./ worth including in his tableaux, perhaps for reasons of violation of sonority sequencing, and rules out the form /l.ɣaʁ.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 (section 1.5); the surfacing of schwa in [vvasiləzaʁ] ‘here is the coincidence’ (section 1.5) 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 tableaux handle the surfacing of schwa. An undominated faithfulness constraint has to rule out \*[kɛləzaʁ], at least in the non-variation account pursued here (cf. sections 1.3 and 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(V), which directly expresses the facts that [lom] is much better than \*[ləom] and [ynfam] is somewhat better than \*[ynəfam].<sup>8</sup>

(9) *Constraints handling the surfacing of schwa*

\*ə: “do not pronounce any schwas.”

MAX(V): “underlying expendable vowels (i.e. schwa and the vowel in [la]) must surface.”<sup>9</sup>

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 [z] in [lez]), 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(?): “underlying glottal stops must surface.”

MAX(C̄): “underlyingly expendable consonants must surface.”

To force the deletion of latent consonants before other consonants, i.e. to rule out \*[lezɣaʁsɔ̃] and \*[lezʔazaʁ], there must be a ranking like { MAX(C), MAX(?) } >> \*CC >> MAX(C̄), which

<sup>8</sup> The difference in grammaticality between \*[ynəide] and [ynəfam] is formalized in section 3.2.

<sup>9</sup> The only non-schwa expendable vowel is that in [la] (except for some restricted cases in [ty] ‘you’ and [si] ‘if’). Encrevé-Lambert (1971, as mentioned by Encrevé, 1988:133) and (apparently independently) Tranel (1996) propose that this vowel is not actually expendable, but that the whole allomorph [la] can be replaced by its masculine counterpart [lə]. An underlying form is then |[l̥]<sub>MASC</sub>, [la]<sub>FEM</sub>|#id̥<sub>FEM</sub>| ‘the idea’, and the /l/ in its surface form /li.de./ then represents an elided version of |[l̥]<sub>MASC</sub>|. This is plausible, since the same gender changing mechanism is required for handling allomorphs like |[m̥]<sub>MASC</sub>, [ma]<sub>FEM</sub>| ‘my’, |[b̥]<sub>MASC</sub>, [be]<sub>FEM</sub>| ‘beautiful’, and |[b̥]<sub>MASC</sub>, [bɔ̃]<sub>FEM</sub>| ‘good’ (Cornulier (1981:194) disagrees because of differences between masculine [bɔ̃] and feminine [bɔ̃]). Tranel (1996) proposes that the candidate generator is allowed to freely combine either of the allomorphs to the noun, so that phonological constraints are allowed to play a role in the decision. As a result, the allomorph that is selected before vowel-initial words is always the one that is capable of providing an onset (/m̥.n-/ with liaison, /be.l-/ with elision). In such a theory of allomorphy (of course not in Tranel’s version of it, which disallows underlying final schwa), the constraint MAX(V) could be replaced with MAX(ə).

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.<sup>10</sup>

- (11) *Constraint handling liaison*  
\*CC: “no consonant clusters.”<sup>11</sup>

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 \*ʔ, but since [kɛlʔazaʁ] is better than [kɛlazaʁ], this constraint must be outranked by MAX(ʔ).

- (12) *Constraint handling the surfacing of glottal stop*  
\*ʔ: “do not pronounce any creaky pauses.”

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

- (13) *A ranking for handling the French h-aspiré effects*  
{ MAX(C), DEP(ə) } >> MAX(ʔ) >> \*ʔ >> \*ə >> { MAX(V), \*CC } >> MAX(C)

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 I will refine the ranking 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 tableaux show that in the usual OT with two levels (underlying form and surface form) ranking (13) predicts only 9 of the 12 forms in (6) correctly. I thus (somewhat unfairly) claim 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 tableaux in (14), the pointing finger (“☞”) indicates the winner according to the grammar, and the check mark (“✓”) indicates the correct French

<sup>10</sup> Other formalizations of liaison are possible. In Tranel’s (1996) view, latent consonants lack an underlying prosodic node 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 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.

<sup>11</sup> Like the liaison rule “\*C/\_C” in (6), this constraint does not handle the phrase-final drop of latent consonants.

form. If the two point at different candidates, this indicates a failure of the speaker-based account.

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

ə#gɑrs̃	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ	*ə	MAX (V)	*CC	MAX (C)
√ <sub>ɛə</sub> ləgɑrs̃					*		*	
ləɑrs̃	*!				*		*	
lɑrs̃	*!					*	*	

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

lə#ʔazɑɾ	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ	*ə	MAX (V)	*CC	MAX (C)
ɛə ləʔazɑɾ				*	*			
√ ləazɑɾ			*!		*			
lazɑɾ			*!			*		

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

lə#om	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ	*ə	MAX (V)	*CC	MAX (C)
ləom					*!			
√ <sub>ɛə</sub> lom						*		
ləʔom				*!	*			

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

kəl#gɑrs̃	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ	*ə	MAX (V)	*CC	MAX (C)
√ <sub>ɛə</sub> kəlɡɑrs̃							**	
kəɡɑrs̃	*!						*	
kələrs̃	*!						*	
kələɡɑrs̃		*!			*		*	

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

kəl#ʔazɑɾ	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ	*ə	MAX (V)	*CC	MAX (C)
√ <sub>ɛə</sub> kəlʔazɑɾ				*			*	
kəʔazɑɾ	*!			*				
kələazɑɾ		*!	*		*			
kələzɑɾ			*!					

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

kɛl#ɔm	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ	*ə	MAX (V)	*CC	MAX (C)
√ <sub>ɛə</sub> kɛlɔm								
kɛɔm	*!							
kɛlʔɔm				*!			*	

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

lez#gɑɾsɔ̃	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ	*ə	MAX (V)	*CC	MAX (C)
lezgɑɾsɔ̃							**!	
√ <sub>ɛə</sub> legɑɾsɔ̃							*	*
lezɑɾsɔ̃	*!						*	
lezəgɑɾsɔ̃		*!			*		*	

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

lez#ʔazɑɾ	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ	*ə	MAX (V)	*CC	MAX (C)
lezʔazɑɾ				*			*!	
√ <sub>ɛə</sub> leʔazɑɾ				*				*
lezazɑɾ			*!					
√ leazɑɾ			*!					*

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

lez#ɔm	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ	*ə	MAX (V)	*CC	MAX (C)
√ <sub>ɛə</sub> lezɔm								
leɔm								*!
leʔɔm				*!				*

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

ynə#fam	MAX (C)	DEP (ə)	MAX (ʔ)	*ʔ	*ə	MAX (V)	*CC	MAX (C)
ynəfam					*!			
√ <sub>ɛə</sub> ynfam						*	*	

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

ynə#ŋos	MAX (C)	DEP (ə)	MAX (ʔ)	*ŋ	*ə	MAX (V)	*CC	MAX (C)
ynəŋos				*	*!			
<sup>h</sup> ynŋos				*		*	*	
√ ynəos			*!		*			
ynos			*!			*		

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

ynə#ide	MAX (C)	DEP (ə)	MAX (ʔ)	*ŋ	*ə	MAX (V)	*CC	MAX (C)
ynəide					*!			
√ <sup>h</sup> ynide						*		
ynŋide				*!		*	*	

From these tableaux 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.<sup>12</sup> But there remain three failures, enumerated in (15). There is no ranking of the constraints that can give the correct forms in all 12 cases.

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

- [ləŋʔazaʁ] instead of [ləazaʁ].
- [leŋʔazaʁ] instead of [leazaʁ].
- [ynŋos] 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 discuss them in sections 2.3 and 2.4, respectively. But first I have to summarize the first (non-OT) constraint-based account available in the literature, because the OT accounts are based on it.

2.2. *The constraint-based account by Cornulier ([1974] 1981)*

Several early accounts of *h-aspiré* phrased the relevant generalization in terms of constraints: the occurrence of hiatus was a *nécessité* for Martinon ([1905] 1962:16) and an *assurance* for

<sup>12</sup> An equally low ranked ONSET could perform the same function.



Damourette and Pichon (1911–1927:I:207). The first completely explicit constraint-based account, however, was that by Cornulier ([1974] 1981), who proposed a set of *postulats*, which were inviolable grammatical constraints, and *contraintes*, which were inviolable word-specific constraints in the lexicon.

Cornulier represents the facts of *h*-aspiré in terms of syllable boundaries, insisting that the phonological surface structures of our example forms look like /kɛ.ləm/ versus /kɛ.l.a.zaʁ/. We can start by observing that the forms that are handled correctly if we regard *h*-aspiré as a segment, as in (14), are also handled correctly if we regard *h*-aspiré as a syllable boundary. For *h*-aspiré words, Cornulier (1981:210) proposes a word-specific *syllable separation constraint*, i.e. *h*-aspiré words are lexically specified for satisfying an output-oriented constraint that says that the left edge of *h*-aspiré words has to be aligned to a syllable boundary.<sup>13</sup> This constraint is therefore identical to ALIGN (Word, Left; Syllable, Left) by McCarthy and Prince (1993). According to Cornulier, this constraint rules out forms like /la.zaʁ/ in (14b), /le.za.zaʁ/ in (14h), and /kɛ.la.zaʁ/ in (14e). According to Cornulier (p. 211), the candidate /kɛ.lə.a.zaʁ/ in (14e) is ruled out by the fact that the masculine |kɛl| does not have an underlying schwa and cannot therefore consummate the ‘right of schwa’ (*droit d’e*), except for those speakers who do not care about (or do not have) underlying final schwas and do say /kɛ.lə.a.zaʁ/. This constraint is therefore the same as DEP(ə).

If surface representations are ‘abstract’ syllable structures rather than the more ‘phonetic’ representations that appear in (14), then the candidate sets in the tableaux will be at the same time larger and smaller than those in (14), because a single phonetic form can sometimes stand for multiple syllable structures, and a single syllable structure can sometimes stand for multiple phonetic forms. An example of an ambiguous phonetic form is [lazaʁ] in (14b). Beside the obvious /la.zaʁ/, which we already discussed, it could be /l.a.zaʁ/, a form that does not violate Cornulier’s alignment constraint. Cornulier (p. 213) rules out this candidate with his ‘Postulate IV’ (p. 201), which says that every syllable has to contain exactly one vowel. Conversely, an example of an auditorily ambiguous syllable structure is /lə.a.zaʁ/. If the question whether the hiatus between /ə/ and /a/ will be pronounced with a glottal stop or not is considered merely a case of phonetic implementation, then this structure can reflect both [ləʔazaʁ] and [ləazaʁ]. As a result, any syllable-based account that ignores phonetic detail automatically takes care of the failures in (15a) and (15b) by considering them irrelevant, leaving only the remaining case of [ynəos] in (15c) unaccounted for.

For the underlying form |ynə#ʔos|, two candidate syllable structures satisfy all of Cornulier’s constraints, namely /yn.os/ and /y.nə.os/. The big question now is: to which of the four ‘phonetic’ forms in (14k) do these two structures correspond? First, presumably, /y.nə.os/ corresponds to both [ynəos] and [ynəʔos]. But what does /yn.os/ correspond to? If it corresponds to both [ynʔos] and [ynos], then Cornulier’s constraints allow all four forms in (14k), which certainly does not improve on the problem in (15c). Cornulier is aware of this problem, and invokes (on p. 211) his independently needed ‘Postulate V.1’ (p. 201), which says that if a CV sequence is pronounced in one stroke (without pause), there cannot be a syllable boundary between the consonant and the vowel. This entails that the pronunciation [ynos] can never reflect the structure /yn.os/. This rules out the arguably worst candidate in (14k), namely [ynos]. Note, however, that Cornulier’s Postulate V.1 is a *listener-oriented* constraint under light disguise, because it can be formulated as stating that syllable boundaries should be *audible* or *recoverable*,

<sup>13</sup> Except for glide-initial *h*-aspiré words, which need not satisfy this constraint (Cornulier, 1981:214), as in [swiski] for [sɛ+ʔwiski] ‘that whisky’. See section 6.4.

i.e. before a vowel they must either be a hiatus (after a vowel) or a pause (after a consonant). The present paper therefore identifies this constraint as a *cue constraint* in sections 4.4, 6.1 and 7.1. A fully formalized listener-oriented account using Cornulier’s constraints and syllable structures with minimal modifications can be done, but only after a detailed presentation of a framework in which it is possible to represent these constraints and structures. The result is in section 7.1.

Finally, we must note that any preference of the ‘official’ form /y.nə.os/ over the attested but less ‘official’ form /yn.os/ cannot be handled by Cornulier’s complete constraint set, so that even with Cornulier’s cue constraint the problem of (15c) is not solved. Any solution requires violable constraints (as is shown in e.g. sections 2.3 and 7.1), and therefore the framework of Optimality Theory.

### 2.3. The patch by Tranel and Del Gobbo (2002)

The first OT account of *une hausse* was provided by Tranel and Del Gobbo (2002), but it was based on earlier adaptations of Cornulier’s proposals by Tranel.

In a less explicit non-OT account than Cornulier’s, Tranel (1995) renames Cornulier’s word-specific syllable separation constraint to a *syllable island constraint*, and arrives at identical syllable structures. There are two differences with Cornulier’s approach, though.

The first difference is that, in contrast with Cornulier, Tranel (1995) does not consider at all the phonetic realization of the abstract syllable structures; still, presumably, a separate process of phonetic implementation will have to turn the prevocalic syllable boundary in /kɛl.a.zaʁ/ into a pause, a glottal stop, and/or creak: [kɛl\_azaʁ], [kɛl\_azaʁ], [kɛl\_ʔazaʁ], [kɛl\_ʔazaʁ], or [kɛlazaʁ] (section 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 contrastive feature of vowels, the realization of a syllable boundary as creak in the following vowel (as in [kɛlazaʁ]) cannot be universal and must be specific to French. In a full account of *h*-aspire, therefore, something similar to Cornulier’s cue constraint, and probably something much more detailed than that, will have to be included.

The second difference is that Tranel (1995) does attempt to get rid of the candidate /yn.os/, in contrast with Cornulier (1981). He proposes (p. 812) that this form is ruled out by “the phonological pressure exerted by forward syllabification in VCV sequences”: /yn.os/ violates this pressure because of the backward syllabification of /n/.<sup>14</sup> The OT account by Tranel (1996), who was apparently unaware of the connection, closely follows that by Cornulier, including the word-specific constraint ALIGN (Word, Left; Syllable, Left) and an every-syllable-one-vowel constraint (‘NUCLEUS/V’) (Tranel does leave out Cornulier’s cue constraint, thus ignoring the phonetics again). In his OT implementation Tranel expressed the ‘phonological pressure of forward syllabification’ with a combination of OT constraints: 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. Despite Tranel’s reluctance, Tranel and Del Gobbo (2002), having no choice, ultimately did formalize the phenomenon with the conjoined constraint ONSET&NoCODA, as shown in (16).

<sup>14</sup> An earlier proposal (Tranel, 1987:95) that a form like /vwa.sil.a.zaʁ/ is ruled out because the // ‘belongs’ to the morpheme [azaʁ] rather than to [vwasil] cannot be correct, given that /vwa.sil.gaʁ.sɔ̃/ is grammatical.

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

ynə#os	ALIGN-L(/os/, σ)	ONSET&NoCODA	*ə	MAX (V)
√ <sub>ε</sub> .y.nə.os.			*	
.yn.os.		*!		*
.y.nos.	*!			*

The tableau shows that what remains is /y.nə.os/. We can also understand why Cornulier, with his inviolable constraints, did not manage to rule out /yn.os./: the relevant constraint ONSET&NoCODA has to be a *violable* constraint, because the optimal form /kɛ.l.a.zaʁ/ does not satisfy it; therefore, the word-specific alignment constraint ALIGN-L has to dominate ONSET&NoCODA, an option not available within Cornulier's theory of inviolable constraints.

Meisenburg and Gabriel (2004) criticize Tranel and 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, because (as Schane (1978) indicates) the underlying form can be specified with an underlying syllable boundary, i.e. as |.os|, so that Cornulier's alignment constraint can be replaced with such faithfulness constraints as McCarthy and Prince's (1995) CONTIGUITY and LINEARITY, i.e. with alignment-as-faithfulness (Boersma, 1998:196–199; Horwood, 2002), and this is pursued in the present paper in section 7. Instead, the real problem in (16) is the constraint ONSET&NoCODA. First, proposing constraints that are conjoined from two heterogeneous 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 [ynʔos] 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 and 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 and Del Gobbo would have to extend their model with a serious formalization of the phonetics. The present paper implements precisely such a formalization (sections 3–8), and achieves the desired coverage with less constraint conjunction and more explanatory force.

2.4. *The patch by Meisenburg and Gabriel (2004)*

The account by Meisenburg and 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 and Del Gobbo did. H-aspiré is represented underlyingly as a (creaky) glottal stop, and the constraint set is similar to the one used here in section 2.1, but with some additional constraints to get rid of the failures in (15).

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

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

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

While this constraint rules out [ləʔazaʁ] and [leʔazaʁ], 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 has to become [lazaʁ], because of the crucial constraint ranking \*ə >> MAX(C). In Meisenburg and Gabriel's view, the form \*[lazaʁ] violates an alignment constraint very similar to Cornulier's (section 2.2), although it refers to a segmental glottal stop rather than to a syllable boundary.

(18) *Meisenburg and 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 [ləʔazaʁ] becomes the winner.<sup>15</sup>

(19) *Meisenburg and Gabriel's speaker-based elision, h-aspiré case*

ə#ʔazaʁ	MAX (C)	ALIGN (ʔ)	*VʔV	DEP (ə)	MAX (ʔ)	*ʔ	*ə	MAX (V)	*CC	MAX (C)
ləʔazaʁ			*!			*	*			
√ <sub>ɛ</sub> ləʔazaʁ					*		*			
lazaʁ		*!			*			*		

The addition of two constraints, then, handles problems (15a) and (15b) (almost<sup>16</sup>). To handle (15c), additional machinery is called for, since the ranking of (19) would still make [ynʔos] the winner in (14k). Meisenburg and Gabriel propose that this candidate is ruled out by the conditional faithfulness constraint in (20).

(20) *Meisenburg and 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.

<sup>15</sup> Meisenburg and Gabriel's account differs from (19) on two minor points. They handle liaison with Tranel's (1996) constraint "avoid integrating floaters" (see footnote 10), and they do not discuss anything like \*ə >> MAX(V) because they do not consider candidates like \*[ləom] and ?[ynəfam].

<sup>16</sup> *Almost*, because although in (14h) [ləʔazaʁ] has now become better than [lezazaʁ], the winning candidate is [lezʔazaʁ], a form that M&G did not take into account. This could be corrected by adding another constraint or by ranking MAX(ʔ) lower (this has become possible because the alignment constraint can now take over some of the effects of MAX(ʔ), and it would also make \*VʔV superfluous thanks to the presence of our \*ʔ). A possible speaker-based ranking that handles all of (14) is { MAX(C), ALIGN(ʔ), MAX(ə/\_ʔ), DEP(ə) } >> { \*ʔ, \*ə } >> \*CC >> { MAX(ʔ), MAX(V), MAX(C) }.

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

[ynə#ʔos]	MAX (C)	ALIGN (ʔ)	MAX (ə/_ʔ)	*VʔV	DEP (ə)	MAX (ʔ)	*ʔ	*ə	MAX (V)	*CC	MAX (C)
ynəʔos				*!			*	*			
ynʔos			*!				*		*	*	
√ <sub>εσ</sub> 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 words: schwa must be saved in [ynəos] because that improves the recoverability of *h-aspiré* (section 1.7). In fact, all three constraints proposed by Meisenburg and Gabriel can be understood in the light of this recoverability, as summarized in (22).

(22) *Informal listener-oriented reformulation of Meisenburg and 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 explicitly. The next section shows that this can be done and that none of the three concocted constraints in (22) is necessary.

**3. Formalization with listener-oriented faithfulness constraints**

The deeper cause of the problems identified in sections 1.6 through 2.4 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 explicitly, and that is the procedure followed in this section. It will turn out not to have the problems of the speaker-based approach.

### 3.1. The simplest listener-oriented OT account

In a grammar consisting only of faithfulness constraints and constraints against certain structures, the locus of recoverability can only be the faithfulness constraints, i.e. it must be up to the faithfulness constraints to evaluate the degree to which the listener will be able to recover the phonological structure. To arrive at such *listener-oriented faithfulness constraints*, very little has to be changed in comparison to McCarthy and Prince's (1995) two-level formulation of faithfulness. In (23) I compare McCarthy and Prince's version of MAX to Boersma's (1998) three-level listener-oriented version.

(23) *Listener-oriented faithfulness as recoverability*

MAX(?) (speaker-based): “pronounce an underlying  $[ʔ]$  as  $/ʔ/$  (or  $[ʔ̥]$ ).”

MAX(?) (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 those psycholinguists who agree that lexical access is not involved call it *prelexical processing* (see section 4.2 for a more detailed discussion of the assumption and references to the literature). If the lexicon is indeed 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 section 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/.<sup>17</sup> This yields the last column in table (6).

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 auditory forms like [teatʁ] and [agʁeabl] must be perceived as /teʔatʁ/ and /agʁeʔabl/ despite the fact that these forms reflect single words ('theatre', 'agreeable') whose word-internal hiatus will never be in a

<sup>17</sup> Or they are perceived with syllable structure as /lɛ.ʔa.zaʁ/, /lɛ.ʔa.zaʁ/ and /y.nə.ʔos/, or as the Cornulier-type surface forms /lɛ.a.zaʁ/, /lɛ.a.zaʁ/ and /y.nə.os/. Which type of surface structure is perceived depends on one's theory of French phonological structures (see sections 2.2, 7 and 8.2).

position that could show any consonant-like *h*-aspiré effects.<sup>18,19</sup> The same automaticity of perception predicts that the candidate [ləʊm] in (14c) is perceived as /ləʊʔm/, with a glottal stop that is not present in the underlying form. Likewise, [ləɔm] and [ynəɔide] would be perceived as /ləʊʔm/ 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(?) (speaker-based): “pronounce an underlying |ʔ| as /ʔ/ (or [ʔ]).”  
 MAX(?) (listener-oriented): “pronounce an underlying |ʔ| as a phonetic form that the listener will perceive as /ʔ/.”  
 DEP(?) (speaker-based): “do not pronounce a /ʔ/ (or [ʔ]) if there is no corresponding underlying |ʔ|.”  
 DEP(?) (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 [ləʌzaʒ], [ləʌzaʒ] and [ynəɔs] satisfy the listener-oriented version of MAX(?), and the forms [ləʊm], [ləɔm] and [ynəɔide] violate the listener-oriented version of DEP(?). This will have repercussions for the tableaux in (14). If we ignore DEP(?) for a while (until section 3.2), we see that with a listener-oriented MAX(?) eight tableaux stay the same, but the remaining four tableaux, namely those with an underlying *h*-aspiré, change their violation patterns. For (14e) the winner does not change, as shown in (25).

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

	kɛl#ʔazaz	MAX (C)	DEP (ə)	MAX (ʔ)	*[ʔ]	*[ə]	MAX (V)	*CC	MAX (C)
√ <sub>ES</sub>	[kɛlʔazaz] → /kɛlʔazaz/				*			*	
	[kɛʔazaz] → /kɛʔazaz/	*!			*				
	[kɛləʌzaz] → /kɛləʌzaz/		*!			*			
	[kɛlazaz] → /kɛlazaz/			*!					

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

<sup>18</sup> 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 [ʒɛʊvɛʒ] from [ʒɛ#ʊvɛʒ] ‘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.” In other words, Bally interprets [ʒɛʊvɛʒ] as /ʒɛʔʊvɛʒ/, where the glottal stop symbol stands for Bally’s zero consonant. Nevertheless, section 8 shows that native speakers are probably correct in rejecting the glottal stop from non-*h*-aspiré hiatus.

<sup>19</sup> About the fact that the /ʔ/ in perceived forms like /aɡʒɛʔabl/ and /ʒɛʔʊvɛʒ/ are not present in the underlying lexical representation, see section 3.2.





In (27) we see that it is crucial that \*ʔ 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 \*ʔ as an articulatory constraint ensures that of the pronunciations [leʔazaʔ] and [leazaʔ], 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 [leʔazaʔ] and [leazaʔ] 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 section 4.4). 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 a 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 |ynə#ʔos|. Tableau (28) shows that even this form is handled correctly by a listener-oriented MAX(?).

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

ynə#ʔos	MAX (C)	DEP (ə)	MAX (ʔ)	*[ʔ]	*[ə]	MAX (V)	*/CC/	MAX (C)
[ynəʔos] → /ynəʔos/				*!	*			
[ynʔos] → /ynʔos/				*!		*	*	
<sup>√</sup> <sub>ε</sub> [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 section 2.1, can now also be addressed, as is done in the next section.

### 3.2. Improving the constraint ranking: the need for DEP(?)

While tableaux (25) through (28) have illustrated a listener-oriented interpretation of MAX(?), a listener-oriented interpretation of its counterpart DEP(?), 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 tableaux (14c), (14j) and (14l), I have attributed the ungrammaticality of [ləm], [ynəfam] and [ynəide] to the ranking of \*[ə] over MAX(V). But as noted in section 1.5, these three forms are not equally ungrammatical. The forms \*[ləm] and \*[ynəide] are much worse than ?[ynəfam], which can occur in practice in a non-negligible minority of cases. The variation between [ynəfam] and ?[ynəfam] can be explained by ranking \*[ə] just a bit above

MAX(V). The stochastic form of evaluation in Optimality Theory (Boersma, 1997; Boersma and Hayes, 2001) then predicts that at evaluation time MAX(V) will outrank \*[ə] in a large minority of cases, so that forms like ?[ynəfam] will occur, although less often than [ynfam].<sup>20</sup> But a close ranking of \*[ə] and MAX(V) 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(?). This is made explicit in tableaux (29) through (31).

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

lə#ɔm	MAX (C)	DEP (ə)	DEP (?)	MAX (?)	*[ʔ]	*[ə]	MAX (V)	*/CC/	MAX (C)
[ləɔm] → /ləʔɔm/			*!			*			
<sup>√</sup> [ləɔm] → /ləɔm/							*		
[ləʔɔm] → /ləʔɔm/			*!		*	*			

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

ynə#ide	MAX (C)	DEP (ə)	DEP (?)	MAX (?)	*[ʔ]	*[ə]	MAX (V)	*/CC/	MAX (C)
[ynəide] → /ynəʔide/			*!			*			
<sup>√</sup> [ynide] → /ynide/							*		
[ynʔide] → /ynʔide/			*!		*		*	*	

(31) *Listener-oriented ‘schwa drop’, consonant case*

ynə#fam	MAX (C)	DEP (ə)	DEP (?)	MAX (?)	*[ʔ]	*[ə]	MAX (V)	*/CC/	MAX (C)
[ynəfam] → /ynəfam/						*!			
<sup>√</sup> [ynfam] → /ynfam/							*	*	

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

The height of DEP(?) can be determined more precisely than just as DEP(?) >> \*[ə]. This constraint cannot be undominated, as can be seen by considering underlying forms like |ʃɔ̃bɛ#ə#æɛ| ‘inner bicycle tube’ or |ʒə#uvɛɛ| ‘I opened’. We know that the morpheme |ɛɛ| ‘air’ in the first form does not have an underlying glottal stop, because |lə#ɛɛ| ‘the air’ is pronounced [lɛ:ɛ], not \*[lə:ɛ]. Nevertheless, the form is pronounced as [ʃɔ̃bɛæ:ɛ], which as a result of its hiatus has to be perceived as [ʃɔ̃bɛəʔɛɛ]. This attested correct French form thus violates DEP(?). The cause must

<sup>20</sup> I ignore here the possibility that it is \*/CC/ rather than MAX(V) that is variably ranked with respect to \*[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).

be that all its competitors violate a constraint that is ranked higher than DEP(?). Most notably, the competitor [ʃǎbʁɛɛʁ], which is perceived as /ʃǎbʁɛʁ/ and therefore satisfies DEP(?) 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 obtained.

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

ʃǎbʁɛ#a#ɛʁ	MAX (C)	MAX (V)	DEP (ə)	DEP (?)	MAX (?)	*[ʔ]	*[ə]	MAX (Y)	*/CC/	MAX (C)
[ʃǎbʁɛaɛʁ] → /ʃǎbʁɛʔaʔɛʁ/				**!			*		*	
√ <sub>ε</sub> [ʃǎbʁaɛʁ] → /ʃǎbʁaʔɛʁ/				*				*	*	
[ʃǎbʁɛʁ] → /ʃǎbʁɛʁ/		*!						*	*	

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

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

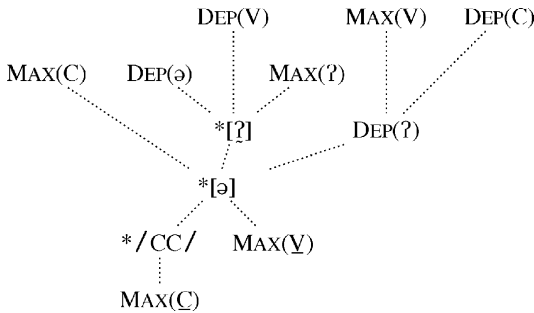
(33) *Listener-oriented liaison*

lez#ɔm	MAX (C)	MAX (V)	DEP (ə)	DEP (?)	MAX (?)	*[ʔ]	*[ə]	MAX (Y)	*/CC/	MAX (C)
√ <sub>ε</sub> [lezɔm] → /lezɔm/										
[leɔm] → /leʔɔm/				*!						*
[leʔɔm] → /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 tableaux, are the cases in which it had to rule out \*[lezgəʁsɔ̃] and \*[lʔazaʁ]. In these tableaux \*/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(?) high-ranked, this problem is out of the way, and liaison could be handled by MAX(C) >> \*/C/ >> MAX(C). In (33), the constraint \*/C/ would be violated three times in each candidate. As we can induce from tableau (25), \*/C/ would have to be ranked below MAX(?).<sup>21</sup>

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 an absurd form like [ʃǎbʁaɛɛʁ] in tableau (32).

<sup>21</sup> Yet another analysis of liaison that has become possible is Tranel’s DEP(X), see footnote 10. In section 2.3, Tranel’s analysis required a ranking like ONSET >> DEP(X), which would be problematic in the present syllable-free framework. Now that DEP(?) is available, however, the same effect can be achieved with the ranking DEP(?) >> DEP(X), i.e. [lezɔm]/leʔɔm/ is better than [leɔm]/leʔɔm/ because the former violates DEP(X), the latter DEP(?).

(34) *A ranking for h-aspiré without optionality*3.3. *Evaluative comparison of the speaker-based and listener-oriented accounts*

When compared to the two-level speaker-based OT accounts of [Tranel and Del Gobbo \(2002\)](#) and [Meisenburg and Gabriel \(2004\)](#), which vary in their degrees of abstractness of surface forms, the three-level listener-oriented account proposed in sections 3.1 and 3.2 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 and Gabriel's proposal, the listener-oriented proposal economizes on three constraints:  $*VʔV$ ,  $ALIGN(ʔ, \sigma)$ , and  $MAX(\partial/_?)$ . From an emergentist viewpoint of OT constraints ([Boersma, 1998](#)), such economy makes for better grammars, because the child has to create fewer constraints that express parochial and coincidental generalizations over the language data; from an innatist viewpoint of OT constraints ([Prince and Smolensky, 1993](#)), this specific economy also makes for better grammars, because innate constraints are assumed to be cross-linguistically useful rather than specific to the language at hand, as especially  $MAX(\partial/_?)$  clearly is. The cause of this economy lies in the fact that these three 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  $[VV] \rightarrow /VʔV/$  seems to be comparable to the cost of the speaker's rule in the fourth column of (6), which can be written  $/VʔV/ \rightarrow [VV]$ : the two are perfect mirror images, as we can expect from their opposite viewpoints (production versus comprehension). When compared to Meisenburg and 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 two grammar elements, namely the constraints  $ALIGN(ʔ, \sigma)$  and  $MAX(\partial/_?)$ , of which especially the latter is problematic as far as universal defensibility is concerned. Three large points of concern remain, however. The first is how to assess the rule  $[VV] \rightarrow /VʔV/$ . Is it extragrammatical, or can it be incorporated into the OT grammar? Is it universally defensible, or is it a quirk specific to French? And is it really needed in more full-fledged formalizations? These questions are answered in sections 4 and 8. The second point of concern is the fact that although section 3.1 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 3.1 has attributed the preference of [ynəos] over [ynʔos] to the speaker-based ranking \*[ʔ] >> \*[ə], contrary to the listener-oriented explanation of section 1.7 in terms of the quality of auditory cues. Can the idea that [ynəos] has better auditory cues for recovering *h*-aspiré than [ynʔos] be formalized? This question is answered affirmatively in section 5, where some explicit listener-oriented models of the interplay between phonology and phonetics are proposed. The third point of concern regards the form [ynəʔos], 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 section 6.

#### 4. Modelling the recovery process

Although the account presented in section 3 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.

##### 4.1. Four representations

As we can see from section 3, 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 and 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 psycholinguists for comprehension (e.g. McQueen and Cutler, 1997: AudF, SF, UF) and for production (e.g. Levelt, 1989: UF, SF, ArtF).

In the tableaux of section 3, 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 \*[ʔ] and \*[ə] evaluate ArtF;

the constraint \*/CC/ evaluates SF; and the faithfulness constraints evaluate the similarity between SF and UF.

#### 4.2. 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  $\text{AudF} \rightarrow \{\text{SF}, \text{UF}\}$ , in which case the listener's perception would be influenced by lexical access. In section 3.1, 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 and 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) and followed here in section 3.1. The comprehension model in that view is therefore  $\text{AudF} \rightarrow \text{SF} \rightarrow \text{UF}$ : it consists of two serial 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:chapter 18, 2000), or the integration of multiple auditory cues into a bivalued or multivalued phonological contrast (Escudero and Boersma, 2003, 2004; Boersma and Escudero, 2004). A very similar OT mapping was proposed by Tesar (1997, 1998, 1999) and Tesar and 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 σ σ], to an abstract hidden prosodic structure with feet and head syllables, e.g. /σ(σσ)/. 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 and Boersma, 2004), although Smolensky (personal communication) maintains that the two work at different levels.

#### 4.3. A grammar for perceiving h-aspiré: one example

In the tableaux of section 3, 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 arrow in the third candidate in tableau (28), namely the French listener's mapping from an overt [VV] to a hidden /V?V/ for the auditory form [ynəos]. The Optimality-Theoretic perception tableau is shown in (35).

(35) *The perception of hiatus*

[ynəos] <sub>Aud</sub>	*/VV/	*/CV/ X  VC	*/VCV/ \\ /  VV	*/V/ \\ /  VV	*/V?V/ \\ /  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 surface 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 phonological surface 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 in production, so it could be capable of influencing perception as well as production. This bidirectional use of constraints at SF was stressed by Tesar and 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 section 3, for instance, \*/VV/ does not make its appearance in the tableaux, 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 \*/V\_V in (6) or Meisenburg and Gabriel’s constraint \*[V?V] in section 2.4. The constraint is top-ranked in (35); I defer a discussion of how (and whether) this high ranking might come about until sections 6.5 and 8.1.

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 and Boersma, 2003, 2004), called from now on *cue constraints*, that militate against mapping any auditory cue to any phonological element.<sup>22</sup> 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).

The second constraint in (35) is about order. The auditory form [ynəos] contains information on the relative timing of sibilance and back vocalicity. Perceiving it as /ynəso/ would discard this information, violating a constraint loosely statable as \*[V<sub>1</sub>C<sub>2</sub>]/C<sub>2</sub>V<sub>1</sub>/, which is short for “do not

<sup>22</sup> One of the reasons for the negative formulation is explained in Boersma and Escudero, 2004, the other here in section 6.1.1.

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 of this constraint (“\*[V<sub>1</sub>C<sub>2</sub>]/C<sub>2</sub>V<sub>1</sub>/”), 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<sub>1</sub>C<sub>2</sub>]/C<sub>2</sub>V<sub>1</sub>/ ranked so high? And the question concerning the explanatory adequacy is: how has \*[V<sub>1</sub>C<sub>2</sub>]/C<sub>2</sub>V<sub>1</sub>/ 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 and Marslen-Wilson, 1991; McQueen and 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. In Optimality-Theoretic terms, the number of violations of SF-UF faithfulness constraints in the recognition phase should be minimized. In the present case of time ordering, the faithfulness constraint to minimize violations of is McCarthy and Prince’s LINEARITY, whose specific instantiation for the present case we could write as \*/C<sub>1</sub>V<sub>2</sub>/|V<sub>2</sub>C<sub>1</sub>|. For French, we observe that this language never uses overt metathesis in alternations, i.e. the order of the elements in an underlying form such as |gɑʁsɔ̃| is always identical to the order of auditory elements in a phonetic form such as [gɑʁsɔ̃]. The simplest way to handle this underlying-to-auditory ordering identity is to have both a high-ranked LINEARITY in recognition (so that lexical access is optimized) and a high-ranked \*[V<sub>1</sub>C<sub>2</sub>]/C<sub>2</sub>V<sub>1</sub>/ in perception, resulting in an SF (such as /gɑʁ.sɔ̃/) that always has its elements in the same order both as the recognized UF and as the cues in AudF. The answer to the second question is that this high ranking is an automatic result of lexicon-driven learning of perception (Boersma, 1997; Escudero and Boersma, 2004). Tableau (36) illustrates what happens if at a certain point during acquisition the learner perceives [gɑʁsɔ̃] incorrectly as /gʁɑ.sɔ̃/, because of a ranking of \*[V<sub>1</sub>C<sub>2</sub>]/C<sub>2</sub>V<sub>1</sub>/ below the structural constraint \*CODA (which evaluates SF).

(36) *Learning the high ranking of auditory-to-phonological ordering identity*

[gɑʁsɔ̃] <sub>Aud</sub> (UF =  gɑʁsɔ̃ )	*CODA	*/CV/ X [VC]
✓ /gɑʁ.sɔ̃/	*!→	
✗ /gʁɑ.sɔ̃/		←*

The auditory form [gɑʁsɔ̃] is perceived as the surface form /gʁɑ.sɔ̃/, but the subsequent process of word recognition has to map this SF to the lexical item |gɑʁsɔ̃| ‘boy’ at UF (thus violating LINEARITY). The learner uses this lexical information to mark the UF-identical form in the tableau with a check mark (✓). Now that the winning candidate /gʁɑ.sɔ̃/ is different from the one that the learner considers correct, namely /gɑʁ.sɔ̃/, the learner will take action by slightly raising the ranking of the constraints that prefer the correct form over the winning form, and slightly lowering the ranking of the constraints that have the opposite preference. The rankings move along a continuous scale, assuming Stochastic OT (Boersma, 1998), so that if they move by small steps, it will take multiple learning data before the order of any constraints is changed; the symmetric ranking update strategy just described is the Gradual Learning Algorithm (Boersma, 1997; Boersma and Hayes, 2001). In tableau (36), \*CODA will fall a bit and \*[V<sub>1</sub>C<sub>2</sub>]/C<sub>2</sub>V<sub>1</sub>/ will rise a



bit, thus making it slightly more likely (given the evaluation noise of Stochastic OT) that the learner will perceive /gɑ̃.s̃/ at future occurrences of [gɑ̃s̃]. Since no cases of metathesis occur in French, \*[V<sub>1</sub>C<sub>2</sub>]/C<sub>2</sub>V<sub>1</sub>/ will ultimately end up ranked far above \*CODA.<sup>23</sup>

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 [ynətos]<sub>Art</sub> and [ynəʔos]<sub>Art</sub>, respectively, the first one has a much smaller chance of being heard as [ynəos]<sub>Aud</sub> in a noisy environment than the second, simply because [t]<sub>Art</sub> produces a superset of the auditory cues that [ʔ]<sub>Art</sub> produces (the shared cue is the silence; the cues that [t]<sub>Art</sub> does and [ʔ]<sub>Art</sub> does not produce are formant contours and release burst). If [t]<sub>Art</sub> is not much more common than [ʔ]<sub>Art</sub> in this over-faithful French, [ynəos]<sub>Aud</sub> 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/.<sup>24</sup> In a bidirectional situation, such a ranking is self-reinforcing: the speaker, knowing that the listener is likely to perceive [ynəos]<sub>Aud</sub> as /ynəʔos/, may simply stop pronouncing the glottal stop and get away with it (Boersma, 1998:182),<sup>25</sup> 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 [ynəos]<sub>Aud</sub> as /ynos/ would discard the auditory information in a pronounced [ə]<sub>Art</sub>, violating \*[V<sub>1</sub>V<sub>2</sub>]/V<sub>2</sub>/. Background noise is much more likely to conceal 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.<sup>26</sup> 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/.<sup>27</sup> If schwa

<sup>23</sup> Comparable results (high ranking of both ordering cue constraints and LINEARITY) will emerge in the parallel comprehension learning algorithm by Boersma (2006), which uses a parallel evaluation of the AudF → SF and SF → UF mappings rather than the serial evaluation discussed here (in that acquisition model, the learner receives pairs of UF and AudF and has to construct SF herself). If the auditory cues for ordering are very poor or ambiguous, the learner will end up with low-ranked ordering cue constraints and LINEARITY, which can lead to metathesis in sound change (this confirms Blevins and Garrett’s (1998) perceptual account of diachronic metathesis) as well as to metathesis in the synchronic phonology and/or phonetic implementation.

<sup>24</sup> The relative commonness of [t] and [ʔ] also plays a role. The more common an underlying element is, the lower the constraints against perceiving it will be ranked.

<sup>25</sup> 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 formal explanations in terms of blind mechanisms, see section 5.3 and especially the computer simulations in Boersma (2006).

<sup>26</sup> 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.

<sup>27</sup> 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.

cues have a comparable quality as glottal stop cues,  $*[V_1V_2]/V_2/$  will therefore be ranked above  $*[VV]/V\text{?}V/$ , as it is in (35). If schwa has better cues than a glottal stop (as proposed in section 1.7), the ranking difference will be even larger. In (35), I have ranked  $*[V_1V_2]/V_2/$  below  $*[VV]/V\text{?}V/$ , because I think  $[\text{yn}\text{?}\text{os}]_{\text{Aud}}$  is more likely to be perceived as  $/\text{ynos}/$  than as  $/\text{yn}\text{?}\text{os}/$ . A perception experiment may shed light on issues like these but is outside the scope of the present paper (see also footnote 28).

#### 4.4. Ranking by cue

This section gives all the constraints responsible for the order of auditory cue quality proposed in section 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 (or syllable boundary). Beside the structural constraint  $*[VV/$  we need eight cue constraints rather than just the two remaining in tableau (35).

The first four cue constraints to consider are those against perceiving a glottal stop. In section 1.7, the cues for the glottal stop were proposed to be better in  $[\text{yn}\text{?}\text{os}]$  than in  $[\text{yn}\text{?}\text{os}]$ , since this difference was able to account for the speaker's preference of  $[\text{yn}\text{?}\text{os}]$  over  $[\text{yn}\text{?}\text{os}]$ . 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.<sup>28</sup> The (perhaps language-specific) quality of the auditory cues in  $[\text{yn}\text{?}\text{os}]$  and  $[\text{yn}\text{?}\text{os}]$  will lead to the ranking  $*[C\text{?}V]/C\text{?}V/ \gg *[VV]/V\text{?}V/$ . 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.  $[\text{yn}\text{?}\text{os}]$ ; the constraint  $*[V\text{?}V]/V\text{?}V/$  must therefore be very low ranked. Finally, the form with neither schwa nor the creaky pause, i.e.  $[\text{ynos}]$ , must have the poorest cues for glottal stop; the constraint  $*[CV]/C\text{?}V/$  must therefore be very high ranked; it corresponds exactly to Cornulier's (1981:201) cue constraint ('Postulate V.1') for syllable boundaries ( $*[CV]/C.V/$ ), which was discussed in section 2.2 and will be discussed again in section 7.1. The ranking of the four constraints, then, is that in (37).

$$(37) \quad \text{Ranking of cue constraints against perceiving a glottal stop in French}$$

$$\begin{array}{ccccccc}
 */[C\text{?}V]/ & & */[C\text{?}V]/ & & */[V\text{?}V]/ & & */[V\text{?}V]/ \\
 \begin{array}{c} \backslash / \\ [CV] \end{array} & \gg & \begin{array}{c} | | | \\ [C\text{?}V] \end{array} & \gg & \begin{array}{c} \backslash / \\ [VV] \end{array} & \gg & \begin{array}{c} | | | \\ [V\text{?}V] \end{array}
 \end{array}$$

This ranking must already be a simplification. For instance, the formulation of the second constraint, the one against perceiving the underlying *h*-spiré in  $[\text{k}\text{e}\text{l}\text{?}\text{azax}]$ , is a summary for five more precise formulations, as shown in (38).

<sup>28</sup> 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. For comparing the *h*-spiré cues of hiatus with those of a creaky pause, a perception experiment would involve having French listeners classify auditory forms like  $[\text{yn}\text{?}\text{otm}\text{?}\text{ta}\text{ŋ}]$  and  $[\text{yn}\text{?}\text{otm}\text{?}\text{ta}\text{ŋ}]$  as either "une haute montagne" (i.e.  $[\text{yn}\text{?}\text{#}\text{ot}\text{?}\text{#m}\text{?}\text{ta}\text{ŋ}]$  'a high mountain') or "une autre montagne" (i.e.  $[\text{yn}\text{?}\text{#ot}\text{?}\text{#m}\text{?}\text{ta}\text{ŋ}]$  'another mountain'). Large methodological challenges include (but seem not to be limited to): varying realistically the degree of reduction of schwa in  $[\text{yn}\text{?}\text{ot}]$ , varying realistically the multitude of realizations of  $[\text{yn}\text{?}\text{ot}]$ , 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.

(38) *Simplifying a cue constraint*

$$\begin{array}{cccccc}
 */C?V/ & = & */C?V/ & \text{or} & */C?V/ & \text{or} & */C?V/ & \text{or} & */C?V/ \\
 \begin{array}{|c|} \hline \text{||} \\ \hline \text{[C?V]} \\ \hline \end{array} & & \begin{array}{|c|} \hline \text{||} \\ \hline \text{[C?V]} \\ \hline \end{array} & & \begin{array}{|c|} \hline \text{||} \\ \hline \text{[C?V]} \\ \hline \end{array} & & \begin{array}{|c|} \hline \text{||} \\ \hline \text{[C?V]} \\ \hline \end{array} & & \begin{array}{|c|} \hline \text{||} \\ \hline \text{[C?V]} \\ \hline \end{array} \\
 \end{array}$$

These examples may be instructive because they show that two consecutive cues can contribute to the perception of a single segment and that, conversely, 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).

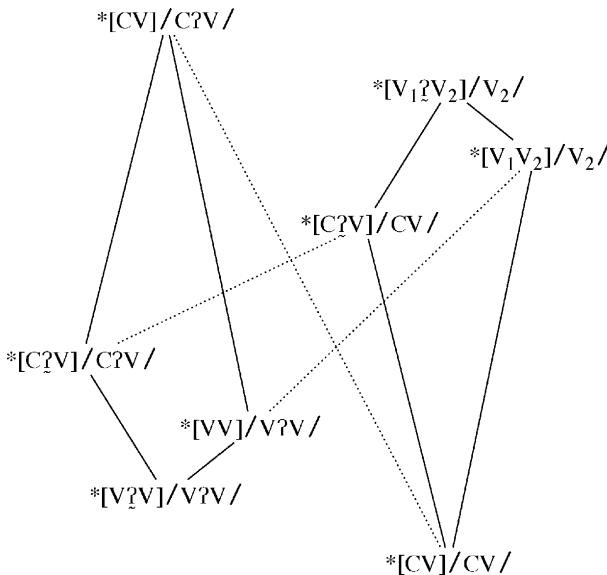
(39) *Ranking of constraints against not perceiving a glottal stop in French*

$$\begin{array}{cccc}
 */V/ & >> & */V/ & >> & */CV/ & >> & */CV/ \\
 \begin{array}{|c|} \hline \text{||} \\ \hline \text{[V?V]} \\ \hline \end{array} & & \begin{array}{|c|} \hline \text{||} \\ \hline \text{[VV]} \\ \hline \end{array} & & \begin{array}{|c|} \hline \text{||} \\ \hline \text{[C?V]} \\ \hline \end{array} & & \begin{array}{|c|} \hline \text{||} \\ \hline \text{[CV]} \\ \hline \end{array} \\
 \end{array}$$

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 [V?V] has a superset of the cues in [VV] and [C?V], which again have more cues than [CV], which has none. The idea that [VV] has better cues for French listeners than [C?V] is again based on the linguistic observation (section 1.5) that [ynəos] is preferred over [ynʔos] (or at least [vwasiləazəʃ] over [vwasilʔazəʃ]). Note that the third constraint is identical to one of Cornulier’s (1981:201) cue constraints for syllable boundaries, namely ‘Postulat II’, which states that a syllable does not contain a pause (\*[C-V]/CV/).

To obtain the relative recoverabilities listed in (7), the eight cue constraints must be ranked along a continuous ranking scale in the vein of Stochastic OT (Boersma, 1997, 1998; Boersma and 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) *Cue constraint ranking for the recovery of a glottal stop in French*



In the figure, the six universal rankings-by-cue are depicted by solid lines. The three dotted lines depict the rankings that make sure that [yn<sub>ə</sub>ʔos] is most often perceived as /yn<sub>ə</sub>ʔos/, that [yn<sub>ɔ̃</sub>ʔos] is most often perceived as /yn<sub>ɔ̃</sub>ʔos/, and that [yn<sub>ə</sub>ʔos] is most often perceived as /yn<sub>ə</sub>ʔos/. The usual perception of [yn<sub>ə</sub>ʔos] as /yn<sub>ə</sub>ʔos/ 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 tableaux in (41) show this for the present case.

(41) Variation in perception

	110.0	98.0	97.0	96.0	95.0	93.0	92.0	91.0	55.0	
[yn <sub>ə</sub> ʔos]	* /VV/	* /CʔV/	* /V/	* /V/	* /CV/	* /CʔV/	* /VʔV/	* /VʔV/	* /CV/	
/yn <sub>ə</sub> ʔos/								*		98%
/yn <sub>ə</sub> ʔos/	*!									2%
/ynos/			*							
	98.0	98.0	97.0	96.0	95.0	93.0	92.0	91.0	55.0	
[yn <sub>ɔ̃</sub> ʔos]	* /VV/	* /CʔV/	* /V/	* /V/	* /CV/	* /CʔV/	* /VʔV/	* /VʔV/	* /CV/	
/yn <sub>ɔ̃</sub> ʔos/						*				76%
/ynos/				*						24%
	98.0	98.0	97.0	96.0	95.0	93.0	92.0	91.0	55.0	
[yn <sub>ə</sub> ʔos]	* /VV/	* /CʔV/	* /V/	* /V/	* /CV/	* /CʔV/	* /VʔV/	* /VʔV/	* /CV/	
/yn <sub>ə</sub> ʔos/							*			92%
/yn <sub>ə</sub> ʔos/	*!									8%
/ynos/			*							
	98.0	98.0	97.0	96.0	95.0	93.0	92.0	91.0	55.0	
[ynos]	* /VV/	* /CʔV/	* /V/	* /V/	* /CV/	* /CʔV/	* /VʔV/	* /VʔV/	* /CV/	
/yn <sub>ɔ̃</sub> ʔos/		*!								0%
/ynos/									*	100%

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 (the structural constraint \*/VV/ is ranked so high that it is never overtaken by any of the other constraints; the very low ranking of \*/CV/ is explained in section 9). These percentages, written to the right of each candidate row, were computed by filtering each auditory form through the tableaux

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 |ynə#ʔos|. The form [ynʔos] 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% 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 [ynəʔos]: if used for implementing a glottal stop, the listener will make an error in no more than 2% of the cases.

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

## 5. Modelling recoverability in production

As noted in section 3.3, the simple listener-oriented grammar in (34) is not listener-oriented enough. In section 1.7, I attributed the preference for [ynəos] over [ynʔos] to the idea that [ə] has better auditory cues than [ʔ], i.e. that [ynəos] is more clearly distinguishable from [ynos] than [ynʔos] is. But in tableau (28), the preference for [ynəos] over [ynʔos] seems to be due to the ranking \*[ʔ] >> \*[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 [ynʔos] is unlikely. The existence of the variants [ynəʔos], [ləʔazaʁ] and [ləʔazaʁ] (section 1.6) cannot have a speaker-based explanation in terms of \*[ʔ] >> \*[ə], since these forms violate *both* articulatory constraints, and are in fact harmonically bounded in (26)–(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%, as (41) shows). I will therefore assume that the preference of [ynəos] over [ynʔos] is wholly or partly due to the same drive that makes [ynəʔos] an option. The theory to be developed has to allow a ranking of \*[ʔ] below \*[ə] and still predict that [ynəos] is the preferred form in (28). It is reassuring to see that in the other 11 tableaux, namely (14acdfgijl) and (25)–(27), the constraint \*[ʔ] 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 section 4.4 can be integrated into an Optimality-Theoretic model of production. Several grammar models have been proposed for production in the literature. I will only discuss models that include phonetic detail, i.e. those enumerated in (42). For the abbreviations see section 4.1.

- (42) *Five production models that integrate phonology and phonetics*
- |    |                                        |                         |
|----|----------------------------------------|-------------------------|
| a. | Serial (section 5.1):                  | UF → SF → phonetic form |
| b. | Two-level (section 5.2):               | UF → phonetic form      |
| c. | Control loop (section 5.3):            | UF → {ArtF → AudF → SF} |
| d. | Stochastic control loop (section 5.4): | UF → {ArtF → AudF → SF} |
| e. | Parallel (section 5.5):                | 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.

### 5.1. The serial production model

The serial model in (42a) is probably the current mainstream view. The first arrow represents *phonology* and the second arrow *phonetic implementation*. ‘Seriality’ here refers to the order of processing in production: the underlying form is first fed into the phonological module, which processes it, and the speaker subsequently uses the output of the phonological module as the input to the phonetic implementation module. In this way, it is impossible for phonetic processing to influence what happens in the phonology. In other words, this view allows no bottom-up processing in production, and no feedback from the phonetics to the phonology.

The separation of the two modules is most vigorously advocated by those researchers who advocate the viewpoint that phonetic detail should not be subjected to investigation by phonological theory at all. A recent instance is Hale and Reiss (1998), who distinguish between the *grammar* (phonology) and the *body* (phonetic implementation). However, even researchers who have worked on the phonetics–phonology interface are able to share the view of a separation of the two modules. For instance, Hayes (1999 [1996:7]) spends a footnote on suggesting that a module of phonetic implementation, separate from the phonological module and following it serially, could perhaps be described by Optimality-Theoretic constraint ranking. According to this view, then, both modules are perhaps regarded as belonging to the grammar, but they are still separate and they are serially related.

Serial grammar models do not seem to work for the case of *h*-spiré, though. I have argued above that the surfacing of the schwa in the form [ynəos] (versus [ynide] and [ynfam]) was partly determined by phonetic considerations of perceptibility, considerations that in the serial model must have their locus in the phonetic implementation module. However, the presence or absence of schwa is a discrete choice that the speaker makes, and all serial models, including that by Hale and Reiss and that by Hayes, would therefore very likely regard this choice as a discrete phonological decision. In other words, the serial grammar models would likely propose that the output of the phonological module is /ynəos/ (versus /ynide/ and /ynfam/). However, if the separability of phonology and phonetics is taken seriously, this discrete phonological decision has no explanation in the module where it occurs, because the explanation is in a module that serially follows it and that it cannot see.

In general, considerations like these provide evidence for non-serial relationships between modules. That is, if a theory of grammar has to propose that a certain phenomenon happens in module A, but the explanation resides in a different module B, then it cannot be the case that module A feeds into module B serially. In some way, the information from module B (in our case, the phonetics) must feed back into module A (the phonology). This leaves several possibilities, each of which I discuss in the following sections: the two modules are not distinct after all (section 5.2), or there is a feedback loop from module B to module A (sections 5.3 and 5.4), or the two modules are processed in parallel (section 5.5).

### 5.2. The two-level production model

The first alternative to the serial production model is the one that does not distinguish between a phonological and a phonetic module, i.e. (42b). Such a production model allows only two levels of representation: it regards the underlying form as the input and the phonetic form as the output, and no privileged intermediate non-lexical discrete level of representation, such as SF, is allowed. Although strong evidence for such an intermediate 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 indeed rejected the role of such a form in production. The first ones to do so were the generative phonologists (Halle, 1959; Chomsky and Halle, 1968), who worked with a derivational production model that contained many intermediate discrete forms but found that none of them could be regarded as privileged enough to warrant being called by a separate name such as ‘phonemic representation’ or ‘phonological surface form’.

Within Optimality Theory as well, 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 they do here in section 4.4), Jun has to propose that these cues are given as invariant surface reflexes for each of the phonological elements that make up the UF; that is, Jun’s model works as if the auditory cues are explicitly specified in the underlying form, and as if faithfulness constraints for these cues regulate their presence on the surface. Other proposals with phonetically detailed output (and therefore often also input) 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 Steriade, the output consists of discrete but low-level features such as [short VOT], which must as a result of their discreteness be regarded as invariantly related to phonological features such as [+voiced]. Steriade admits that her assumption of invariance could be a weak point in her theory (chapter 1:p. 25), and so it turns out to be in the case of *h*-aspiré. In the *h*-aspiré case, after all, one and the same discrete phonological element (/ʔ/) has a large variety of discrete and continuous auditory cues (hiatus, creak, pause, glottal stop), which are related to that phonological element in a French-specific way (most notably the creak, see section 2.3).

While all 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 in some way (either underlyingly or by universal rule) for the auditory cues [Y] and [Z]. It is understandable that Jun and Steriade maintained this simplified view of perception, 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. It seems, then, that accounting for the language-specificity of auditory cues is possible only if we include a third level of representation, the discrete phonological Surface Form, which is related to the underlying form by faithfulness constraints, and to the continuous phonetic form by cue constraints. This idea is pursued in the following three sections, which consider various ways in which the faithfulness constraints (‘phonology’) can interact with the cue constraints (‘phonetics’): either in a phonetic–phonological control loop (sections 5.3 and 5.4), or directly in parallel phonetic–phonological evaluation (section 5.5).

### 5.3. *The control-loop production 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 comprehensive 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 (section 5.2). Since phonetic considerations can influence discrete phonological decisions, this SF level cannot, however, be an

intermediate representation in the serial processing of two modules (section 5.1). In the remaining three models in (42), SF does exist but is not an intermediate representation in processing. All three models are listener-oriented to some extent. The present section considers the control-loop model of (42c), which is the most strongly listener-oriented of the three.

The listener-oriented production model of section 3 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 other words, for every articulatory candidate the speaker computes the SF that she thinks the listener will perceive, and this hypothesized perception is then compared by the speaker to the underlying form; in this way, articulatory constraints can interact directly with faithfulness constraints.

In the control-loop production model, the idea that [ynəos] has better auditory cues for *h*-aspiré than [ynʔos] can be expressed with *probabilistic faithfulness constraints* (Boersma, 2000; Boersma, 2003a:42–44; Boersma and Hamann, 2005). These are listener-oriented faithfulness constraints that take into account the probability (between 0 and 100%) 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*% probability of failing to recover /ʔ/.” As an example, suppose that a speaker is considering to pronounce an underlying |ynəʔos| as the auditory form [ynəos]. From tableau (41) we see that the speaker can compute that the listener will perceive this auditory form as /ynəʔos/ 92% of the time and as /ynos/ 8% of the time. When comparing these hypothetically perceived surface forms to the underlying form |ynəʔos|, the speaker can conclude that if she says [ynəos], glottal-stop faithfulness will be satisfied 92% of the time and violated 8% of the time; this knowledge is represented in the third candidate row in tableau (43). The knowledge about the 8% violation chance means that the speaker knows that pronouncing |ynəʔos| as [ynəos] violates the probabilistic faithfulness constraint MAX (ʔ, 3%) but does not violate the constraint MAX (ʔ, 10%); this information is included in the violation pattern in the third candidate of tableau (43). Tableau (43) also indicates that the candidate [ynʔos], which has poorer *h*-aspiré cues than [ynəos], violates both MAX (ʔ, 3%) and MAX (ʔ, 10%), because the chances of non-transmission of the glottal stop, according to (41), are 24% in this case. When compared to the winning candidate in (43), the candidate [ynʔos] is therefore ruled out by its more severe violation of probabilistic faithfulness. The candidate [ynos], finally, has, according to (41), 100% chance of being perceived without a glottal stop, so it certainly violates the high-ranked MAX (ʔ, 80%).

(43) *Une hausse in the probabilistic faithfulness model*

ynəʔos	MAX (C)	DEP (ə)	MAX (ʔ, 80%)	MAX (ʔ, 10%)	*[ə]	*[ʔ]	MAX (ʔ, 3%)	MAX (V)	*/CC/	MAX (C)
[ynəʔos] → /ynəʔos/ 98% /ynos/ 2%					*	*!				
[ynʔos] → /ynʔos/ 76% /ynos/ 24%				*!		*	*	*	*	
☞ [ynəos] → /ynəʔos/ 92% /ynos/ 8%					*		*			
[ynos] → /ynos/ 100%			*!	*			*	*		



Tableau (43) exemplifies a more listener-oriented solution than tableau (28), where the greater harmony of [ynəos] as compared to [ynʔos] was ascribed to the ranking  $*[\ʔ] \gg *[\ə]$ , i.e. to a difference in articulatory effort. In tableau (43), where  $*[\ə]$  happens to outrank  $*[\ʔ]$ , the form [ynʔos] is instead ruled out as a result of its lower perceptibility. This tableau, therefore, gives an affirmative answer to one of the questions posed at the end of section 3.3, namely whether the analysis could be made even more listener-oriented than the one in (28) or (34). The listener-orientedness of (43) finally corresponds to the informal account proposed in section 1.7, especially to the recoverability hierarchy presented in (7).

The control-loop model with probabilistic faithfulness can also account for variation. In (43) we can see that occurrences of [ynʔos] can be described by a variable ranking of MAX (ʔ, 10%) and  $*[\ə]$ , i.e. by a variable weighing of the enhanced perceptibility brought about by producing schwa and the increased effort required for articulating it, and that the occasional occurrence of the enhanced form [ynəʔos] can be described by a variable ranking of  $*[\ʔ]$  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. This answers another one of the questions posed at the end of section 3.3 in the affirmative.

To sum up, probabilistic faithfulness constraints seem to have several kinds of advantages. On the empirical level, they directly take into account percentages of perception such as those computed in section 4.4, and therefore provide a close link between constraint ranking in production and universal and language-specific cue quality; in our case, the result was a fully listener-oriented account of the preference for [ynəos] and the occasional occurrence of the enhanced [ynəʔos]. On the theoretical level, probabilistic faithfulness constraints provide a link between two seemingly disparate phenomena: *auditory enhancement* and *Steriade's (1995) licensing by cue*. This link is a result of their universal ranking: it is worse to pronounce an underlying element as something that has a larger probability of being perceived as a different category, than to pronounce it as something that has a lower probability of being perceived as a different category. For low  $p$  values (as in the *h*-aspiré example), the probabilistic faithfulness constraints express the auditory enhancement of phonological elements, as in (43), while for high  $p$  values, they turn out to express licensing by cue. For example, the fact that plosives have better place cues than nasals is reflected in their recoverabilities: if you pronounce a [t] as [p], it may have only 3% chance of being perceived as /t/, whereas if you pronounce an [n] as [m], it may still have 10% chance of being perceived as /n/. Therefore, the universal ranking IDENT (place, 97%)  $\gg$  IDENT (place, 90%) reflects Steriade's ranking-by-cue, which says that IDENT (place/plosive)  $\gg$  IDENT (place/nasal).

However, probabilistic faithfulness constraints also have several disadvantages, the largest being that there is no known on-line learning algorithm for these 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 and 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.<sup>29</sup> The problem now is that this whole procedure is possible only if there is a single virtually produced SF; with the multiple Surface Forms in every candidate cell of (43), there is no single virtually produced SF, and hence no way to determine whether the virtually produced SF is identical to the perceived SF or not.

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 cue constraints, i.e. (41), in a set of heterogenous constraints (namely, faithfulness constraints) in the production grammar.

Fortunately, there are ways to get rid of the disadvantages of probabilistic faithfulness and keeping either some of the advantages (section 5.4) or all of them (section 5.5).

#### 5.4. The stochastic control-loop production model

The problems with probabilistic faithfulness can be solved by allowing a stochastic evaluation of the perception mapping in the candidate cells. From the tableaux in (41) we see, for instance, that the four candidate phonetic forms [ynəʔos], [ynʔos], [ynəos], and [ynos] are perceived as the four surface forms /ynəʔos/, /ynʔos/, /ynəʔos/, and /ynos/, respectively, in  $98 \cdot 76\% \cdot 92\% \cdot 100\% = 68.5\%$  of the cases (if the perceptions of the four candidates are independent of each other). It can now be argued that in those 68.5% of the production cases, tableau (28) has to be changed to (44).

#### (44) The stochastic control loop, 68.5% of the time

[ynəʔos]	MAX (C)	DEP (ə)	MAX (ʔ)	*[ʔ]	*[ə]	MAX (V)	*/CC/	MAX (C)
[ynəʔos] → /ynəʔos/				*!	*			
[ynʔos] → /ynʔos/				*!		*	*	
<b>[ynəos] → /ynəʔos/</b>					*	*		
[ynos] → /ynos/			*!			*		

#### (45) The stochastic control loop, 5.9% of the time (minority perceptions written in bold)

[ynəʔos]	MAX (C)	DEP (ə)	MAX (ʔ)	*[ʔ]	*[ə]	MAX (V)	*/CC/	MAX (C)
[ynəʔos] → /ynəʔos/				*	*!			
<b>[ynʔos] → /ynʔos/</b>				*		*	*	
[ynəos] → / <b>ynos</b> /			*!		*	*		
[ynos] → /ynos/			*!			*		

<sup>29</sup> The comparison strategy is similar to that in Tesar and Smolensky’s (1998, 2000) Robust Interpretive Parsing with Constraint Demotion.

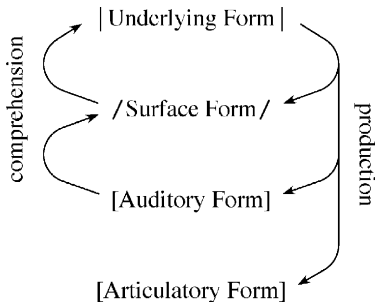
In this way we can create 16 tableaux, one for every possible combination of perceptions of the four candidates. From tableaux (41), for instance, we see that the four candidate phonetic forms [ynəʔos], [ynʔos], [ynəos], and [ynos] are perceived as the four surface forms /ynəʔos/, /ynʔos/, /ynos/ (!), and /ynos/, respectively, in  $98\% \cdot 76\% \cdot 8\% \cdot 100\% = 5.9\%$  of the cases, as illustrated in (45). Tedious computations show that if the constraints have a fixed ranking as in (44) and (45), the phonetic–surface pair [ynəos] → /ynəʔos/ wins in total 92.0% of the time, [ynʔos] → /ynʔos/ 6.1% of the time, and [ynəʔos] → /ynəʔos/ 1.9% of the time. In this ‘stochastic control loop’ model, the variation in production therefore comes to reflect the variation in perception, even if the constraints in (44) and (45) have a fixed ranking (any variable ranking of these constraints would provide an additional source of variability).

When assessing this model in comparison with the probabilistic faithfulness model of section 5.3, we see that learnability is no longer a problem, because every candidate cell contains a single SF. There is a drawback, however. It turns out to be no longer possible for the speaker to base an obligatory decision on gradient perceptibility. One can see, for instance, that making [ynəos] the winner in (44) and in several of the remaining 14 tableaux heavily relies on the ranking  $*[\ʔ] \gg *[\ə]$ . In other words, the ranking  $*[\ə] \gg *[\ʔ]$  is no longer compatible with a preference for [ynəos] over [ynʔos], a fact that brings us back from the listener-orientedness of sections 5.3 and 1.7 to that of (28) and (34). The degree of listener-orientedness of this model is therefore smaller than that of the model in section 5.3. The next section presents a model that combines the advantages of the models discussed in section 5.3 and the present section.

5.5. *The parallel phonological–phonetic production model*

Given the disadvantages of the first four production models in (42), I propose that the remaining production 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 sections 5.3 and 5.4. A summary is shown in (46).

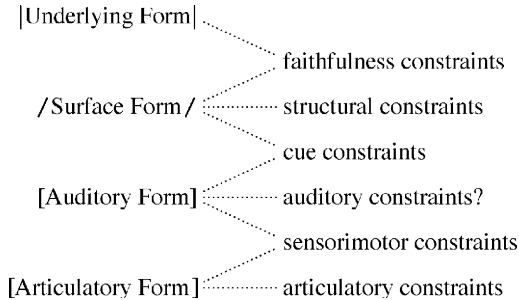
(46) *A bidirectional model of phonology and phonetics with parallel production*



This picture combines the comprehension and production models. As in the other models, comprehension consists of two serial modules, prelexical perception and word recognition. In

this model, however, phonological and phonetic production together consist of a single parallel mapping. The input of production is the usual Underlying Form, but its output consists of freely combined triplets of Articulatory Form, Auditory Form and Surface Form. The four representations are related by constraints that work in both directions of processing (comprehension and production), as shown in (47).

(47) *Constraints in bidirectional phonology and phonetics*



The constraints in (47) are shared with the control-loop models. Four of the constraint types have been discussed before: faithfulness constraints evaluate the similarity between UF and SF both in word recognition and in production, structural constraints evaluate SF both in prelexical perception and in production, articulatory constraints evaluate ArtF (in production only), and the cue constraints evaluate the relation between AudF and SF both in prelexical perception and in production. This model is thus truly bidirectional, as will be stressed repeatedly below.<sup>30</sup> 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 assume 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, and also any representations above UF such as lexical meaning, which could be connected to UF via lexical constraints (Boersma, 2001; Apoussidou, 2006).

<sup>30</sup> 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.  $|je + \lambda k + hrek + ?| \rightarrow /j\lambda khrek?/$ . In the comprehension direction, the same constraint can only express *optional* vowel insertion, i.e. a perceived  $/j\lambda k/$  must be compatible with  $|j\lambda k|$  as well as with  $|je\lambda k|$ . 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  $*/|V_1V_2|/V_1/ \gg */|V_1V_2|/V_2/$ , but this would require OT, which was not available to Lakoff.

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

## 6. A parallel phonological–phonetic account of *h*-aspiré

This section shows that the main facts about *h*-aspiré can be handled well with the parallel phonological–phonetic production model of section 5.5. A comparison of the three listener-oriented models is given in section 6.6.

### 6.1. The variable production of *h*-aspiré

There is variation not only in the perception of *h*-aspiré (section 4.4), but as mentioned in section 1.6, there is a high degree of variation in its production as well. I will ignore here the phonetically detailed variability in what the symbol [ʔ] stands for, and discuss only the variability with respect to what Meisenburg and Gabriel (2004) regarded as categorical variation, namely the variability in whether something transcribable as [ʔ] is present or not, and the variability in whether [ə] is present or not.

#### 6.1.1. Variation after a consonant

The first case of variation to discuss is that between [kɛlʔazaʁ] and [kɛlazaʁ]. As discussed in section 1.3, the pronunciation with the glottal stop is obligatory for some speakers, optional for others. I will first discuss the grammar for those speakers for whom it is optional, deferring the obligatory case to section 6.3. Meisenburg and Gabriel propose that Tranel and 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 /kɛ.l.a.zaʁ./ and /kɛ.la.zaʁ./, as can be inferred from our tableau (16). Meisenburg and Gabriel themselves do discuss the variation (by noticing in their data 6 occurrences of [keleʁo], from |kɛl#ʔeʁo| ‘what a hero’), but do not provide an analysis.<sup>31</sup> In the present framework, the variation can be attributed to a variable ranking of MAX(?) among the cue constraints of section 4.4. This is shown in tableau (48), where it is further assumed that most constraints undominated in (34), namely MAX(C), MAX(V), DEP(C), DEP(V), and DEP(ə) are ranked at 110.0 (thinkable candidates that violate any of these constraints were excluded from the tableau), and the constraints ranked low in (34), namely \*/CC/, MAX(V), and MAX(C), are ranked at 75.0, 75.0, and 65.0, so that they are irrelevant to determining the winner (see sections 6.2 and 9 for details).

The parallel production tableau (48) is slightly more complicated than the control-loop tableaux (43)–(45) in that it employs four rather than three types of constraints. For most constraint types, the formulation in terms of brackets or slashes makes explicit what form or relation the constraint evaluates. Thus, there are structural constraints written like \*/X/ that evaluate the phonological part of each candidate, articulatory constraints written like \*[X] that evaluate the phonetic (articulatory) part of each candidate, and cue constraints written like \*[X]/Y/ (or an equivalent arboreal version) that evaluate the relation between the phonetic (auditory) part and the phonological part of each candidate. Finally, there are faithfulness

<sup>31</sup> They do analyse the variation between [nɛtʃɟɛʁwa] and [nɛtʃɟɛʁwa] ‘clear Hungarian’, but attribute it to variation in the lexicon, i.e. |nɛt#(?)ɟɛʁwa|, on the basis of the variation between [tuʔɟɛʁwa] and [tutʃɟɛʁwa] for |tut#(?)ɟɛʁwa| ‘all Hungarian’, a phenomenon apparently not found with |tut#ʔazaʁ| (\*[tutazaʁ]), which must therefore have an invariable underlying glottal stop.



constraints that evaluate the relation between the phonological part of each candidate and the underlying form, and that could therefore have been written explicitly in the form  $*|X|/Y/$  but are instead written more traditionally as  $\text{MAX}(X)$  and  $\text{DEP}(X)$  (short for  $*|X|//$  and  $*| |X/$ , respectively).

In tableau (48) the cue constraints are ranked at the same height as in section 4.4. Their negative formulation (e.g.  $*[C\text{?}V]/C\text{?}V/$  means ‘‘a postconsonantal creaky pause at AudF is not perceived as a glottal stop at SF’’) allows a full bidirectional interpretation (e.g.  $*[C\text{?}V]/C\text{?}V/$  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 section 3 have been interspersed among the eight cue constraints. The structural constraint  $*/VV/$  has been put at the top, as in section 4.4. 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. I will discuss the three main candidates.

The most common winning candidate is the phonetic–phonological pair  $[k\text{el}\text{?}aza\text{?}]/k\text{el}\text{?}aza\text{?}/$ . Phonetically, this form is articulated and heard with a creaky pause, i.e. the articulatory and auditory forms can be simplifyingly abbreviated together as  $[k\text{el}\text{?}aza\text{?}]$ . Phonologically, it has an SF ( $/k\text{el}\text{?}aza\text{?}/$ ) that contains a glottal stop. It is important to realize that the phonetic form and the phonological form are related bidirectionally. From the speaker’s standpoint, one can say that the phonological SF  $/k\text{el}\text{?}aza\text{?}/$  is phonetically implemented as the ArtF  $[k\text{el}\text{?}aza\text{?}]$ . From a listener-oriented standpoint, one can say that the speaker thinks that the phonetic AudF  $[k\text{el}\text{?}aza\text{?}]$  will be perceived as the phonological SF  $/k\text{el}\text{?}aza\text{?}/$ . These two statements are equivalent. A parallel model of phonological and phonetic production necessarily comes with this dual speaker-listener orientation.

The second grammatical output form for an underlying  $|k\text{el}\#?aza\text{?}|$  is the pair  $[k\text{el}aza\text{?}]/k\text{el}aza\text{?}/$ . If this candidate is chosen, the speaker phonetically pronounces  $[k\text{el}aza\text{?}]$  and either ‘thinks’ that she has phonologically produced  $/k\text{el}aza\text{?}/$  or (equivalently) ‘knows’ that the listener will phonologically perceive  $/k\text{el}aza\text{?}/$ ; by comparing the underlying  $|k\text{el}\#?aza\text{?}|$  with the surface  $/k\text{el}aza\text{?}/$ , the speaker ‘knows’ that she has violated  $*|?|/$ , i.e.  $\text{MAX}(?)$ . This pair will win, then, if at evaluation time the importance of phonologically producing (or transmitting) the underlying glottal stop, i.e. the ranking of  $\text{MAX}(?)$ , is less than the articulatory effort of the creaky pause, i.e. the ranking of  $*[?] /$ , or less than the perceptibility of a postconsonantal creaky pause, i.e. the ranking of  $*[C\text{?}V]/C\text{?}V/$ .

The third possible candidate, which is predicted to occur in 4% of the cases, is  $[k\text{el}aza\text{?}]/k\text{el}\text{?}aza\text{?}/$ . This form is interesting, since it did not occur as a candidate in tableau (25). Both of the candidates  $[k\text{el}\text{?}aza\text{?}]/k\text{el}\text{?}aza\text{?}/$  and  $[k\text{el}aza\text{?}]/k\text{el}aza\text{?}/$  are *listener-optimal* (the term is by Jäger (2003)), i.e. their SF part is the winner (or perhaps most common winner) in the perception tableau where their AudF part is the input. The form  $[k\text{el}aza\text{?}]/k\text{el}\text{?}aza\text{?}/$  is not listener-optimal. Instead, this looks very much like a speaker who decides to produce  $/k\text{el}\text{?}aza\text{?}/$  but manages only to articulate the impoverished  $[k\text{el}aza\text{?}]$ . If we assume that  $/k\text{el}aza\text{?}/$  is also implemented as  $[k\text{el}aza\text{?}]$ , 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  $[k\text{el}aza\text{?}]$ , has hallucinated that she has produced the faithful  $/k\text{el}\text{?}aza\text{?}/$ , thereby violating the cue constraint  $*[CV]/C\text{?}V/$ , which is identical to Cornulier’s (1981:201) Postulate V.1 and to Cornulier’s (1981:210) statement that syllable boundaries between C and V should have a phonetic (I would say: auditory) correlate. By allowing such hallucinations in production, the parallel model is thus less listener-oriented than the control-loop model of section 3. See section 6.6 for more discussion.

### 6.1.2. Intervocalic variation

The second case of variation is that between hiatus and creaky pause. The variation between [leazɤ] /leʔazɤ/ and [leʔazɤ] /leʔazɤ/ corresponds to that between Meisenburg and Gabriel's forms [tuɔ̃gɤwa], which they observed three times in their data, and [tuʔɔ̃gɤwa], which they observed five times. For Meisenburg and Gabriel, forms like [leazɤ] violate MAX(?), and forms like [leʔazɤ] violate \*VʔV. Gabriel and Meisenburg (2005) model this variation with Stochastic OT by ranking MAX(?) just above \*VʔV. In the present framework, the account for the variation must be different, since both [leazɤ] and [leʔazɤ] tend to be perceived with a glottal stop. According to section 4.4, however, [leʔazɤ] has even better auditory cues for the underlying glottal stop than [leazɤ] has, which according to (41) should lead to a glottal stop recoverability of 98% for [leʔazɤ]. The form [leʔazɤ] now becomes a possible realization if its gain in auditory cues by satisfying \*[VV]/VʔV/ approaches the articulatory cost of pronouncing the creaky pause, i.e. \*[ʔ]. A similar variation is that between [ləazɤ] and [ləʔazɤ]. Tableau (49) formalizes this.

In tableau (49), the choice between [ləazɤ] and [ləʔazɤ] is determined by the relative ranking of \*[ʔ] 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.

### 6.1.3. Variation in *une hausse*

The largest variation is found in the form |ynə#os|, which Meisenburg and Gabriel (2004) report can be pronounced as [ynəos] (twice), [ynʔos] (6 times), and [ynəʔos] (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 (section 1.5), I will assume that some of the 6 tokens of [ynʔos] must be ascribed to those speakers, and therefore model a speaker for whom [ynəos] is the preferred pronunciation and [ynʔos] is just a minority pronunciation. Tableau (50) gives the analysis.

Although \*[ə] and \*[ʔ] are ranked equally high, [ynəos] is now finally preferred over [ynʔos] because the former is favoured by the cue constraints (cf. section 5.4). The phonetically enhanced form [ynəʔos] 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 (50), 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 (50), 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 tableaux of section 3. Six more candidates appear in tableau (50), since AudF and SF can be combined freely. Nevertheless, five of these forms have a frequency below 0.1%, although four 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.

### 6.1.4. Evaluation

Gabriel and Meisenburg (2005) modelled their observed variation by applying Stochastic OT to their two-representation grammar model (section 2.4), 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,



(49) *Variation between creaky pause and hiatus*

	110.0	98.0	97.0	96.5	96.0	95.0	93.0	92.5	92.5	92.0	91.0	55.0	
$ \text{[əʔ}^{\text{h}}\text{azazək} $	*/VV/	* /CʔV/ \\ [CV]	* /V/ \\ [VʔV]	M <sub>AX</sub> (ʔ)	* /V/ \\ [VV]	* /CV/ \\ [CʔV]	* /CʔV/          [CʔV]	*[ə]	*[ʔ]	* /VʔV/ \\ [VV]	* /VʔV/          [VʔV]	* /CV/    [CV]	
$ \text{[əʔ}^{\text{h}}\text{azazək}  / \text{[əʔ}^{\text{h}}\text{azazək} $								*	*		*		31%
$ \text{[əʔ}^{\text{h}}\text{azazək}  / \text{[ə}^{\text{h}}\text{azazək} $	*!			*				*	*				
$ \text{[əʔ}^{\text{h}}\text{azazək}  / \text{[ə}^{\text{h}}\text{azazək} $			*	*				*	*				
$ \text{[ə}^{\text{h}}\text{azazək}  / \text{[əʔ}^{\text{h}}\text{azazək} $					*			*		*			58%
$ \text{[ə}^{\text{h}}\text{azazək}  / \text{[ə}^{\text{h}}\text{azazək} $	*!			*	*			*					
$ \text{[ə}^{\text{h}}\text{azazək}  / \text{[ə}^{\text{h}}\text{azazək} $				*	*			*					
$ \text{[ə}^{\text{h}}\text{azazək}  / \text{[ə}^{\text{h}}\text{azazək} $				*								*	9%
$ \text{[ə}^{\text{h}}\text{azazək}  / \text{[əʔ}^{\text{h}}\text{azazək} $		*											2%

(50) Triple variation

	110.0	98.0	97.0	96.5	96.0	95.0	93.0	92.5	92.5	92.0	91.0	55.0
	* /VV/	* /C <sup>2</sup> V/   /CV/   [CV]	* /V/   [V <sup>2</sup> V]	MAX (V)	* /V/   [VV]	* /CV/   [C <sup>2</sup> V]	* /C <sup>2</sup> V/   [C <sup>2</sup> V]	* [a]	* [ɔ]	* /V <sup>2</sup> V/   [VV]	* /V <sup>2</sup> V/   [V <sup>2</sup> V]	* /CV/   [CV]
[yna#ʔos]												
ɛ <sup>3</sup> [ynaʔos] /ynaʔos/								*	*		*	
[ynaʔos] /ynaos/	*!			*				*	*			
[ynaʔos] /ynos/			*					*	*			
ɛ <sup>3</sup> [ynaʔos] /ynʔos/							*		*			
[ynʔos] /ynos/				*				*	*			
ɛ <sup>3</sup> [ynaos] /ynaʔos/								*	*	*		
[ynaos] /ynaos/	*!			*				*	*			
[ynaos] /ynos/				*				*	*			
[ynos] /ynʔos/		*										
[ynos] /ynos/				*								*

17%

32%

45%

1%

5%

whereas the present section assumes the more commonly held view that all forms in the language should derive from a single constraint ranking, so Gabriel and 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 [kɛlʔazaʁ] ~ [kɛlazaʁ] and [ynəos] ~ [ynʔos], it probably is so in the present case. The thing not explained in this section is the grammaticality difference between [kɛlazaʁ] and \*[lazaʁ], because the difference between their predicted frequencies (19% and 11%, respectively) is not large. This is addressed in the next section.

6.2. Improving the ranking: low \*[ə]

Whereas [kɛlazaʁ] is a form both accepted in the literature and attested in reality (sections 1.3, 6.1.1), the form \*[lazaʁ] is rejected by all authors mentioned (especially clear about this difference is Encrevé, 1988:197). This differential grammaticality is not yet fully accounted for by the ranking in section 6.1. But now that the point has been proven that a preference of [ynəos] over [ynʔos] need not be due to a ranking of \*[ʔ] over \*[ə], i.e. that it can be due to a difference in auditory cue quality, we are free to rank these two constraints lower than in (48) through (50). The trick to reduce the occurrences of \*[lazaʁ] is to move \*[ə] further down the hierarchy, namely to 80.0 (still above \*/CC/ and MAX(V)). The frequencies of the resulting forms are listed in (51).

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

UF	{ArtF/AudF, SF}	Frequency
kɛl#ʔazaʁ	[kɛlʔazaʁ] /kɛlʔazaʁ/	81%
	[kɛlazaʁ] /kɛlʔazaʁ/	4%
	[kɛlazaʁ] /kɛlazaʁ/	15%
lə#ʔazaʁ	[ləʔazaʁ] /ləʔazaʁ/	34%
	[ləazaʁ] /ləʔazaʁ/	64%
	[lazaʁ] /lazaʁ/	2%
	[lazaʁ] /lʔazaʁ/	0.4%
ynə#ʔos	[ynəʔos] /ynəʔos/	28%
	[ynʔos] /ynʔos/	8%
	[ynəos] /ynəʔos/	61%
	[ynos] /ynʔos/	0.4%
	[ynos] /ynos/	2%

The illegal forms have indeed shrunk to no more than 2% occurrence, while [kɛlazaʁ] still works as before. All of the predicted frequencies in (51) are now plausible, i.e. they are compatible with the relative acceptability judgments by Cornulier (1981), with the corpus frequencies found by Encrevé (1988), and with the frequencies found in the elicitation tasks by Meisenburg and Gabriel (2004). The preference of [ynəos] over [ynʔos] has become greater than it was in section 6.1.3. 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 (51) 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 [ʔ] may not be the articulatorily most effortful vowel and consonant, respectively. In fact, [ə] is likely to be the *easiest* vowel. In all tableaux from section 3.2 on, \*[ə] could be replaced with the more general \*[V], and \*[ʔ] 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 (an equally high-ranked) \*[V<sub>1</sub>V<sub>2</sub>]/V<sub>1</sub>V<sub>2</sub>/ there would be no structural constraints left at SF (which according to Wheeler and Touretzky (1993) would be advantageous for connectionist modelling). Future theorizing will tell us whether such simplifications are viable.<sup>32</sup>

### 6.3. Variations in enchainment between speakers and within the lexicon

I will now discuss the grammars of several French speakers who diverge from the speakers modelled in sections 6.1 and 6.2.

The speakers of sections 6.1 and 6.2 had an optional phonetic glottal stop for |kɛl#ʔazaʁ|. Speakers for whom this glottal stop is obligatory (as must exist according to Dell (1973:256)) simply have a higher ranking of MAX(ʔ). If we rank this constraint at 99.0 rather than at 96.5, and have all other constraints ranked equally high as in (51), the pair [kɛlazaʁ]/kɛlazaʁ/ is reduced from 15% to a mere 2%, and everything else stays the same.

A more complicated grammar is that in which there are two groups of lexical items. According to Cornulier (1981:210–221), enchainment is forbidden for some lexical items, e.g. in \*/.kɛ.le.ʁo./ for |kɛl#ʔeʁo| ‘what a hero’, while it is optional in /.kɛ.la.zaʁ./ for |kɛl#ʔazaʁ|. Cornulier states that this lexical contrast exists for only some speakers, and indeed Encrevé (1988:200) and Meisenburg and Gabriel (2004) did observe many instances of [kɛleʁo] in their elicitations. For the speakers under discussion, Cornulier analyzes the behaviour of |ʔeʁo| as being due to a ‘(strong) initial syllable separation constraint’, i.e. to a high-ranked CONTIGUITY(V) in OT terms (given the syllable-island analysis), and the behaviour of |ʔazaʁ| as being due to a ‘(weak) initial syllable separability constraint’, i.e. to a low-ranked CONTIGUITY(V) in OT terms. Our analysis of underlying |ʔ| will have to utilize two separate underlying arbitrary symbols, say |ʔ<sub>1</sub>| and |ʔ<sub>2</sub>| (a solution attributed to Freeman (1975) by Tranel (1981:305)), so that the lexicon can distinguish between |ʔ<sub>1</sub>azaʁ| and |ʔ<sub>2</sub>eʁo|. If MAX(ʔ<sub>1</sub>) is ranked at 96.5 (as in (51)) and MAX(ʔ<sub>2</sub>) is ranked at 99.0 (cf. the previous paragraph), the underlying form |kɛl#ʔ<sub>1</sub>azaʁ| will be realized as [kɛlazaʁ]/kɛlazaʁ/ in 15% of the cases, whereas |kɛl#ʔ<sub>2</sub>eʁo| will be realized as [kɛleʁo]/kɛleʁo/ in only 2% of the cases.

A third group of speakers could be those who freely insert a schwa in [kɛləazaʁ] (section 1.3). Lowering DEP(ə) to 99.0 causes 1.5% cases of [kɛləazaʁ]/kɛləʔazaʁ/. According to Tranel (1981:287) the occurrence of such forms is evidence against underlying schwas in general. That conclusion does not seem to be warranted. If Tranel were right, the masculine [kɛlə(ʔ)azaʁ] and the feminine [kɛlə(ʔ)ʔt] (from |kɛl(ə) + ʔt|) ‘what a shame’ should have the same percentage of occurrence. If, on the other hand, Cornulier is right about the existence of underlying schwas, then the two percentages will be different. From (51) we can see that with our ranking, [kɛlə(ʔ)ʔt]

<sup>32</sup> For instance, collapsing \*[ʔ] with \*[C] seems to be incompatible with the low ranking of \*/VV/ proposed in sections 8.1 and 9, because this combination would not let liaison consonants surface.

occurs in  $28 + 61 = 89\%$  of the cases, i.e. much more often than  $[\text{k}\epsilon\text{l}\text{ə}(\text{ʔ})\text{azax}]$ . Even according to Tranel (1981:287), the two percentages are different. I interpret this as evidence for underlying schwa in the feminine  $[\text{k}\epsilon\text{l}\text{ə}]$  but not in the masculine  $[\text{k}\epsilon\text{l}]$  (Tranel (1981) ascribes the difference to orthographic influence).

Both the observed between-speaker variation and the observed lexical variation of enchainment thus turn out to reflect straightforward differences in the grammar and in the lexicon.

#### 6.4. Glides

A possible special case of *h*-aspiré that has been ignored in this article is that of the glides  $[\text{j}]$ ,  $[\text{w}]$ , and  $[\text{ɥ}]$ . Many words that in isolation would start with one of these glides, exhibit elision and liaison, just like vowel-initial words:  $[\text{lezj}\text{ø}]$  ‘the eyes’,  $[\text{l}\text{ɥ}\text{i}\text{l}]$  ‘the oil’,  $[\text{l}\text{waz}\text{o}]$  ‘the bird’. Other words that in isolation start with one of these glides, fail to exhibit elision or liaison, just as consonant-initial or *h*-aspiré words:  $[\text{lej}\text{øt}]$  ‘the yachts’,  $[\text{l}\text{ø}\text{ɥ}\text{i}\text{t}]$  ‘the eight’,  $[\text{l}\text{ø}\text{w}\text{i}\text{s}\text{k}\text{i}]$  ‘the whisky’. Several analyses have been given for this contrast, either in terms of an *h*-aspiré contrast, or in terms of a vowel-consonant contrast, or in terms of both.

For Martinet (1933), all these words start with an underlying vowel, and the contrast is one of *h*-aspiré. With our notation for the morphology, the six underlying forms would be  $|\text{lez}\#i\text{ø}|$ ,  $|\text{la}\#y\text{i}\text{l}|$ ,  $|\text{l}\text{ə}\#u\text{az}\text{o}|$ ,  $|\text{lez}\#hi\text{øt}|$ ,  $|\text{l}\text{ə}\#hy\text{i}\text{t}|$ ,  $|\text{l}\text{ə}\#hui\text{s}\text{k}\text{i}|$ . Martinet favours the *h*-aspiré analysis for economical reasons, because French can then do without the phonemes  $/\text{w}/$  and  $/\text{ɥ}/$ .<sup>33</sup> Martinet is aware of the fact, however, that by regarding all glides (in onsets) as vowels, he ignores any potential distinction between the monosyllabic  $[\text{pje}]$  ‘foot’ and a possibly disyllabic  $[\text{nie}]$  ‘deny’.

Another possible analysis, which Martinet considers but rejects (because it requires the otherwise superfluous phonemes  $/\text{w}/$  and  $/\text{ɥ}/$ ), is that of a contrast between vowels and glides:  $|\text{lez}\#i\text{ø}|$ ,  $|\text{la}\#y\text{i}\text{l}|$ ,  $|\text{l}\text{ə}\#u\text{az}\text{o}|$ ,  $|\text{lez}\#j\text{øt}|$ ,  $|\text{l}\text{ə}\#ɥ\text{i}\text{t}|$ ,  $|\text{l}\text{ə}\#w\text{i}\text{s}\text{k}\text{i}|$ . Such an analysis (or better, a nearly equivalent analysis in terms of a contrast between glides in the nucleus and glides in the onset) is favoured by Kaye and Lowenstamm (1984:135) and Encrevé (1988:200).

Cornulier (1981:226) is not satisfied with either proposal, because for him the vowel-glide distinction has to exist independently of the *h*-aspiré contrast. Thus,  $|\text{ye}|$  ‘hiss at’ has an underlying syllable boundary (i.e. it is an *h*-aspiré word) because of the way it is pronounced after pronouns like  $|\text{vu}\text{z}|$  ‘you’ (i.e.  $[\text{vu}\text{t}\text{e}]$  ‘you hiss at’, not  $[\text{vuz}\text{t}\text{e}]$ ), although it must have an underlying vowel because it can optionally be pronounced with a vowel (i.e. as either  $[\text{t}\text{e}]$  or  $[\text{ye}]$ ). By contrast,  $|\text{j}\text{ø}|$  ‘eyes’ does not have an underlying syllable boundary (i.e. it is not an *h*-aspiré word), although it is never pronounced with a vowel (i.e. always as  $[\text{j}\text{ø}]$ ). Therefore, there is no correlation between having an *h*-aspiré and being a glide: words without *h*-aspiré can have non-optional glides ( $|\text{waz}\text{o}| \rightarrow [\text{waz}\text{o}]$ ) or optional glides ( $|\text{ui}|$  ‘hearing’  $\rightarrow$   $[\text{ui}, \text{wi}]$ ), and words with *h*-aspiré can equally have non-optional glides ( $|\text{w}\text{i}\text{s}\text{k}\text{i}| \rightarrow [\text{w}\text{i}\text{s}\text{k}\text{i}]$ ) or optional glides ( $|\text{ye}| \rightarrow [\text{ye}, \text{t}\text{e}]$ ). Following Cornulier’s distinctions, I write the six underlying forms as  $|\text{lez}\#j\text{ø}|$ ,  $|\text{la}\#ɥ\text{i}\text{l}|$ ,  $|\text{l}\text{ə}\#w\text{az}\text{o}|$ ,  $|\text{lez}\#\text{ʔ}\text{øt}|$ ,  $|\text{l}\text{ə}\#\text{ɥ}\text{i}\text{t}|$ ,  $|\text{l}\text{ə}\#\text{w}\text{i}\text{s}\text{k}\text{i}|$ .

It has been noted that *h*-aspiré words with glides behave differently in several ways from *h*-aspiré words with vowels. Typically, schwa or pause can be dispensed with after  $|\text{ə}|$  words if a vowel precedes. According to Cornulier (1981:206),  $|\text{by}\#\text{l}\text{ə}\#\text{jau}\text{xt}|$  ‘drank the yoghurt’ can be pronounced  $[\text{by}\text{l}\text{jau}\text{xt}]$ , and  $|\text{vy}\#\text{s}\text{ə}\#\text{w}\text{est}\text{ɛ}\text{x}\text{n}|$  ‘seen that western’ can be pronounced  $[\text{vys}\text{w}\text{est}\text{ɛ}\text{x}\text{n}]$ ; this contrasts with the inacceptability of  $*[\text{palazax}]$  for  $|\text{paz}\#\text{l}\text{ə}\#\text{azax}|$  ‘not the

<sup>33</sup> Martinet still needs the phoneme  $/\text{j}/$ , but only in final position, e.g. to distinguish  $|\text{p}\text{ei}|$  ‘land’ from  $|\text{p}\text{ej}|$  ‘pay-3Sg’.

risk', and even that of \*[ʒəlɥɛ] for [ʒə#lə#.yɛ] 'I hissed at it' (phonetic glide but underlying vowel). Cornulier (1981:214) even mentions the possibility ("with some speakers") of utterance-initial [ljauʁt] for [lə#.jauʁt] 'the yoghurt' and [swɛstɛʁn] for [sə#.wɛstɛʁn] 'that western'.

The question is how these facts can be incorporated in the present theory. First, the ungrammaticality of \*[ʒəlɥɛ] can straightforwardly be accounted for by a tableau like (50) (just as \*[palazaʁ]), because the underlying form [ʔyɛ] has a vowel after the glottal stop, just as [ʔos] does. That is, forms pronounced [ʒəlɥɛ] will occur in only 6% of the cases, just as forms pronounced [ynos] will. The grammaticality of [byljauʁt] and [vyswɛstɛʁn], on the other hand, will have to be accounted for by the fact that the underlying forms are not vowel-initial. If we assume that the surface structures of these forms contain glides, the cue constraints in tableaux (48) through (50) are no longer valid, because these constraints refer to vowels at SF. Nevertheless, the data can be accounted for by assuming that most cue constraints are ranked equally high for glides as for vowels, e.g. the constraint \*[CʔG]/CʔG/ is ranked equally high as \*[CʔV]/CʔV/, where "G" stands for a glide. The only constraint for which we have to assume a different ranking is \*[CG]/CʔG/. This has to be ranked lower than \*[CV]/CʔV/, e.g. at 91.5 instead of at 98.0. This lower ranking is natural, especially in the syllable-boundary view of *h*-aspire: it is cross-linguistically (and probably also in French) much more likely that the auditory form [swɛ] contains a syllable boundary than that the auditory form [sɛ] contains a syllable boundary; hence, the listener's standpoint requires that \*[CV]/C.V/ >> \*[CG]/C.G/.

With the ranking just proposed (and \*[ə] still ranked at 80.0), the production percentages are the following. In a tableau analogous to (48), the candidate [kɛlwɛstɛʁn] /kɛlʔwɛstɛʁn/ would win in 81% of the cases (cf. 4% in (51) for [kɛlazazɛ] /kɛlʔazazɛ/), so that pauses before glides are much less likely than before vowels. In a tableau analogous to (49), the candidate [səʔwɛstɛʁn] /səʔwɛstɛʁn/ would be reduced to 15% (cf. 34% in (51) before vowels), the candidate [səwɛstɛʁn] /səʔwɛstɛʁn/ would be reduced to 36% (cf. 64% before vowels), and the candidate [swɛstɛʁn] /swɛstɛʁn/ would be reduced to 0.7% (cf. 2%). The major candidate would become [swɛstɛʁn] /sʔwɛstɛʁn/, now 48% (cf. 0.4%). If "for some speakers" such forms are not ruled out separately by a high-ranked structural constraint (perhaps \*/.Cʔ/ or \*/.C./), they can now surface, and both pauses and schwas are less likely before glides (15% and 51%) than before vowels (34% and 98%). In a tableau analogous to (50), finally, the candidate [vysəʔwɛstɛʁn] /vysəʔwɛstɛʁn/ would win in 13% of the cases (cf. 28%), the candidate [vysʔwɛstɛʁn] /vysʔwɛstɛʁn/ in 4% of the cases (cf. 8%), the candidate [vysəwɛstɛʁn] /vysəʔwɛstɛʁn/ in 35% of the cases (cf. 61%), and the candidate [vyswɛstɛʁn] /vyswɛstɛʁn/ in 0.6% of the cases (cf. 2%). The big winner in this case is [vyswɛstɛʁn] /vysʔwɛstɛʁn/, which wins in 48% of the cases (cf. 0.4%). Again, pauses are less likely before glides than before vowels (17% versus 36%), and so are schwas (48% versus 89%).

By making the ranking of one single cue constraint dependent in a cross-linguistically plausible way on whether it references a vowel or a glide, we have successfully modelled existing observations on the reduction of the number of schwas and pauses in [kɛlwɛstɛʁn], [swɛstɛʁn], and [vyswɛstɛʁn].<sup>34</sup>

<sup>34</sup> The reader may object that the schwas have been reduced more thoroughly than the pauses. One could argue, for instance, that a creaky pause will *never* be pronounced in [vys(\*ʔ)wɛstɛʁn]. If so, then the likely cause of this is the same as the likely cause of the asymmetry between the cross-linguistic abundance of phonetic glottal stop before utterance-initial vowels and the cross-linguistic rarity of phonetic glottal stop before utterance-initial glides, namely that *articulatory* glottal stops are much more audible before vowels than before glides. To model this asymmetry in the relation between Articulatory Form and Auditory Form, we would have to take seriously the sensorimotor constraints in (47). This is outside the scope of the present article, in which Articulatory Form and Auditory Form have simplifyingly be collapsed into a single phonetic form.

### 6.5. Stability as a result of four learning algorithms

It is of some concern how a ranking like that in (50) can be learned. I propose that four learning algorithms help to keep it stable. This section discusses them briefly.

The first algorithm to consider is lexicon-driven learning of perception (section 4.3; Boersma, 1997). If a child's ranking of \*/VV/ in (35) and (50) is too low, she will perceive an incoming [ynəos]<sub>Aud</sub> as /ynəos/. If she subsequently accesses the correct underlying form |ynə#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 (50), where we can also see that if comprehension were just as parallel as production, the constraint MAX(?) would have to rise as well. There is one problem with the ranking of \*/VV/ in this account: in the French language environment the auditory form [VV] is much more likely to represent an underlying |V(#)V| than an underlying |V(#)ʔV|, which should lead to a low ranking of \*/VV/ according to lexicon-driven learning of perception; this problem is addressed in section 8.1.

The second algorithm is meaning-driven learning of recognition. Semantic levels of representation above UF, such as the meaning of a lexical item and the meaning of the sentence (e.g. Logical Form), can correct errors in the SF-to-UF mapping. This can be applied to minimal pairs (Boersma, 2001), which are of little relevance to the present paper, and to establishing the underlying forms themselves (Escudero, 2005; Apoussidou, 2006), which is relevant to the question of whether *h*-aspiré is better represented underlyingly as a glottal stop or as a syllable boundary.

The third algorithm is learning by virtual production (section 5.3; Boersma, 1998). If the perception system works well, [ynəos]<sub>Aud</sub> will be perceived as /ynəʔos/ and lead to accessing |ynə#ʔos| in the lexicon. If the child's ranking of MAX(?) is too low, tableau (50) makes her compute from this UF the virtual production [ynos] /ynos/, independently of where \*[ə] and \*[ʔ] are ranked.<sup>35</sup> The comparison of a 'correct' [ynəos]/ynəʔos/ and an 'incorrect' [ynos]/ynos/ will lead to a rise of MAX(?) and \*[CV]/CV/ and to a fall of \*[ə] and \*[VV]/VʔV/, as can be seen from (50).

The fourth algorithm has to be learning by self-perception, the mirror-image of learning by virtual production. If MAX(?), \*[ə] and \*[ʔ] are ranked too high, the winning form in production, given the underlying form |ynə#ʔos|, will be [ynos] /ynəʔos/, as can be seen from (50). If the child's ranking of the cue constraints is adultlike, she will subsequently perceive her own [ynos] as /ynos/. The comparison of a 'correct' [ynos] /ynos/ and an 'incorrect' [ynos] /ynəʔos/ will lead to a rise of \*[CV]/CʔV/ and to a fall of MAX(?) and \*[CV]/CV/, as can be seen from (50).

Technically, these four learning algorithms can probably be subsumed under a scheme where the 'correct' intermediate representations are based on an optimization where two more peripheral representations (e.g. auditory form and meaning) are kept fixed, and two possibly 'incorrect' candidates are computed by keeping only one of the peripheral representations fixed (Boersma, 2006). In any case, the learning algorithms seem to lead together to a situation in which the cue constraints are ranked optimally for perception and MAX(?) is ranked neither too low nor too high. A solid proof that this optimization really works would involve a whole-language computer simulation that is far beyond the scope of the present paper.

<sup>35</sup> 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 et al. (2004).

### 6.6. Assessment of the three listener-oriented models

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  $*[\ə] \gg *[\ʔ]$ . If one multiplies all the ranking values in (50) by 100, the form [ynəos] /ynəʔos/ will always win.<sup>36</sup> 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  $*[\ʔ] \gg *[\ə]$  in (28).

The parallel model also seems to be *less* listener-oriented than the stochastic control-loop model, because of the possibility of hallucinations (a speaker's incorrect views of what the listener will perceive), as explained in section 6.1.1.

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. The parallel model also has a theoretical advantage, in that it is not *explicitly* listener-oriented: it just works by using the same constraints in production as in perception, and any listener-oriented properties turn up as automatic side effects.

## 7. A listener-oriented account in terms of syllable boundaries

As suggested in section 2.2, Cornulier's (1981) syllable-boundary account can be translated into a listener-oriented OT version with only slight modifications. I will first explain how this can be done, then discuss some advantages and disadvantages of the syllable-boundary account as compared to the glottal-stop account.

### 7.1. Cornulier's syllable-boundary account in Optimality Theory

In order to avoid the word-specific alignment constraints of Cornulier (1981) and Tranel and Del Gobbo (2002) or the word-specific constraint ranking of Tranel (1996), we can express *h*-aspiré as an underlying syllable boundary (Schane, 1978), which according to Cornulier (1981:210,227) is an equivalent solution.

To rule out candidates like /y.nos/, /la.zaʁ/, /le.za.zaʁ/, and /ke.le.ʁo/, we can use a CONTIGUITY constraint (McCarthy and Prince, 1995) for syllable boundaries, namely CONTIG(.V) or, more explicitly,  $*|_{-1}V_2|/.1XV_2/$ , which states that an adjacent sequence of syllable boundary plus vowel at UF should correspond at SF to a sequence of syllable boundary plus vowel without any intervening material. This is the alignment-as-faithfulness solution by Boersma (1998:196–199) and Horwood (2002). It replaces Cornulier's word-specific alignment constraint ALIGN-L (/os/, Syllable) and the segmental faithfulness constraint MAX(?) earlier in the present paper.

As far as hiatus is concerned, both auditory forms with underlying syllable boundaries like [leazaʁ] and auditory forms without underlying syllable boundaries like [agreabl] have to be interpreted as having a syllable boundary within the vowel sequence, i.e. as /le.a.zaʁ/

<sup>36</sup> It may look problematic to propose that the rankings in (50) 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 section 4.4. An alternative possibility worth pursuing is that the evaluation noise could be smaller in production than in perception.



and /a.gre.abl/.<sup>37</sup> This can be achieved with a constraint at SF, namely Cornulier’s Postulate IV “every syllable contains exactly one vowel”, which has already been seen to be capable of ruling out forms like /l.a.zaʁ./ (section 2.2). This constraint can be formulated in SPE-style as /σV<sub>1</sub>/ and replaces \*/VV/ (section 4.3).

As argued in section 2.2, it is crucial that Cornulier’s cue constraint (“an auditory CV sequence without a pause does not contain a syllable boundary”) has to be included in the story. Because even for Cornulier this constraint relates two different representations (phonetics and syllable structure), the correct way of including it in tableaux is to have a pair of phonetic and phonological representations in every candidate cell of the tableau, similarly to tableaux (48)–(50). Cornulier’s cue constraint can then be formulated as \*[CV]/C.V/. This replaces \*[CV]/CʔV/ (section 4.4).

To rule out forms like /kɛ.lə.a.zaʁ./, we need DEP(ə) or, more explicitly, \*|/ə/, which is identical to the same constraint in section 2.1 and reflects Cornulier’s “droit d’e”. In order to prevent forms like [ynəfam] from being optimal, any non-consummation of the “droit d’e” has to be enforced by an additional constraint like \*ə.

The four undominated OT constraints just mentioned, each of which directly expresses one of Cornulier’s inviolable constraints and each of which corresponds to a single segmental constraint in section 2.1 through section 4.4, can take care of almost all forms. In (52) and (53) we see that the underlying form |kɛl#.azaʁ| surfaces with an audible syllable boundary as [kɛl<sub>2</sub>ʔazaʁ] /kɛl.a.zaʁ./, and the underlying form |ynə#.os| surfaces either with a schwa as [ynəos] /y.nə.os./ or with an audible syllable boundary as [yn<sub>2</sub>ʔos] /yn.os./, just as in Cornulier’s paper.

(52) Cornulier in OT: enchainment

kɛl#.azaʁ	CONTIG(.V)	/σV <sub>1</sub> /	*[CV]/C.V/	DEP(ə)	* <sub>2</sub>	*ə
√ <sub>☞</sub> [kɛl <sub>2</sub> ʔazaʁ] /kɛl.a.zaʁ./					*	
[kɛləazaʁ] /kɛ.lə.a.zaʁ./				*!		*
[kɛlazazʁ] /kɛ.la.zaʁ./	*!					
[kɛlazazʁ] /kɛl.a.zaʁ./			*!			

(53) Cornulier in OT: une hausse

ynə#.os	CONTIG(.V)	/σV <sub>1</sub> /	*[CV]/C.V/	DEP(ə)	* <sub>2</sub>	*ə
[ynəʔos] /y.nə.os./					*(!)	*(!)
√ <sub>☞</sub> [ynəos] /y.nə.os./						*
[ynəos] /y.nə.os./	*!	*				
√ <sub>☞</sub> [yn <sub>2</sub> ʔos] /yn.os./					*	
[ynos] /y.nos./	*!					
[ynos] /yn.os./			*!			

To see why this account works, consider the cue constraint \*[CV]/C.V/, which is based on Cornulier’s observation that syllable boundaries have to be audible, i.e. if there is no pause or

<sup>37</sup> The present paper does not attempt to address what happens at the end of this last word.

other break within a phonetic [CV], the phonological structure cannot be /C.V/. Crucially, this cue constraint cannot be replaced with Tranel and Del Gobbo's (2002) conjoined structural constraint ONSET&NoCODA, i.e. \*/C.V/. If tableaux (52) and (53) used \*/C.V/ instead of \*/[CV]/C.V/, then the first candidate in (52) and the fourth candidate in (53) would be incorrectly ruled out. Unlike Tranel and Del Gobbo, Cornulier argued that it is all right to have the onsetless coda structure /C.V/, as long as the syllable boundary is audible in the phonetic form. This link between the phonology and the phonetics gives the correct results, and is therefore crucial here.

A full-fledged syllable-based account would require more than the constraints in (52) and (53). In order to rule out /kɛ.a.zaʁ/, which would satisfy all four high-ranked constraints in (52), the constraint MAX(C) would be needed again. Likewise, we would need MAX(V) and the constraints that handle liaison, and we would again need the cue constraint \*/[C̥V]/CV/ to rule out [yn̥ɔs] /y.nos/. And in order to make sure that the phonetically enhanced form [yn̥ɔs] /y.n̥ɔs/, which is harmonically bounded in (53), could sometimes win, we would need to add a ranking like { \*[VV]/V.V/, \*[C̥V]/C.V/ } >> \*[V̥V]/V.V/, analogous to { \*[VV]/V̥V/, \*[C̥V]/C̥V/ } >> \*[V̥V]/V̥V/ in (50), expressing the idea that a creaky pause is a better cue for a syllable boundary than a naked hiatus or a postconsonantal pause are. Ultimately, the full syllable-based analysis would be very close to the one in tableaux (48) through (50).

## 7.2. The underlying representation of *h*-aspire

Is *h*-aspire better represented as an underlying glottal stop or as an underlying syllable boundary? With the account of *h*-aspire as an underlying syllable boundary (section 7.1), the explanation for the surfacing of schwa in *une hausse* is 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 empirical advantage of the syllable-island approach would be that it could immediately explain the case of initial neutralization (section 1.1), since every phrase boundary must necessarily be a syllable boundary. According to Cornulier (1981:183), it may also account for four occurrence restrictions, namely that *h*-aspire cannot occur before a consonant or morpheme-finally, that it cannot be followed by schwa, and that it cannot be geminated; however, these four points could also be adduced *in favour* of an analysis as [ʔ] or [h], as the same four restrictions apply to laryngeal consonants (both [ʔ] and [h]) in nearly all Germanic languages, where because of their contrastivity only one of the two (usually [ʔ]) can be dismissed as a mere juncture phenomenon. Finally, it would also account for other phenomena in which *h*-aspire words act as if they start with a vowel, such as 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*-aspire is realized as a glottal stop but not as a syllable boundary; as a case in point, Meisenburg and Gabriel (2004) mention “tʁwaʔɑ̃bœʁgœʁ” for [tʁwaz#ʔɑ̃bœʁgœʁ] ‘three hamburgers’, in which the glottal stop ends up in the middle of a monosyllabic diphthong. Another problem is that, ironically, Cornulier’s distinction between |.ɛʁ| and |.azaʁ| is easier to represent as [ʔ<sub>1</sub>] versus [ʔ<sub>2</sub>] than as |.<sub>1</sub>| versus |.<sub>2</sub>|, since only the former contrast can be expressed in terms of substantive phonological features.

The biggest difference between the segmental and the syllable-boundary approach is their handling of non-*h*-aspire hiatus. As mentioned in section 1.5, the creaky pause in hiatus is typical of *h*-aspire (and hesitations and other breaks) but not of intervocalic syllable boundaries in

general. Section 8.2 shows that even in the case of our parallel model of phonology and phonetics, the specificity of the creaky pause is an argument in favour of the segmental approach.

### 8. Modelling hallucination in production

As shown in section 6, the bidirectional model of phonology and phonetics predicts seemingly listener-oriented effects in production. However, we have also seen a few cases where the speaker’s produced phonological surface structure is different from the phonological surface structure that the listener will recover from the speaker’s production. In these cases, therefore, the speaker can be said to *hallucinate* that the listener will be able to recover a certain surface structure. Less dramatically, these cases could be called cases of *non-recoverability* in production. In the serial production model of section 5.1, these would be cases of neutralization in phonetic implementation. In the parallel production model of section 6.1, these cases are forced by high-ranked faithfulness constraints. The present section shows that allowing a case of hallucination in production solves one of the remaining problems of *h*-aspire.

#### 8.1. The creaky pause: specific to *h*-aspire words?

In this section I try to account for the observation that the intervocalic creaky pause occurs before *h*-aspire words but not in the more common case of vowel–vowel hiatus (see the discussions in sections 1.5 and 7.2). There is a question mark in the title of this section because the constraint ranking behind (51) predicts that French speakers use the creaky pause for enhancing any kind of hiatus, not just *h*-aspire. This is shown in tableaux (54) and (55), where the constraint DEP(?) has been added at the same height as MAX(?).

(54) Variation between creaky pause and hiatus

		110.0	96.5	96.5	92.5	92.0	91.0	
	lez#?azaz	*/VV/	MAX (?)	DEP (?)	*[ʔ]	*/VʔV/ \\ / [VV]	*/VʔV/     [VʔV]	
☞	[leʔazaz] /leʔazaz/				*		*	35%
	[leʔazaz] /leazaz/	*!	*		*			
☞	[leazaz] /leʔazaz/					*		65%
	[leazaz] /leazaz/	*!	*					

(55) Variation between creaky pause and hiatus

		110.0	96.5	96.5	92.5	92.0	91.0	
	agʁeabl	*/VV/	MAX (?)	DEP (?)	*[ʔ]	*/VʔV/ \\ / [VV]	*/VʔV/     [VʔV]	
☞	[agʁeʔabl] /agʁeʔabl/			*	*		*	35%
	[agʁeʔabl] /agʁeabl/	*!			*			
☞	[agʁeabl] /agʁeʔabl/			*		*		65%
	[agʁeabl] /agʁeabl/	*!						

Although the violation patterns are different, (54) and (55) lead to the same degree of variation, because of the high ranking of \*/VV/. They predict that |leʔ#ʔazaʔ| will be pronounced as [leazaʔ] in 65% of the cases and as the creaky-pause-enhanced [leʔazaʔ] in 35% of the cases. But they also predict the same enhancement in cases of underlying non-*h*-aspire hiatus: |agʔeabl| will be pronounced as [agʔeabl] in 65% of the cases and as the creaky-pause-enhanced [agʔeʔabl] in 35% of the cases.

It is possible to change the constraint ranking in such a way that the two hiatuses in |leʔ#ʔazaʔ| and |agʔeabl| are treated differently. The trick is to move \*/VV/ to the bottom of the hierarchy, i.e. to freely allow sequences of vowels in the phonological surface structure.

The first result of moving \*/VV/ to the bottom is that this constraint does not pose any restrictions on the perception of hiatus any longer. The listener is now free to perceive a phonetic sequence of vowels as a phonological sequence of vowels. This is shown in (56) and (57), which, other than (41), now contain feasible candidates with a phonological hiatus (i.e. /leazaʔ/). With the loss of the general anti-hiatus constraint \*/VV/ we now need to take seriously the auditory cues that handle the perception of /VV/, so we need at least the two cue constraints \*[VʔV]/VV/ and \*[VV]/VV/, which have been included in (56) and (57), ranked according to the principles of section 4.4.

(56) *The perception of hiatus*

	94.0	92.0	91.0	90.0	30.0	
[leazaʔ]	*/VV/ / \\ [VʔV]	*/VʔV/ \ /\ [VV]	*/VʔV/     \\\ [VʔV]	*/VV/    \\\ [VV]	*/VV/	
ʔ /leʔazaʔ/		*				24%
/leazaʔ/				*	*	76%

(57) *The perception of the creaky pause*

	94.0	92.0	91.0	90.0	30.0	
[leʔazaʔ]	*/VV/ / \\ [VʔV]	*/VʔV/ \ /\ [VV]	*/VʔV/     \\\ [VʔV]	*/VV/    \\\ [VV]	*/VV/	
/leʔazaʔ/			*			86%
(ʔ) /leazaʔ/	*			*	*	14%

With \*/VV/ out of the way, the auditory form [leazaʔ] will now be perceived as the naked hiatus /leazaʔ/ most of the time, while [leʔazaʔ] will still mostly be perceived as /leʔazaʔ/. The new possibility of perceiving /leazaʔ/ in (56) contradicts my assertions and formalizations in sections 3.1, 3.3, 4.4 and 7.1, but heeds the native-speaker objection in footnote 18. Most importantly, the low ranking of \*/VV/ is compatible with the observation made in section 6.5 that in a French language environment the auditory form [VV] much more often represents an underlying |V(#)| than an underlying |V(#)ʔV|.

The next question is how the removal of \*/VV/ influences production. Tableau (58) shows that the percentages of [leʔazaʔ] and [leazaʔ] pronunciations closely mirror those in (54).

(58) Underlying glottal stop: hallucinated glottal stop

	96.5	96.5	94.0	92.5	92.0	91.0	90.0	
$ \underline{le}z\#?azaz\bar{z} $	MAX (?)	DEP (?)	*/VV/ / \ \ [V <sub>2</sub> V]	*[ʔ]	*/VʔV/ \ / \ [VV]	*/VʔV/       \ [V <sub>2</sub> V]	*/VV/     \ [VV]	
☞ [leʔazaz̩] /leʔazaz̩/				*		*		34%
[leʔazaz̩] /leazaz̩/	*		*	*				
☞ [leazaz̩] /leʔazaz̩/					*			64%
[leazaz̩] /leazaz̩/	*						*	2%

The result is 64% unenhanced hallucinatory [leazaz̩] /leʔazaz̩/ and 34% enhanced [leʔazaz̩] /leʔazaz̩/. I speak of a hallucination here because the most common winning candidate in the speaker’s production is [leazaz̩] /leʔazaz̩/ (i.e. the speaker ‘thinks’ the listener will perceive /leʔazaz̩/), although the speaker herself will perceive an incoming [leazaz̩] most frequently as /leazaz̩/, not as /leʔazaz̩/.

Tableaux analogous to (58) can be made for all the other forms. Their percentages are very close to those in (51).

The removal of \*/VV/ to the bottom of the constraint hierarchy has a much larger influence on forms without an underlying glottal stop. This is because MAX(?) can now no longer suppress the forms with a phonological vowel sequence. This is illustrated in (59).

(59) Underlying vowel sequence: naked hiatus

	96.5	96.5	94.0	92.5	92.0	91.0	90.0	
$ agʁeabl $	MAX (?)	DEP (?)	*/VV/ / \ \ [V <sub>2</sub> V]	*[ʔ]	*/VʔV/ \ / \ [VV]	*/VʔV/       \ [V <sub>2</sub> V]	*/VV/     \ [VV]	
[agʁeʔabl] /agʁeʔabl/		*		*		*		0.2%
[agʁeʔabl] /agʁeabl/			*	*				4%
[agʁeabl] /agʁeʔabl/		*			*			0.5%
☞ [agʁeabl] /agʁeabl/							*	95%

The result is 95% unenhanced non-hallucinatory [agʁeabl] /agʁeabl/. In terms of actual pronunciation, the auditorily enhanced [leʔazaz̩] occurred in 34% of the cases, the auditorily enhanced [agʁeʔabl] only in 4.2% of the cases. This difference fulfils my wish to devise an account that restricts the creaky pause to cases of underlying *h*-aspire.

Ironically, the removal of \*/VV/ disposes of a constraint that was introduced in section 4.3 as the constraint that drove the whole listener-oriented account; the cause why it can now be disposed of is that the parallel phonological–phonetic framework has speaker-based hallucinatory aspects as well. The reason we would want to dispose of it is that a low \*/VV/ is predicted by the gradual lexicon-driven learning of perception (section 6.5). A low \*/VV/ is also compatible with native-speaker intuitions that say that an auditory hiatus is not usually perceived as a glottal stop.

## 8.2. The creaky pause: specific to the glottal-stop view of *h*-aspiré?

The next question to answer (section 7.2) is whether the difference in creaky pauses between [leʔazax̥] and [aɔʁeabl̥] is possible in the syllable-boundary approach. The answer is that it is possible, but unlikely to be correct. The first thing to note is that with the constraints in (53) and below, the syllable-boundary approach leads to the same results as tableaux (54) and (55). The two views of *h*-aspiré do not differ in this respect. The next step is to move  $*/\sigma V_1^1/$  to the bottom of the hierarchy, thereby allowing multiple vowels within a syllable. The result is 64% [leazax̥] / .le.a.zaɔx̥./, 34% [leʔazax̥] / .le.a.zaɔx̥./, and 83% [.a.ɔʁeabl̥.] / .a.ɔʁeabl̥./; in the case of [leazax̥], it is this time CONTIG(V) that forces the speaker to hallucinate that the underlying syllable boundary has surfaced. The disadvantage of this approach, however, is that it predicts that Grammont's (1948) auditory vowel sequence [eayœ] mentioned in section 1.5 constitutes a single syllable, something unheard of in the literature on French phonology.

Although the case of enhancing creak thus seems to favour the glottal-stop view of *h*-aspiré, the issue has by no means been settled, because I may have missed a possible ranking or constraint. In the end, a whole-language computer simulation of the acquisition of the phonology of French may provide a partial answer to the question of whether *h*-aspiré is an underlying segment or an underlying syllable boundary, or whether French learners are just as unsure about this as linguists are, and that they therefore come to vary in their underlying representations. Such a computer simulation could be performed with the doubly-hidden learning algorithm used by Apoussidou (2006), i.e. with learners who are given pairs of auditory form and meaning and who have to construct the two intermediate representations (SF and UF) from those.

## 9. A possible complete ranking

Throughout this article I have proposed many partial rankings, and many modifications: after the categorical (non-varying) ranking in (34), a set of more finely-grained cue constraints was introduced in section 4.4; then  $*/[\partial]/$  was moved down a bit in section 6.2, and DEP( $\partial$ ) in section 6.3; cue constraints for glides were introduced in section 6.4; and the structural constraint  $*/VV/$  was moved down (or away) in section 8.1. A possible complete ranking of all these constraints is shown in (60), where a ranking difference of 10.0 or more represents categorical ranking.

### (60) Final best constraint ranking

Constraints	Ranking
MAX(C), DEP(C), MAX(V), DEP(V)	110.0
DEP( $\partial$ )	99.0
$*/[CV]/C^?V/$	98.0
$*/[V^?V]/V/$	97.0
MAX(?), DEP(?)	96.5
$*/[VV]/V/$	96.0
$*/[C^?V]/CV/, */[C^?G]/CG/$	95.0
$*/[V^?V]/VV/, */[V^?G]/VG/$	94.0
$*/[C^?V]/C^?V/, */[C^?G]/C^?G/$	93.0
$*/[\text{?}]/$	92.5
$*/[VV]/V^?V/, */[VG]/V^?G/$	92.0
$*/[CG]/C^?G/$	91.5
$*/[V^?V]/V^?V/, */[V^?G]/V^?G/$	91.0

*[VV]/VV/, *[VG]/VG/	90.0
*[ə]	80.0
*/CC/, MAX(V)	75.0
MAX(C)	65.0
*[CV]/CV/, *[CG]/CG/	55.0
(*VV/)	(30.0)

This ranking handles all the forms discussed in the present paper: 100% [ləgaks̥] /ləgaks̥/, 64% hallucinatory [ləzak] /ləzak/ (34% enhanced [ləzak] /ləzak/, 2% elided [lazak] /lazak/), 100% elided [lɔm] /lɔm/, 100% [kɛlgaks̥] /kɛlgaks̥/, 80% non-enchained [kɛlʔazak] /kɛlʔazak/ (15% enchained [kɛlazak] /kɛlazak/, 4% hallucinatory [kɛlazak] /kɛlʔazak/, 1.5% schwa-inserted [kɛləzak] /kɛləzak/), 100% [kɛlɔm] /kɛlɔm/, 100% dropped floaters in [legaks̥] /legaks̥/, 64% hallucinatory [lezak] /lezak/ (34% enhanced [ləzak] /lezak/, 2% non-hallucinatory [lezak] /lezak/), 100% liaison in [lezɔm] /lezɔm/, 96% schwa drop in [ɲnfam] /ɲnfam/ (4% [ɲnɛfam] /ɲnɛfam/), 61% hallucinatory [ɲnɔs] /ɲnɔs/ (28% enhanced [ɲnɔs] /ɲnɔs/, 8% non-enchained schwa-dropped [ɲnɔs] /ɲnɔs/, 2% elided [ɲnos] /ɲnos/), 95% [agvɛabl] /agvɛabl/ (4% [agvɛabl] /agvɛabl/, 0.5% [agvɛabl] /agvɛabl/, 0.2% [agvɛabl] /agvɛabl/).

The reason for the low ranking of \*[CV]/CV/ can now be seen: it is needed to disallow avoidance of [lɔm] /lɔm/, [kɛlɔm] /kɛlɔm/, and [lezɔm] /lezɔm/; if ranked at 90.0, it would lead to such monstrosities as 49% [lɔɔm] /lɔɔm/, 2% [kɛlʔɔm] /kɛlɔm/, and 49% [leɔm] /leɔm/.

It may seem that the number of constraints is large, but the formulation of the constraints is straightforward. No parochial or concocted constraints, such as those in sections 2.3 and 2.4, are needed any longer. Using a large number of simple constraints instead of a small number of complex constraints seems to require a massively parallel processing system. The human cortex seems to be a good candidate.

## 10. Conclusion

This paper was about the observation that in the French form *une hausse* the first *e* tends to be pronounced, a fact for which no satisfactory explanation had been given yet. All speaker-based accounts of *h*-aspiré (sections 1.6, 2.3 and 2.4) are observationally more or less adequate, but to account for the schwa in *une hausse* they require a special rule or constraint that does not work in accounting for any of the other forms in the language. Descriptive adequacy has been improved by introducing an Optimality-Theoretic listener-oriented account, in which the surfacing of the schwa in *une hausse* is predicted on the basis of the same constraint set that is crucial to handling the other forms in the language (sections 3.1, 3.2, 4.4, 5.3, 5.4 and 6.1.3). The surfacing schwa ultimately falls out naturally as a means to improve the perceptibility of an underlying glottal stop or syllable boundary. Explanatory adequacy has partly been achieved by discussing the acquisition of the constraint ranking in perception (section 4.3) and in production (section 6.5), which supplies the non-goal-oriented automatic mechanism that underlies what I teleologically called ‘improving perceptibility’ in the previous sentence. It is simply on the basis of non-goal-oriented learning mechanisms that the ranking in (60) has been established. The cause why it works is that the same constraints are used in production as in comprehension.

The present paper does not provide the final answer to the question of *h*-aspiré. A question remaining to be answered is the question of the underlying representation of *h*-aspiré as a syllable boundary or a segment (section 7.2), although the observations about the enhancing creaky pause seem to favour the segmental approach (section 8.2).

The remaining question is which of the listener-oriented grammar models is correct. All three can handle most of the observed facts of *h*-aspire. 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 *bidirectional phonology and phonetics* where in the production direction *phonology and phonetics are evaluated in parallel*, looks most promising. In this paper it turned out to be the only one that can account for both enhancement and hallucination, leading to a correct treatment of the naked hiatus cases (section 8). By using the same constraints in production as in perception, this model seems to be able to handle several more production phenomena such as incomplete neutralization, licensing by cue, counterbleeding opacity, the evolution of auditory contrast (Boersma and Hamann, 2007), and the emergence of universal faithfulness rankings (Boersma, 2006) in ways that require fewer stipulations than former grammar models.

## References

- Apoussidou, D., 2006. On-line learning of underlying forms. Rutgers Optimality Archive 835.
- Apoussidou, D., Boersma, P., 2004. Comparing two Optimality-Theoretic learning algorithms for Latin stress. WCCFL 23, 29–42.
- Bagemihl, B., 1995. Language games and related areas. In: Goldsmith, J.A. (Ed.), *The Handbook of Phonological Theory*. Blackwell, Cambridge, MA, Oxford, pp. 697–712.
- Bally, C., 1944. *Linguistique générale et linguistique française*, second ed. Francke, Berne.
- Best, C.T., 1995. A direct realist view of cross-language speech perception. In: Strange, W. (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. York Press, Baltimore, pp. 171–203.
- Blevins, J., Garrett, A., 1998. The origins of consonant-vowel metathesis. *Language* 74, 508–556.
- Boersma, P., 1989. Modelling the distribution of consonant inventories by taking a functionalist approach to sound change. *Proceedings of the Institute of Phonetic Sciences, Amsterdam* 13, 107–123.
- Boersma, P., 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam* 21, 43–58.
- Boersma, P., 1998. *Functional phonology*. Ph.D. thesis, University of Amsterdam.
- Boersma, P., 2000. The OCP in the perception grammar. Rutgers Optimality Archive 435.
- Boersma, P., 2001. Phonology–semantics interaction in OT, and its acquisition. In: Kirchner, R., Wikeley, W., Pater, J. (Eds.), *Papers in Experimental and Theoretical Linguistics*, vol. 6. University of Alberta, Edmonton, pp. 24–35 [Rutgers Optimality Archive 369, 1999].
- Boersma, P., 2003a. The odds of eternal optimization in Optimality Theory. In: Holt, D.E. (Ed.), *Optimality Theory and Language Change*. Kluwer, Dordrecht, pp. 31–65.
- Boersma, P., 2003b. Overt forms and the control of comprehension. In: Spenader, J., Eriksson, A., Dahl, Ö. (Eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. Department of Linguistics, Stockholm University, pp. 47–56 [Rutgers Optimality Archive 597].
- Boersma, P., 2006. The acquisition and evolution of faithfulness rankings. Talk presented at the 14th Manchester Phonology Meeting, May 28.
- Boersma, P., Escudero, P., 2004. Learning to perceive a smaller L2 vowel inventory: an Optimality Theory account. Rutgers Optimality Archive 684 [to appear in a book edited by E. Dresher, P. Avery, K. Rice].
- Boersma, P., Hamann, S., 2005. The violability of backness in retroflex consonants. Rutgers Optimality Archive 713.
- Boersma, P., Hamann, S., 2007. The evolution of auditory contrast. Rutgers Optimality Archive 909.
- Boersma, P., Hayes, B., 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32, 45–86.
- Chao, Y.-r., 1934. The non-uniqueness of phonemic solutions to phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica IV* (part 4), 363–397 [Reprinted in: M. Joos (Ed.), 1958, *Readings in Linguistics: The Development of Descriptive Linguistics in America since 1925*, second ed. American Council of Learned Societies: New York, pp. 38–54].
- Chomsky, N., Halle, M., 1968. *The Sound Pattern of English*. Harper and Row, New York.
- Clements, G.N., 1988. The role of the sonority cycle in core syllabification. Working Papers of the Cornell Phonetics Laboratory 2, pp. 1–68 [not seen].
- Clements, G.N., Keyser, S.J., 1983. *CV Phonology: A Generative Theory of the Syllable*. MIT Press, Cambridge, MA.



- de Cornulier, B., 1981. H-aspirée et la syllabation: expressions disjonctives. In: Goyvaerts, D.L. (Ed.), *Phonology in the 1980s*. Story-Scientia, Ghent, pp. 183–230 [Many references in the literature are to a version entitled: Expressions disjonctives: h et la syllababilité, handout workshop Méthodes en Grammaire Française, Paris, 6–8 December 1974].
- Damourette, J., Pichon, E., 1911–1952. *Des mots à la pensée: essai de grammaire de la langue française*, 8 volumes. d'Artrey, Paris.
- Davidson, L., Smolensky, P., Jusczyk, P., 2004. The initial and final states: theoretical implications and experimental explorations of Richness of the Base. In: Kager, R., Pater, J., Zonneveld, W. (Eds.), *Fixing Priorities: Constraints in Phonological Acquisition*. Cambridge University Press, Cambridge.
- Dell, F., 1973. *Les règles et les sons*. Hermann, Paris.
- Encrevé, P., 1988. *La liaison avec et sans enchaînement: phonologie tridimensionnelle et usages du français*. Seuil, Paris.
- Encrevé-Lambert, M.-H., 1971. A propos de l'élision en Français. *Rééducation Orthophonique* 60, 245–251 [not seen].
- Escudero, P., 2005. The attainment of optimal perception in second-language acquisition. Ph.D. thesis, University of Utrecht.
- Escudero, P., Boersma, P., 2003. Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In: S. Arunachalam, E. Kaiser, A. Williams (Eds.), *Proceedings of the 25th Annual Penn Linguistics Colloquium*. Penn Working Papers in Linguistics 8.1, pp. 71–85 [Rutgers Optimality Archive 439, 2001].
- Escudero, P., Boersma, P., 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26, 551–585.
- Flemming, E., 1995. *Auditory representations in phonology*. Ph.D. thesis, UCLA.
- Flemming, E., 2003. The relationship between coronal place and vowel backness. *Phonology* 20, 335–373.
- Fouché, P., 1959. *Traité de prononciation française*, second ed. Klincksieck, Paris [printing 1969].
- Fowler, C.A., 1986. An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* 14, 3–28.
- Freeman, M., 1975. Is French phonology abstract or just elsewhere: boundary phenomena and 'h aspiré' = [ʔ], not #?!. Unpublished ms., Harvard University [not seen].
- Fromkin, V., 1971. The non-anomalous nature of anomalous utterances. *Language* 47, 27–52.
- Gabriel, C., Meisenburg, T., 2005. Silent onsets? An Optimality-Theoretic approach to French h aspiré words. Poster presented at OCP 2, Tromsø, 20–22 January [Rutgers Optimality Archive 709].
- Grammont, M., 1948. *Traité pratique de prononciation française*. Delagrave, Paris.
- Hale, M., Reiss, C., 1998. Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry* 29, 656–683.
- Hall, R., 1948. *French*. Language monograph 24 (Structural sketch 1).
- Halle, M., 1959. *The Sound Pattern of Russian*. Mouton, The Hague.
- Hayes, B., 1999. Phonetically-driven phonology: the role of Optimality Theory and inductive grounding. In: Darnell, M., Moravcsik, E., Noonan, M., Newmeyer, F., Wheatley, K. (Eds.), *Functionalism and Formalism in Linguistics*, vol. I: General Papers. John Benjamins, Amsterdam, pp. 243–285 [Rutgers Optimality Archive 158, 1996].
- Horwood, G., 2002. Precedence faithfulness governs morpheme position. In: Mikkelsen, L., Potts, C. (Eds.), *Proceedings WCCFL 21*, Cascadilla, Somerville, pp. 166–179.
- Hyman, L., 1985. *A Theory of Phonological Weight*. Foris, Dordrecht.
- Jäger, G., 2003. Learning constraint sub-hierarchies: the Bidirectional Gradual Learning Algorithm. In: Zeevat, H., Blutner, R. (Eds.), *Optimality Theory and Pragmatics*. Palgrave Macmillan, Basingstoke, pp. 251–287.
- Jun, J., 1995. Place assimilation as the result of conflicting perceptual and articulatory constraints. *Proceedings of WCCFL 14*, 221–237.
- Kaye, J., Lowenstamm, J., 1984. De la syllababilité. In: Dell, F., Hirst, D., Vergnaud, J.-R. (Eds.), *Forme sonore du langage*. Hermann, Paris, pp. 123–155.
- Kirchner, R., 1998. *Lenition in phonetically-based Optimality Theory*. Ph.D. dissertation, UCLA.
- Lahiri, A., Marslen-Wilson, W., 1991. The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition* 38, 245–294.
- Lakoff, G., 1993. *Cognitive phonology*. In: Goldsmith, J. (Ed.), *The Last Phonological Rule: Reflections on Constraints and Derivations*. The University of Chicago Press, Chicago, London, pp. 117–145.
- Levelt, W., 1989. *Speaking: From Intention to Articulation*. MIT Press, Cambridge, MA.
- Malécot, A., 1975. The glottal stop in French. *Phonetica* 31, 51–63.
- Martinet, A., 1933. Remarques sur le système phonologique du français. *Bulletin de la Société de Linguistique de Paris*, vol. 34. Klincksieck, Paris pp. 191–202..
- Martinet, Ph., 1905. *Dictionnaire méthodique et pratique des rimes françaises*. Larousse [seen edition 1962].

- McCarthy, J.J., 2003. OT constraints are categorical. *Phonology* 20, 75–138.
- McCarthy, J., Prince, A., 1993. Generalized alignment. In: Booij, G., van Marle, J. (Eds.), *Yearbook of Morphology 1993*. Kluwer, Dordrecht, pp. 79–153.
- McCarthy, J., Prince, A., 1995. Faithfulness and reduplicative identity. In: J. Beckman, L.W. Dickey, S. Urbanczyk (Eds.), *Papers in Optimality Theory*. University of Massachusetts Occasional Papers 18. Graduate Linguistic Student Association, Amherst, MA, pp. 249–384.
- McQueen, J.M., Cutler, A., 1997. Cognitive processes in speech perception. In: Hardcastle, W.J., Laver, J. (Eds.), *The Handbook of Phonetic Sciences*. Blackwell, Oxford, pp. 566–585.
- Meisenburg, T., Christoph, G., 2004. Silent onsets? The case of French h aspiré words. Talk presented at the workshop *Phonetik und Phonologie 1*, Potsdam, 19 June.
- Pater, J., 2004. Bridging the gap between receptive and productive development with minimally violable constraints. In: Kager, R., Pater, J., Zonneveld, W. (Eds.), *Constraints in Phonological Acquisition*. Cambridge University Press, Cambridge, pp. 219–244.
- Prince, A., Smolensky, P., 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. Technical Report TR-2, Rutgers University Center for Cognitive Science.
- Prunet, J.-F., 1986. *Spreading and locality domains in phonology*. Ph.D. thesis, McGill.
- Schane, S., 1968. *French Phonology and Morphology*. MIT Press, Cambridge, MA.
- Schane, S., 1978. L'emploi des frontières de mots en Français. In: de Cornulier, B., Dell, F. (Eds.), *Études de phonologie française*. CNRS, Paris, pp. 133–147.
- Selkirk, E., Vergnaud, J.-R., 1973. How abstract is French phonology? *Foundations of Language* 10, 249–254.
- Steriade, D., 1995. Positional neutralization. Two chapters of an unfinished manuscript, UCLA.
- Tesar, B., 1997. An iterative strategy for learning metrical stress in Optimality Theory. In: Hughes, E., Hughes, M., Greenhill, A. (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development*, Cascadilla, Somerville, MA, pp. 615–626.
- Tesar, B., 1998. An iterative strategy for language learning. *Lingua* 104, 131–145.
- Tesar, B., 1999. Robust interpretive parsing in metrical stress theory. In: Shahin, K., Blake, S., Kim, E.-S. (Eds.), *Proceedings WCCFL 17, CSLI, Stanford, CA*, pp. 625–639.
- Tesar, B., Smolensky, P., 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29, 229–268.
- Tesar, B., Smolensky, P., 2000. *Learnability in Optimality Theory*. MIT Press, Cambridge, MA.
- Tranel, B., 1974. *The phonology of nasal vowels in Modern French*. Doctoral thesis, UC San Diego.
- Tranel, B., 1981. *Concreteness in Generative Phonology: Evidence from French*. University of California Press, Berkeley, Los Angeles, London.
- Tranel, B., 1987. *The Sounds of French: An Introduction*. Cambridge University Press, Cambridge.
- Tranel, B., 1995. Current issues in French phonology. In: Goldsmith, J.A. (Ed.), *The Handbook of Phonological Theory*. Blackwell, Cambridge, MA, Oxford, pp. 798–816.
- Tranel, B., 1996. French liaison and elision revisited: a unified account within Optimality Theory. In: Parodi, T., Quicoli, C., Saltarelli, M., Zubizarreta, M.L. (Eds.), *Aspects of Romance Linguistics*. Georgetown University Press, Washington, pp. 433–455 [Rutgers Optimality Archive 15, 1994].
- Tranel, B., Del Gobbo, F., 2002. Local conjunction in Italian and French phonology. In: Wiltshire, C.R., Camps, J. (Eds.), *Romance Phonology and Variation*. John Benjamins, Amsterdam, pp. 191–218.
- Wheeler, D.W., Touretzky, D.S., 1993. A connectionist implementation of cognitive phonology. In: Goldsmith, J. (Ed.), *The Last Phonological Rule: Reflections on Constraints and Derivations*. The University of Chicago Press, Chicago, London, pp. 146–172.