On the deletion of word-final schwa in Southern French*

Julien Eychenne
Hankuk University of Foreign Studies

To appear in *Phonology*

Abstract

Although a growing body of literature in formal phonology has espoused the view that phonological knowledge is gradient and probabilistic, this perspective remains somewhat controversial. This paper provides further empirical support to this strand of work: it offers an analysis of the gradient deletion of word-final schwa in Southern French, using a corpus containing 7,787 data points obtained from 45 subjects spread over three dialectal areas (the Basque Country, Languedoc and Provence). In addition to confirming or nuancing previous findings about the role of several phonological and non-phonological variables, this study demonstrates the influence of lexical frequency, grammatical category, sonority and the feature specification of the consonant before schwa.

Keywords – schwa, deletion, Southern French, variation, gradience

1 Introduction

The variability of schwa is one of the most well-known phenomena in the phonology of French and it has been investigated from many different theoretical perspectives (see among many others Selkirk 1978, Tranel 1981, Anderson 1982, Dell 1985, Côté 2000, van Oostendorp 2000, Eychenne 2006, Pustka 2007, Kaplan 2011, and references therein). Although many theoretical accounts of this phenomenon are based on idealised data that can often be traced back to early normative and descriptive works (see Morin 1987), a number of recent corpus-based studies have broadened our knowledge about the realisation of this vowel within and across varieties of French. (See in particular the results from the project *Phonologie du français contemporain*[1] (PFC), especially the contributions in Detey et al. 2016, Durand et al. 2009 and Gess et al. 2012; see also Bürki et al. 2011a,b and Bayles et al. 2016.) The main goal of this paper is to contribute to these recent

*I would like to thank Jacques Durand, Bernard Laks, Chantal Lyche, an associate editor at *Phonology*, and three anonymous reviewers for their helpful and constructive feedback on previous versions of this paper.

[1] `http://www.projet-pfc.net` (Durand et al. 2002).

efforts by clarifying the factors that influence the loss of word-final lexical schwa in a relatively under-studied variety, namely Southern French.

The behaviour of French schwa is extremely complex and displays a wide range of variation, as it is influenced by phonotactic, situational, social and dialectal factors, to name but a few. Since the realisation of schwa is one of the key characteristics of the phonology of Southern French, it is useful to first review the core properties of this vowel in other varieties, to which I will collectively refer as 'Northern French' for the sake of convenience. Although French 'schwa' historically derives from a bona fide central vowel (hence its name), it merged phonetically with the mid front rounded vowel /œ/ (or sometimes /ø/, depending on the variety) as early as the beginning of the 15th century, and possibly even earlier (Fouché 1969: 519). The crucial difference between schwa and stable /œ/ is thus phonological: schwa can alternate with zero, whereas stable /œ/ cannot. Consider the examples in (1), which represent typical realisations of the verb *relancer* /rəlɑ̃se/ 'to restart' in different phonological environments[2]:

(1) a. *Elle veut r⫽lancer la machine* [ɛlvørlɑ̃selamaʃin]
'She wants to restart the machine'

b. *Elles veulent r̲elancer la machine* [ɛlvœlrœlɑ̃selamaʃin]
'They want to restart the machine'

Schwa is usually not realised in (1a), since its omission does not raise any issue with respect to syllabification, whereas it is usually realised (as a surface [œ]) in (1b) because its absence would yield an illicit consonant cluster, namely *[lrl]. Some authors, such as Scheer (2011), consider that this alternation with zero is a necessary and sufficient condition to identify a phonological schwa, whereas others consider that orthography can also be a useful guide (see for instance Gess et al. 2012).

Schwa is generally found lexically in monosyllabic clitics[3] (e.g. *je* /ʒə/ 'I', *me* /mə/ 'me', etc.) and in initial or internal position of polysyllabic words (e.g. *venir* /vənir/ 'to come', *appeler* /apəle/ 'to call'). Its status in word-final position is still somewhat debated but, following Tranel (1981), most researchers regard it as epenthetic in this position. Indeed, schwa usually does not occur word-finally unless its presence is required by phonotactic constraints to prevent the occurrence of a complex cluster (e.g. *parc _ naturel* [parkœnatyrɛl] 'nature park'); its realisation can also be driven by prosody at the end of an intonational phrase (IP), a phenomenon known as *prepausal schwa* in the literature (e.g. *bonjour_!* [bɔ̃ʒurə] 'Hello!', see for instance Fagyal 2000, Biers 2017). In any case, as both examples show, the presence (or absence) of a schwa is not correlated with the presence of a word-final <e> (Durand & Eychenne 2004).

The behaviour of schwa in Southern French, which is markedly different from that of other varieties, has received comparatively little attention (see nonetheless Durand

[2]French spelling is notoriously opaque. Although schwa generally corresponds to an <e>, the reverse is not true and there are many instances of <e> which correspond to [ɛ] (e.g. *fer* [fɛr] 'iron') or are part of a grapheme (e.g. *eau* /o/ 'water'). There are also a few cases where schwa is spelled differently due to sporadic historical change (e.g. *m̲o̲n̲sieur* /məsjø/ 'sir', *f̲a̲isais* /fəzɛ/ '(I) did'). In order to guide the reader, I will use *e̲*, *e⫽* and *(e)* in orthographic transcriptions to refer to a schwa which is present, absent or optional, respectively.

[3]In some varieties, such as Laurentian French, there may be arguments to analyse schwas in monosyllabic clitics as epenthetic (Côté 2000: 82-83).

1976, 1988, 1995, Moreux 1985a, Watbled 1995, Eychenne 2006, 2014, 2015, Courdès-Murphy 2018, and references therein). Although Southern French is not a homogeneous entity and is undergoing a number of changes, most southern varieties share a cluster of core properties, which Pooley (2007) calls the 'Dominant Southern Pattern'. Following Armstrong & Pooley (2010: 189), we can identify three important characteristics for this 'supra-regional form'. First, schwa is much more stable in Southern French than it is in Northern French. Second, there is a lack of contrast between mid-high and mid-low vowels: mid-high vowels appear in an open syllable not followed by schwa (e.g. *seau* [so] 'bucket', *épais* [e.pe] 'thick.MASC') whereas mid-low vowels appear in a closed syllable (e.g. *sol* [sɔl] 'floor', *sec* [sɛk] 'dry.MASC') or in an open syllable followed by a schwa[4] (e.g. *sole* ['sɔlə] '(fish) sole', *sèche* ['sɛʃə] 'dry.FEM'), a pattern known as the *loi de position* (Rochet 1980, Moreux 1985a,b, Durand 1995, Eychenne 2014, Storme 2017)[5]. Third, 'nasal vowels' are actually realised as oral vowels followed by a nasal glide (or 'nasal appendage'), which is usually velar, especially in word-final position ([ɛ^{ŋ} œ^{ŋ} ɔ^{ŋ} a^{ŋ}]), or sometimes homorganic with the following consonant if it is a stop (e.g. *compter* [kɔ^{n}te] 'to count', *tomber* [tɔ^{m}be] 'to fall') (Coquillon & Turcsan 2012)[6]. The relation between schwa and mid vowels has been extensively discussed in the literature (Durand 1976, 1995, Rochet 1980, Moreux 1985a,b, Watbled 1995, Andreassen & Eychenne 2013, Eychenne 2006, 2014, 2015, Courdès-Murphy 2018) and I will have nothing to add about this issue. On the other hand, the phonological status of nasal vowels will be relevant for this paper and we will discuss it in more detail below (see §2.2.4).

In the traditional supra-regional form of Southern French, a lexical schwa is systematically realised, as shown in (2), unless it occurs in morpheme-final position and is followed by a vowel within the same IP, as in (3) (see Durand et al. 1987, Durand 1995: 987). In addition, this variety displays a contrast between words ending with a consonant (e.g. *net* [nɛt] 'neat.MASC', *golf* [gɔlf] 'golf') and words ending with a consonant followed by schwa (e.g. *nette* ['nɛtə] 'neat.FEM', *golfe* ['gɔlfə] 'gulf'). This phonological contrast generally correlates with the absence vs presence of an <e> in the spelling, although there are sporadic exceptions for some speakers (Durand & Eychenne 2004).

(2) a. *bête* /bEtə/ → [bɛtə]
'foolish'

b. *petite fille* /pətit+ə#fijə/ → [pøtitəfijə]
'little girl'

(3) a. *bêtise* /bEtə+izə/ → [betizə]
'foolishness'

b. *petite ardoise* /pətit+ə#ardwazə/ → [pøtitardwazə]
'small slate'

[4]Word-final schwa in Southern French is usually represented with the symbol ə in both phonetic and phonological forms, and I will not depart from this convention. According to Durand (1995), the quality of this vowel is quite variable and varies between [ə], [ʌ], [œ] and [ø], depending on the speaker and/or dialect.

[5]Since there is no contrast between mid-high and mid-low vowels, the pairs [e ∼ ɛ], [ø ∼ œ] and [o ∼ ɔ] are commonly represented with the symbols /E/, /Œ/ and /O/ respectively in underlying forms. I will follow this convention in the remainder of this paper.

[6]Since the precise realisation of the nasal glide is not the focus of this paper, I will use a superscript /N/ to represent it.

A number of studies have shown that the realisation of this vowel has now become variable in non-pre-vocalic contexts for many Southern French speakers, especially younger ones (Durand et al. 1987, Taylor 1996, Armstrong & Unsworth 1999, Eychenne 2009a,b, Lonnemann & Meisenburg 2009, Coquillon & Turcsan 2012, Courdès-Murphy 2018). This change in progress has been described in the sociolinguistic literature as an instance of dialect levelling, whereby the supra-local norm is spreading southwards (Pooley 2007). It is undeniable that sociolinguistic pressures must play a role in this change, since levelling in metropolitan French is reported to be widespread and to affect many aspects of phonological structure (Armstrong & Pooley 2010). However, the precise factors which condition the loss of schwa in this variety are still not fully understood, and the main goal of this study is to provide a more accurate picture of the variables that influence its deletion in word-final position, which is the position where deletion is in the most advanced stage (Durand et al. 1987). To this end, this paper offers a reanalysis of data from four surveys from the PFC database, which are presented in more detail in §3.1. There are several reasons why a reanalysis of these data is called for. First, while each individual study offers valuable information about the variety it focuses on, the factors investigated, as well as the methods used vary, making it difficult to know what holds for Southern French as a whole and what might be more specific to a given variety. Second, the statistical methods which are now available (generalised linear mixed-effects models, henceforth GLMM's) allow us to probe phonological variation with increased precision, and in much greater depth, than was previously possible (Bürki et al. 2011a, Bayles et al. 2016). Third, as reviewed below, there are several potential factors whose influence (or lack thereof) is simply unknown.

The remainder of this paper is organised as follows. Section §2 reviews the potential factors influencing the realisation of schwa that have been taken into account in this study. Section §3 introduces the corpus that was constructed from the PFC database and the statistical methods that were employed to analyse it. Section §4 presents the results from this study. Section §5 provides a general discussion of schwa in Southern French in light of these results, and Section §6 offers some brief concluding remarks.

2 Potential factors affecting the realisation of schwa

There are a number of well-known factors which affect the presence/absence of schwa, both in Southern French alone and across varieties. I review some of these factors[7] below and discuss others which, to the best of my knowledge, have never been investigated empirically in this variety but are known (or argued) to have a potential effect, either in French or in other languages.

2.1 Social and situational factors

2.1.1 Sociolinguistic variation

Sociolinguistic variables, especially age and gender, have been singularly well documented in southern varieties of French: this is most likely due to the fact that schwa in these

[7] An important variable that could not be taken into account in this study is speech rate. See for instance Bürki et al. (2011a), Dell (1985: 230) and Malécot (1976), among others.

varieties represents an excellent test ground for sociolinguistic theories of dialect convergence and levelling (Durand et al. 1987, Taylor 1996, Pooley 2007, Armstrong & Pooley 2010, Courdès-Murphy 2018). While the effect of age generally agrees with expectations (younger speakers tend to delete schwa more than older speakers), the effect of gender is unclear. Armstrong & Unsworth (1999), who analysed a corpus of speakers aged 16-25 in the Aude *département* (one of the four localities analysed in the current study), reported a significant effect of gender within this age group: female speakers showed a greater tendency to delete schwa than male speakers. However, Courdès-Murphy (2018) compared the realisation of schwa in Toulouse (South West) and Marseilles (South East) across a much broader age range and found no main effect of gender in either location. She noted a mild interaction between age and gender in Marseilles for her second youngest age group (out of four), males showing a *greater* rate of deletion than females. However, she pointed out that this apparent effect was most likely a by-product of these male subjects having lived in Northern France for several years and showing less regional attachment. (On the possible role of regional attachment, see Armstrong & Pooley 2010: 253–254.)

2.1.2 Register and style

Like most linguistically variable phenomena, the realisation of schwa is partly conditioned by situational factors (Malécot 1976, Lucci 1983). For instance, Lucci (1983) compared the rate of realisation of schwa among several professional speakers across several settings (conference, reading task, interview and conversation). His results for proclitics (*je* 'I', *me* 'me', *te* 'you (object)', *ne* 'not', *le* 'him, it', *se* 'self (third person)' and the complementizer *que*) in non-initial position, a position where schwa is highly variable, indicate that the presence of schwa is sensitive to the level of formality. Schwa is most frequent in the reading task (59.90%), followed by the interview (55.84%), the conference speech (50.90%) and the conversation (32.29%). Although it is difficult to know to what extent these differences are significant, they are at least suggestive. A number of analyses of surveys from the PFC database give further support to Lucci's and Malécot's findings. Schwa has been reported to be more frequent in reading than in conversation in Belgian French (Hambye & Simon 2009), in Northern French (Pustka 2009) as well as in Southern French (Eychenne 2006, Lonnemann & Meisenburg 2009, Courdès-Murphy 2018). However, none of these studies found a difference between the formal and informal conversation styles in the PFC data.

2.2 Phonological factors

2.2.1 Phonotactic and prosodic conditioning

Among the linguistic factors, the best known linguistic variable influencing the realisation of schwa is undoubtedly its segmental context, an effect which has been documented since the earliest descriptive and prescriptive works (e.g. Fouché 1956, Delattre 1966). The most famous statement is Grammont's (1920) *loi des trois consonnes* ('law of three consonants', henceforth Grammont's law)[8], according to which schwa is usually realised when it is preceded by two consonants and followed by one (or more) consonants, but

[8]See Durand & Laks (2000) for a historical review of Grammont's law.

may be omitted when it is preceded by a single consonant. (See the exemples in (1) discussed in the introduction.) Although Grammont's law, stated in its simplest form, generally holds at a very coarse level for non-Southern varieties, the precise nature of the consonants surrounding schwa plays an important role, as Grammont himself had already noted (see also Malécot 1976). Dell (1985) points out that in his own dialect, the realisation of schwa is obligatory when it is preceded by an obstruent+liquid (OL) cluster (e.g. *arbre pourri* [arbrəpuri] 'rotten tree'), unless the cluster is simplified and the liquid is deleted (e.g. [arbpuri]). More generally, schwa is more likely to drop if the cluster resulting from its absence obeys the SONORITY SEQUENCING PRINCIPLE (Côté 2000: 112–119; Bürki et al. 2011a). However, there is cross-dialectal variation and Morin (1983) observes that in the variety of French spoken in Saint-Étienne, schwa can delete in contexts where it is usually expected to maintain (e.g. *la mère de Guy* [lamɛrdgi] 'Guy's mother'; *des vagues de quinze mètres* [devagdkɛ̃zmɛtr] 'fifteen-meters high waves'; *porte-feuille* [pɔrtfœj] 'wallet').

The (non-)realisation of schwa within a complex consonant cluster is further influenced by its prosodic position. It has been repeatedly observed that schwa is more likely to surface in the penultimate position of an intonational phrase (Léon 1966; Morin 1983: 82; Dell 1985; Côté 2007). For instance, Dell (1985: 231) suggests that schwa is more likely to be realised in *la terre se vend* 'land sells', where the following syllable is stressed, than in *la terre s(e) vend bien* 'land sells well', where the syllable following the potential schwa is unstressed. Côté (2007) reinterprets this phenomenon in terms of minimality constraints on the prosodic word and suggests that the realisation of schwa can also be influenced by the number of syllables *preceding* it: according to her, schwa is more likely to be realised in *l'acte commence* 'the act begins' than in *l'entract(e) commence* 'the intermission begins'. Another context where prosody happens to play a role is the right edge of the IP. As pointed out in the introduction, a schwa-like vowel is often realised in this context, and this can happen irrespective of the number of consonants that precede it. This phenomenon has been reported in the speech of young Parisians (Hansen 1997), but it seems to be quite widespread, at least in metropolitan France. Eychenne (2006) observes that a similar phenomenon is at play in Southern French: word-final lexical schwa is more likely to be preserved at the end of an IP than in non-final position, and some speakers even display schwa epenthesis at the end of an IP even when it is preceded by a single consonant (e.g. *vin blanc sec_* [vɛ$^{\text{N}}$bla$^{\text{N}}$sɛkə //] 'dry white wine'), although the final trochee is not realised with the pitch contour typically found with prepausal schwa in Northern French (Fagyal 2000, Biers 2017).

2.2.2 Consonant preceding schwa

Durand et al. (1987) made an interesting observation about words ending in VC# in Southern French. In this variety, voiced obstruents (especially voiced fricatives) are rare in this position and they tend to be avoided in surface representations. For example, the word *merguez* 'spicy sausage', a borrowing which is realised as [mɛrgɛz] in Northern French, is variably assimilated as [mɛrgɛs], by devoicing the final /z/, or [mɛrgɛzə], by inserting a schwa. Both repair strategies prevent the appearance of the voiced fricative /z/ in word-final position, and suggest the existence of a constraint such as *z# in Southern French. If phonological constraints of this type are indeed active in this variety, we should

expect them to play a role in the deletion/retention of lexical schwa. The final schwa in a form such as *chemise* [ʃømizə] 'shirt' would be less likely to drop since its deletion would cause /z/ to occur word-finally, thus violating the above-mentioned constraint. Two competing hypotheses are available to account for this effect.

Under the first hypothesis, the constraint against schwa would be a function of the frequency of word-final consonants. Words ending with a consonant which is frequent in word-final position (e.g. words ending with /l/ such as *sel* /sEl/ 'salt', *fil* /fil/ 'thread', *mal* /mal/ 'badly', etc.) would 'attract' words ending with a schwa that is preceded by the same consonant (e.g. *selle* /sElə/ 'saddle', *docile* /dOsilə/ 'docile'): the loss of schwa would be facilitated by the fact that the consonant is well attested in the lexicon in word-final position. Conversely, schwa deletion in a word such as *rêve* /rEvə/ 'dream' would be hindered by the fact that the consonant that precedes schwa (/v/ in this case) is very rare in word-final position (it is only found in borrowings such as *leitmotiv* /lEjtmOtiv/ 'leitmotiv' and *lev* /lEv/ 'lev (Bulgarian currency)'), since the deletion of schwa would create a highly marked or unattested word-final coda. The same reasoning can be straightforwardly extended to word-final consonant clusters, albeit with the provision that these are relatively rare in the first place. For lack of a better term, I shall call this putative effect of final consonants (and consonant clusters) the ATTRACTOR FREQUENCY HYPOTHESIS. To the best of my knowledge, this hypothesis has never been tested empirically. In order to get a better sense of how it might be at work in Southern French, I extracted frequency counts from the 50,000 most frequent words in Lexique (New et al. 2001), which were phonetised according to a typical Southern French pronunciation, as explained in §3.1.4. There were 4,337 words ending in a single pronounced consonant (8.7%), such as *net* [nɛt]. Their frequency is reported in Table 1. As we can see, the liquids /l r/ account for more than two thirds of the occurrences (68.43%). At the other end of the spectrum, voiced fricatives are very rare, and /ʒ/ does not appear at all. The nasal palatal does not appear either in word-final position. In order to ensure that this distribution is not simply a by-product of the consonants' type frequency, I have also extracted the overall type frequency of consonants. Figure 1 represents a plot of the log type frequency of consonants across all contexts against the log frequency of consonants in the V_# context. The solid line represents the expected log frequency of a consonant in V_# context given its observed frequency across the board. Even though /r/ is the most frequent consonant, its frequency in word-final position is far greater than expected. On the other hand, the natural class of voiced fricatives is clearly under-represented word-finally, and their rarity in this context is not simply due to their relatively lower frequency. Therefore, it seems reasonable to consider attractor frequency as a potential predictor of the realization of schwa.

An alternative explanation for Durand et al.'s remarks could be found in the consonant's voicing specification. There is at least suggestive evidence that schwa might interact with consonant voicing. Van Eibergen and Belrhali (1994: 279), based on a corpus study of spontaneous speech in Northern French, observed that voiceless segments tended to favour the non-realisation of schwa in their data. More recently, Bürki et al. (2011a) found that the voicing of the consonant preceding schwa (as well as that of the one following it) affected schwa's duration, although it did not affect schwa deletion. These observations make it worth investigating whether consonant voicing influences the realisation of schwa in Southern French. I shall call this putative effect of voicing the

consonant	frequency	log(frequency+1)	%
r	2,281	7.73	52.59
l	687	6.53	15.84
f	311	5.74	7.17
k	233	5.46	5.37
s	156	5.06	3.60
m	124	4.83	2.86
t	117	4.77	2.70
n	83	4.43	1.91
j	74	4.32	1.71
ŋ	69	4.25	1.59
p	59	4.09	1.36
g	42	3.76	0.97
d	35	3.58	0.81
b	32	3.50	0.74
ʃ	24	3.22	0.55
z	8	2.20	0.18
v	2	1.10	0.05
ʒ	0	0.00	0.00
ɲ	0	0.00	0.00

Table 1. Frequency of word final consonants in Southern French (V_# context)

LARYNGEAL SPECIFICATION HYPOTHESIS.

Although the attractor frequency hypothesis and the laryngeal specification hypothesis more or less agree on the consonants that favour schwa retention (roughly, the class of voiced obstruents), they clearly make different predictions for consonants that favour deletion: according to the attractor frequency hypothesis, the liquids should be the consonants that favour schwa deletion the most, and /r/ more so than /l/, whereas according to the laryngeal specification hypothesis, it is the natural class of voiceless obstruents which favours schwa deletion, irrespective of each consonant's type frequency.

Besides voicing, there is at least indicative evidence that, in Southern French, sonorants and coronals can favour schwa deletion (Eychenne 2006: 189) in VC_# context and that fricatives favour deletion in monosyllabic clitics (Coquillon & Turcsan 2012: 118). Therefore, the place of articulation, continuancy and sonorancy of the consonant before schwa have also been taken into account in the corpus analysis reported below.

2.2.3 Syllable economy

The deletion of schwa always leads to the loss of a syllable and the onset of the schwa-headed syllable must be resyllabified, either in the coda of the preceding syllable or as part of the onset of the following one. Within the framework of Optimality Theory (Prince & Smolensky 2004), Tranel (1999) built upon this observation and proposed that

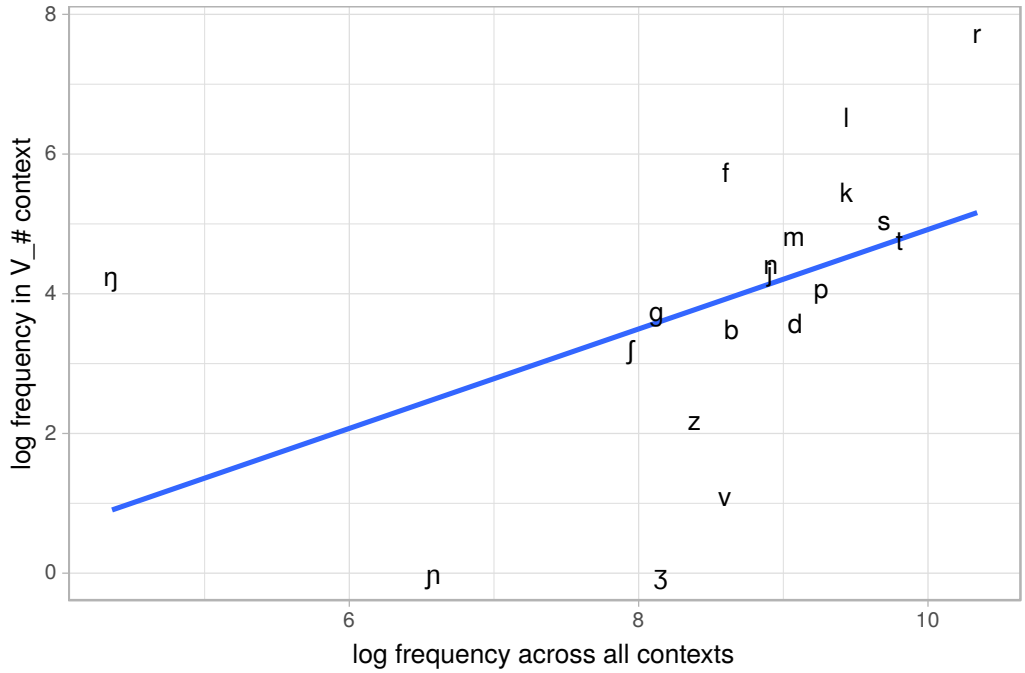

Figure 1. Frequency of consonants across all contexts vs in V_# context

schwa deletion (and epenthesis) is the result of a conflict between a requirement to avoid marked surface structures, such as complex clusters, and the need for the grammar to use no more structure than necessary. More specifically, Tranel's analysis crucially relies on the notion of SYLLABLE ECONOMY, a constraint of the *STRUCT family which penalises syllabic nodes. Under this view, schwa is deleted whenever possible to minimise the number of syllables.

Whether or not one accepts Tranel's formal analysis (see for instance Gouskova 2003 for a criticism of *STRUCT constraints), it does explicitly suggest a link between word size (i.e. the number of syllables) and the (non-)realisation of schwa. A strong interpretation of the role of syllable economy would be that the effect of this constraint is only categorically visible when it dominates the relevant faithfulness and markedness constraints, as proposed by Tranel. In that case, the deletion of schwa would always remove one violation of *STRUCT no matter how long the word is. It is however possible to relax this assumption and to interpret the role of syllable economy as stochastic and gradient. Under this view, longer words could be more likely to lose their schwa than shorter words, everything being equal elsewhere, because they violate syllable economy to a greater extent: the prediction in that case is that there should be a positive correlation between word length and schwa's propensity to be deleted. An anonymous reviewer points out that, if one is to abandon the categorical nature of syllable economy in favour of a gradient version, another interpretation is available: since the deleted syllable in a shorter word represents a larger portion of that word's length, one might expect deletion to be more common in shorter words than in longer words because deletion would in a sense achieve a greater effect in shorter words. In that case, the prediction is that there

should be a negative correlation between word length and schwa deletion.

2.2.4 Nasality of the preceding vowel

As pointed out in the introduction, Southern French is characterized by the absence of genuine nasal vowels, and this is indeed a stereotypical feature of Southern varieties (Pustka 2007). So-called 'nasal vowels' are usually realised as an oral vowel, which may be (partly) nasalised under certain conditions, followed by a more or less prominent nasal glide (Durand 1988, Clairet 2008, Carignan 2017). Durand (1988) put forward several arguments supporting an analysis as /VN/ (instead of /$\tilde{\text{V}}$/) at the phonological level (see also Eychenne 2006 and Courdès-Murphy 2018): although this has not been investigated empirically, a prediction of this analysis is that, everything being equal elsewhere, schwa should be less likely to drop when it is preceded by a VNC sequence, as in *compte* ['kɔ$^{\text{N}}$tə] 'count.3SG.PRES.IND', than when it is preceded by VC, as in *côte* ['kɔtə] 'coast', since the NC sequence should behave like a CC cluster. (Recall that schwa is more likely to be present after CC than after VC in all varieties.)

2.3 Lexical and morphological factors

2.3.1 Word frequency

Word frequency is known to play an important role both in synchrony and in diachrony. As they get automated through repetition, word forms tend to become realised with greater gestural overlap and less precise articulatory control, which can lead to assimilation and phonetic reduction (Bybee 2001, 2015). Such frequency effects have been demonstrated to affect homophone words: Gahl (2008) showed that high frequency words such as *time* tend to be shorter in duration than lower frequency homophones (e.g. *thyme*), which suggests that word forms are associated with frequency information in the mental lexicon (Bybee 2001, Pierrehumbert 2001).

The effect of word frequency on French schwa in currently unclear: a number of studies did point out that the likelihood of schwa deletion was positively correlated with lexical frequency (Dauses 1973, Lucci 1983, Hansen 1994, Racine & Grosjean 2002). However, in a recent corpus study, Bürki et al. (2011a) found that this effect disappeared once the word in which schwa occurred was controlled for by adding a random effect for WORD in their statistical model. This suggests that "there is no effect of lexical frequency on schwa alternation and that its apparent effect is simply attributable to the behavior of some specific (high frequency) words" (Bürki et al. 2011b: 3987). According to these results, there might be more lexical idiosyncrasy in the realisation of schwa than previously thought: while a number of frequent words happen to be particularly prone to schwa deletion, not all frequent words are.

Frequency effects have not been investigated in Southern French in word-final syllable, but they have been in word-initial syllable. Pustka (2007) found that schwa, which is usually stable in this position[9], could only be deleted in frequent words such as *s@maine* 'week' or *p@tit* 'small.MASC' (see also Eychenne 2006, Courdès-Murphy 2018). Pustka

[9]Several authors consider that 'schwa' is a stable /Œ/ in word-initial syllable in Southern French (Durand et al. 1987, Eychenne 2006, 2015, Pustka 2007, Courdès-Murphy 2018).

(2007) further showed that this vowel was more often deleted in highly frequent constructions such as *un p̸etit peu* 'a little bit' than in other constructions (e.g. *il est p̲etit* 'he is small').

Taken together, these results suggest that it is worth exploring frequency effects on the realisation of word-final schwa in Southern French, taking into account the possibility that these effects might in fact be an artefact of lexical idiosyncrasy.

2.3.2 Phonological neighbourhood

In addition to frequency, research has shown that the phonetic realisation and lexical retrieval of a word could be affected by the richness of its phonological neighbourhood, i.e. by words which are similar to it. The most commonly used measure of similarity is NEIGHBOURHOOD DENSITY, which is usually operationalised as the number of words which differ from a target word by the insertion, deletion or substitution of exactly one segment. For instance, in Northern French, the word *mal* /mal/ 'badly' has among its many neighbours the words *mat* /mat/ 'matte' (substitution of /l/), *mule* /myl/ 'mule' (substitution of /a/), *ma* /ma/ 'my.FEM' (deletion of /l/) and *malin* /malɛ̃/ 'cunning' (insertion of /ɛ̃/). Interestingly, since schwa is a feminine marker in Southern French (see §2.3.4 below), there are a number of pairs of words whose masculine and feminine forms are neighbours of each other in this variety, as the only difference is the addition of a schwa in the feminine form (e.g. *net* /nEt/ 'neat.MASC' and *nette* /nEtə/ 'neat.FEM'). This is however not systematic, since some words have the same form in both genders (e.g. *sale* /salə/ 'dirty') while others differ by more than one segment if there is a latent consonant in the paradigm of the word (e.g. *blanc* /blaᴺ/ 'white.MASC' vs *blanche* /blaᴺʃə/ 'white.FEM')[10] or in case of suppletion (e.g. *acteur* /aktŒr/ 'actor' vs *actrice* /aktrisə/ 'actress').

Neighbourhood density has been shown to play a role in speech production, as words with a higher neighbourhood density show evidence for hyperarticulation (Munson & Solomon 2004, Scarborough 2004). If the presence/absence of schwa is to be understood as a continuum, with the possibility of intermediate reduced realisations as some authors have suggested (Eychenne 2006, Bürki et al. 2011b), it is plausible that word-final schwas that appear in a word with a high neighbourhood density would be less prone to phonetic reduction, and hence deletion, than schwas appearing in words with a lower neighbourhood density.

Regarding French schwa, psycholinguistic research has suggested that, at least in Swiss French, neighbourhood density may influence latencies in lexical naming tasks for words that have a schwa in their initial syllable (Bürki et al. 2010); this evidence is only suggestive, however, since the effect of neighbourhood density only showed a statistically non-significant trend. This remark notwithstanding, it seems appropriate to investigate whether neighbourhood density might influence the deletion of schwa in the variety under study.

[10]Whether such latent consonants are part of the underlying representation of the masculine form is an issue which is beyond the scope of this paper. As far as I am aware, latent consonants are never taken into account in the calculation of neighbourhood density in French.

2.3.3 Grammatical category

Bürki et al. (2011a) tested whether the distinction between function words and content words could explain part of the behaviour of schwa in Northern French. They reported that it had no significant effect on alternation with zero and duration. As far as I am aware, the effect of this variable has not been rigorously tested in Southern French but it has been pointed out that schwa deletion appears to be particularly frequent in function words such as *une* 'a.FEM' and *elle* 'she' (Eychenne 2006: 186; Coquillon & Turcsan 2012: 117–118). It is not entirely clear whether this is due to the word's grammatical category per se and/or to the fact that these function words are realised as prosodic clitics. These two variables are in fact very difficult to tease apart since function words are normally realised as clitics. Nonetheless, it is still worth exploring whether there is a difference between these two word classes, keeping in mind that this difference might be (partially) attributable to prosodic structure.

2.3.4 Morphological structure

Morphological structure has been argued to affect the realisation of schwa when it is preceded by two consonants. Dell (1985) observes that in the future and conditional forms, the schwa that appears between the base and the future/conditional suffix can be deleted even if it is preceded by two consonants, as in (4a), but cannot be deleted in morpheme-internal position, as in (4b) (see also Dauses 1973: 52–53). Not everyone agrees that schwas in (4a) should be regarded as lexical, and some authors consider that 'stable schwas', such as those in (4b), have merged with /œ/ in non-alternating contexts since the two vowels cannot be distinguished phonetically and phonologically (Morin 1978). Whatever the case may be, it seems that for some speakers at least, historical schwas can be optionally deleted when they appear in an internal sandhi context and are preceded by two consonants, whereas they cannot in morpheme-internal position in the same segmental environment.

(4) Influence of morphology on schwa deletion (Dell 1985: 231)

a. Optional schwa in internal sandhi:

largu(e)ra	[larg(ə)ra]	'(it) will cast off'	(base: /larg/)
calm(e)rai	[kalm(ə)re]	'(I) will soothe'	(base: /kalm/)
forg(e)ront	[fɔrʒ(ə)rɔ̃]	'(they) will forge'	(base: /fɔrʒ/)
insist(e)ra	[ɛ̃sist(ə)ra]	'(he/she) will insist'	(base: /ɛ̃sist/)

b. Stable schwa in morpheme-internal position:

marguerite	[margərit]	'daisy'
palmeraie	[palmərɛ]	'palm grove'
forgeron	[fɔrʒərɔ̃]	'blacksmith'
fumisterie	[fymistəri]	'sham'

Although the focus of this paper is on the word-final position in Southern French, morphology might play a role there as well. The opposition between word-final consonants (*net* type) and word-final consonants followed by schwa (*nette* type) correlates very well with the orthography, and it is supported by flexional morphology since schwa can be a feminine marker, as in (5), or a verbal suffix for some finite forms, as in (6). It is therefore

possible that morphemic and non-morphemic schwas behave differently. However, the role of morphology could cut both ways.

(5) Schwa as a feminine marker:

correct	[korɛkt]	'correct.MASC'	vs	*correcte*	[korɛktə]	'correct.FEM'
blond	[blɔ^{N}]	'blond.MASC'	vs	*blonde*	[blɔ^{N}də]	'blond.FEM'
Michel	[miʃɛl]	male name	vs	*Michelle*	[miʃɛlə]	female name

(6) Schwa as a verbal suffix:

parler	[parle]	'to talk'	vs	*parle*	[parlə]	'talk.1SG.PRES.IND'
aimer	[eme]	'to like'	vs	*aime*	[ɛmə]	'like.1SG.PRES.IND'
venir	[vønir]	'to come'	vs	*vienne*	[vjɛnə]	'come.3SG.PRES.SUBJ'

On the one hand, it is possible that schwa is more robust to deletion in forms where it realises a morphological category, such as the examples above, than in forms where it does not (e.g. *pôle* [pɔlə] 'pole', *heure* [œrə] 'hour'). It has indeed been argued that morphological and phonological constraints interact within the same grammar. For instance, Rice (2007) used constraints of the MAX{CAT} family, which require the preservation of morphological categories in surface forms, to model paradigmatic gaps in a number of languages. (See also the REALIZEMORPHEME constraint in Kurisu 2001.) If constraints of this type are active in Southern French, they could protect morphological schwas to a certain extent.

On the other hand, there is evidence that morphemic segments can be *shorter* than non-morphemic ones. In a carefully controlled corpus study, Plag et al. (2017) found that the duration of word-final [s] and [z] in English was shorter when these consonants were morphemic (contracted form of *is* and *has*, genitive) than when they were not. If the same phenomenon is at play in Southern French, it is possible that the kind of durational difference that Plag et al. (2017) observed could facilitate schwa deletion: morphemic schwas would be expected to be shorter on average and thus more prone to deletion than non-morphemic ones. Both predictions will be put to a test below.

3 Materials and methods

3.1 Data

3.1.1 PFC surveys

The data used in this study come from the PFC database, a large phonological database which contains material from over 600 French speakers from all over the world, sampled according to a common protocol. For each subject, the database contains the recording of a word list, a short read text, a formal interview and an informal interview. All the files are transcribed orthographically, and transcriptions are time-aligned. In addition, the text and the interviews are coded for two phonological variables, namely liaison and schwa. Each survey typically contains ten to twelve subjects, chosen so as to obtain a balanced sample with respect to gender and age. Although the database contains some socio-economic indicators such as occupation, social class is not rigorously controlled for.

I used the four survey points which were readily available in the PFC database for the southern part of France. These surveys come from four different administrative

districts (French *départements*). Two of the four locations represent rural areas and are small towns with less than 3,000 inhabitants, namely Douzens (département: Aude) and Lacaune (département: Tarn). These two locations are both geographically and linguistically close. Geographically, they are less than 100 km apart and belong to the same administrative *région* called Occitanie. Linguistically, they share the same Romance dialectal substrate (Languedocian Occitan), and I will from now consider these surveys as representative of the Languedocian sub-dialectal area.

The other two surveys represent urban centers. Marseilles (département: Bouches du Rhône) is the second largest French city. It is a large port city which has received a large influx of immigrants, mostly from Italy and North Africa. The local substrate variety of Occitan is known as Provençal. Finally, the last survey point is Bayonne-Anglet-Biarritz (département: Pyrénées Atlantiques), a former conurbation totalling about 120,000 inhabitants. It is located in the French part of the Basque country: the substrate language is Basque, which is genetically unrelated to the surrounding Romance languages spoken in France and Spain. Although this region is also influenced by the Gascon Romance dialect (Haase 1981), all the subjects in this survey point showed a strong attachment to the Basque language and culture[11].

Since, as noted above, the survey points in Lacaune and Douzens share the same substrate, the four survey points were grouped into three dialectal areas: the survey point in Bayonne-Anglet-Biarritz (French département: Pyrénées Atlantiques) represents the Basque dialectal area, Marseilles (Bouches du Rhône) represents the Provençal dialectal area, and Douzens (Aude) and Lacaune (Tarn) represent the Languedocian area. Several aspects of the phonology of the varieties analysed in this study have been described in a number of works. See in particular Durand & Tarrier (2016) and Eychenne (2006, 2009b) for Douzens; Durand et al. (2004) and Eychenne (2006, 2009a) for Bayonne- Anglet-Biarritz; Lonnemann & Meisenburg (2009) for Lacaune; Coquillon & Turcsan (2012) and Courdès-Murphy (2018) for Marseilles. However, the present study is the first attempt to model the behaviour of word-final schwa in Southern French using such a geographically broad sample.

3.1.2 Subjects and material

The four surveys that were selected contained recordings and transcriptions from 45 subjects (10 in Douzens, 10 in Marseilles, 13 in Lacaune and 12 and Bayonne-Anglet-Biarritz). An overview of the distribution of the participants' age and gender in the four locations is provided in Figure 2. All the material coded for schwa from all available subjects was used in this study. For each participant, this corresponded to about 10 minutes of conversation (5 minutes from the formal interview and 5 minutes from the informal interview) and one reading of the PFC text. The PFC word list was not used since it is not coded for schwa.

[11]Six of the twelve subjects in this survey are self-reportedly bilingual (French/Basque), two have a good working knowledge of Basque, and the other four have some notions of the language.

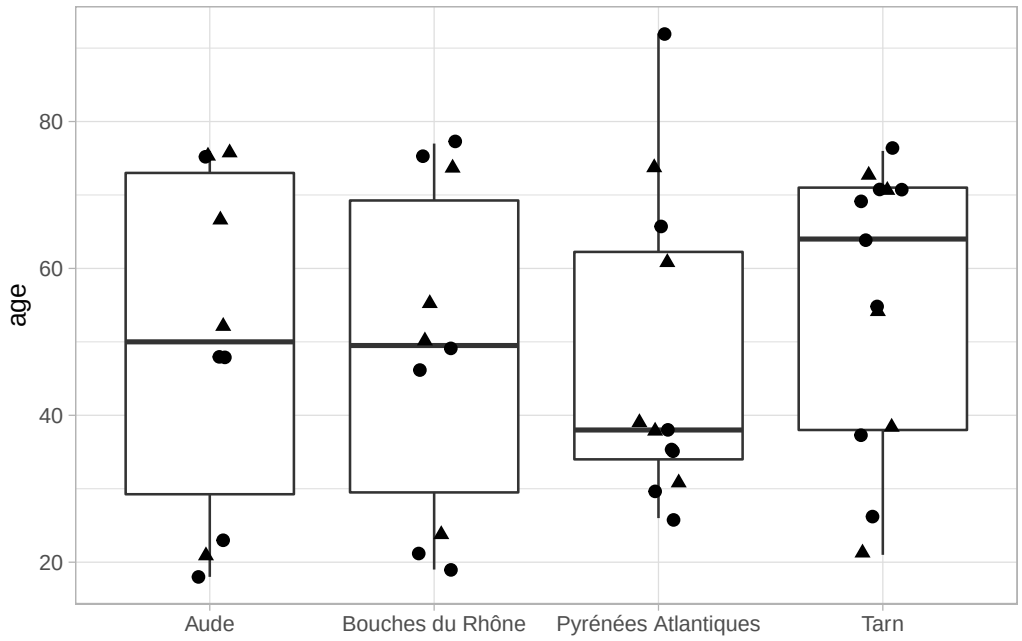

Figure 2. Age of the subjects across the four survey points (● = female, ▲ = male)

3.1.3 Extraction of word-final lexical schwas

The four surveys altogether contain 23,294 schwa codings across all contexts, including 15,261 tokens for the word-final position[12]. (See Lyche 2016 for an overview of the coding system.) The PFC protocol allows for three values for the realisation of schwa: present, absent and uncertain. The 326 tokens which were coded as uncertain were discarded. Since the coding scheme does not presuppose that there is a correlation between orthography (*nette* vs *net*) and phonology (presence vs absence of schwa, [nɛtə] vs [nɛt], respectively), all word-final consonants are coded, whether they are followed by an orthographic <e> or not. Given that the correlation between orthography and the presence of word-final schwa is robust in Southern French, I only considered words which did have an <e> in their spelling, i.e. those words where schwa is assumed to be lexically present ($N = 10,033$). Epenthetic schwas do occur but are quite rare: for words with a final <e>, schwa is realised in 57.6% cases (5,775/10,033) in the database, whereas it only appears in 2.5% cases (124/4,909) when there is no final <e>. The study of epenthetic schwas would certainly be of interest, but with only 124 tokens, it would very difficult, if not impossible, to draw any sort of reliable inference given the number of predictors considered in this study.

Among the words which do have a lexical schwa, many of them (2,224 tokens) appear in prevocalic context, which is a position where schwa is normally categorically deleted (Durand et al. 1987). The database contains 176 tokens (7.9%) coded with a schwa realised in this context, a rate of 'realisation' which is surprisingly high. After listening

[12]Tokens were extracted with Dolmen using the PFC plugin (Eychenne & Paternostro 2016).

to all of these tokens, I determined that most if not all of them belonged to one or more of the following categories: (i) presence of the filler *euh*, often realised with creaky voice; (ii) noticeable prosodic break between schwa and the following vowel, which corresponds to an IP boundary; (iii) presence of an epenthetic glottal stop between schwa and the following vowel (e.g. *pâtes italiennes* [patəʔitaljɛnə] 'Italian pasta'); (iv) slow speech rate and hyperarticulation in the reading task. As a result, the prevocalic context was excluded from the analysis.

Summing up the above, out of the 15,261 tokens coded in word-final position[13], the following types of codings were excluded from the analysis: (i) tokens coded as uncertain ($N = 326$) (i) tokens which did not correspond to an orthographic <e> ($N = 4,902$) (iii) tokens in prevocalic context ($N = 2,224$). Additionally, 5 more tokens were discarded because their left context was coded as the left edge of an IP, which is a mistake, or as uncertain, and three of these tokens were in fact monosyllables. The final dataset only contains schwa codings in word-final position for which there is an <e> in the spelling (which is assumed to correspond to a phonological schwa), preceded by VC or CC and followed by a consonant or at the right edge of an IP. This dataset contained 7,804 data points in total, with information about the coding (split by field), the left and right context of the concordance, and additional metadata about the survey, the task and the speaker's identifier, gender and age. This dataset contained 17 tokens for which the position had been erroneously coded as word-final: these tokens were removed. During exploratory analysis, I also recoded 46 cases for the left context for which I did not agree with the original coders. The final dataset contains 7,787 data points: there are 6,062 tokens (77.8%) where schwa is preceded by one consonant and 1,725 tokens (22.2%) where it is preceded by two or three consonants. A more detailed overview of the distribution of tokens by dialect and task is provided in Table 2. The full dataset is available in the supplementary materials.

		Task			
		Formal	Informal	Text	Total
Dialect	Languedocian	1,405	947	1,707	4,059
	Basque	665	536	817	2,018
	Provençal	542	449	719	1,710
	Total	2,612	1,932	3,243	7,787

Table 2. Distribution of tokens by dialect and task

3.1.4 Lexical and morphological information

The PFC coding scheme for schwa only provides information about schwa realisation, the position of the vowel in the word and broad information about the left and right phonological contexts (e.g. number of consonants, presence of a prosodic boundary). In order to obtain more information about the factors that may contribute to the deletion

[13]Note that monosyllabic clitics such as *je* 'I', *me* 'me' etc. were *not* included in this analysis: they are not coded as word-final in the PFC protocol and there is evidence that schwas in monosyllables behave differently from word-final schwas (Eychenne 2006, Courdès-Murphy 2018).

of schwa, I used the Lexique database[14] (New et al. 2001), a large lexical database which contains, among other things, phonetic transcriptions for all entries. Since these phonetic forms represent standard French, I automatically converted them using a custom script so that they represent canonical southern phonological forms, as found in the Dominant Southern Pattern: contrasts between mid-high and mid-low vowels were neutralized and a schwa was inserted at the end of a word if the standard phonological form ended with a consonant but the orthographic form ended with <e>, <es> or <ent>. For example, a final schwa was added to /ʃɑ̃t/ in the following words: *chante* 'sing.1SG.PRES.IND', *chantes* 'sing.3SG.PRES.IND' and *chantent* 'sing.3PL.PRES.IND'.

Next, for each token in the corpus, I extracted the word in which the schwa coding appeared along with its phonological form, syllable count and token frequency. (Words which were absent from Lexique were processed manually.) Neighbourhood density statistics, although readily available in Lexique, were re-calculated using the southern phonological forms to ensure that they were adequate for this corpus[15]. Two words were considered neighbours if their phonological forms differed by exactly one insertion, one deletion or one substitution (in other words, if their edit distance was equal to 1).

In order to obtain frequency estimates, I first considered using the frequencies provided by Lexique but this proved problematic since the corpus contained a number of words which were absent from this database. These words were mostly proper nouns, such as *Toulouse* and *Ardennes*, which are quite common in that speech community. (Note that Lexique does not contain any proper noun.) Therefore, I simply used frequency estimates from the PFC data: I calculated the token frequency using the transcriptions of the conversations in the four surveys, and I added one instance of the PFC text in order to ensure that all the words which appeared in the reading task, but not in conversation (e.g. *ministre* 'minister'), had a non-null frequency. This represented a total of 158,680 tokens, corresponding to 7,068 types. A list of the 30 most frequent words containing a final schwa is included in the supplementary material.

Finally, I used the information about parts of speech in Lexique to create a categorical predictor for GRAMMATICAL CATEGORY with two levels: function word vs lexical word. Function words represented 13.4% of the tokens in the dataset ($N = 1045$). The presence/absence of a morphological boundary before the schwa was coded manually by the author, who is a native speaker of French.

3.1.5 Phonological information

Using the phonological forms gathered from the previous step, I identified the nature of the two segments that precede schwa in the phonological representation. This was either a VC sequence, where V was oral (e.g. *côte* 'coast') or nasal (e.g. *compte* 'count.1SG.PRES.IND'), or a CC cluster; in the latter case, I also identified the nasality of the preceding vowel (e.g. *possible* 'possible' vs *membre* 'membre'). The consonant before schwa was decomposed into features for VOICING[16], CONTINUANCY, and SONO-

[14] http://www.lexique.org

[15] Since Lexique is quite large and contains many rare words, I only used the 50,000 most frequent words to derive neighbourhood density statistics.

[16] Sonorants were treated as [+voice] since there is evidence that this feature can spread in sonorants in Southern French. For example, several of the speakers in the corpus pronounced the words *slip* 'pants', *socialisme* 'socialism' and *islamique* 'Islamic', which are present in the PFC word list, with a regressive

RANCY, as shown in Table 3. I also added the feature VOCOID to distinguish glides from liquids. Place of articulation was coded according to the following privative features: [labial], [coronal], [palatal], [dorsal]. The phonological status of /r/ is somewhat ambiguous in French since it is generally realised as a uvular fricative or approximant, even though it patterns with /l/ as a liquid (Côté 2000: 114). Some of the older speakers in this corpus have an alveolar flap/trill, and one, in the Basque Country, displays free variation between uvular and alveolar realisations. Therefore, following Russell Webb (2009), I treated /r/ as a placeless approximant at the phonological level.

The LEFT CONTEXT and RIGHT CONTEXT were coded as categorical factors (VC_# vs CC_#, and consonant vs IP edge, respectively). I considered using a more fine-grained coding for the left context in order to take into account the nature of the consonant and/or cluster that precedes schwa, but the resulting models did not converge when the interaction between the left and right context was added. Instead, I adopted a strategy loosely inspired by Bürki et al. (2011b) and decided to code the SONORITY of the consonant/cluster before schwa. Consonants were grouped into broad natural classes according to the following scale: obstruents $\prec$ nasals $\prec$ liquids. A cluster had a rising sonority profile (+1) if the first consonant had a lower sonority than the second (e.g. *autre* /Otrə/ 'other', *rythme* /ritmə/ 'rhythm'), a falling sonority profile (-1) if the first consonant had a greater sonority than the second (e.g. *adulte* /adyltə/ 'adult', *énorme* /EnOrmə/ 'huge'), and a level sonority (0) if both consonants had the same sonority (e.g. *reste* /rEstə/ 'stay', *parle* /parlə/ 'speak'). Single consonants were assigned a level sonority profile.

Consonant class	Voiced	Continuant	Sonorant	Vocoid
Voiceless plosives	-	-	-	-
Voiced plosives	+	-	-	-
Voiceless fricatives	-	+	-	-
Voiced fricatives	+	+	-	-
Nasals	+	-	+	-
Liquids	+	+	+	-
Glides	+	+	+	+

Table 3. Feature specification of consonants

Finally, the type frequency of attractor consonants and consonant clusters in V_# context was calculated as discussed in §2.2.2 and presented in Table 1.

3.2 Statistical analysis

The statistical analysis was conducted in R (R Development Core Team 2008), using logistic mixed-effects models fit with the package `glmmTMB` (Brooks et al. 2017). The REALISATION of schwa was treated as the outcome variable of the model. Table 4 offers a summary of the predictors that were taken into account as fixed effects. In addition to the main effects, I considered the interaction AGE × GENDER, since these variables have

voicing assimilation ([zlip], [sosjalizmə] and [izlamikə], respectively).

often been reported to interact in the sociolinguistic literature, as well LEFT CONTEXT × RIGHT CONTEXT since according to Grammont's law the effect of the right context (consonant or IP edge) crucially depends on what precedes schwa. Moreover, I took into account the interaction between DIALECT and all the linguistic predictors in order to check whether the importance of these predictors differed across dialects.

Variable	Comment
AGE	centred around the median ($M = 50$)
GENDER	ref. = male
DIALECT	ref. = Languedocian
TASK	ref. = formal interview
WORD LENGTH	number of syllables
WORD FREQUENCY	log-transformed
NEIGHBOURHOOD DENSITY	log-transformed
RIGHT CONTEXT	ref. = consonant
GRAMMATICAL CATEGORY	ref. = content word
VOWEL NASALITY	ref. = oral
CONSONANT VOICING	ref. = [-voice]
CONSONANT CONTINUANCY	ref. = [-continuant]
CONSONANT SONORANCY	ref. = [-sonorant]
CONSONANT VOCOIDNESS	ref. = [-vocoid]
CONSONANT PLACE	ref. = [coronal]
MORPHOLOGY	ref. = no morpheme boundary
ATTRACTOR FREQUENCY	log-transformed
SONORITY	

Table 4. Summary of the predictors considered in this study

Subject-level and word-level variation was controlled for in all models by including a random intercept for SUBJECT and for WORD. In addition, a random slope for TASK by SUBJECT was added to model the fact that speakers appeared to differ in their sensitivity to the effect of TASK[17]. In order to find the most parsimonious model for a given context, I adopted a stepwise selection approach similar to Gries (2015): I started with a maximal model containing all the fixed effects for that context and removed non-significant factors one at a time. Fixed effects were considered significant if their p-value was less than 0.05 and if they improved the fit of the model, as estimated by the Akaike Information Criterion and χ^2 likelihood ratios. Post-hoc tests for multi-level categorical predictors were conducted with the package `lsmeans` (Lenth 2016).

One non-negligible issue with generalised linear (mixed-effects) models is that model fit can be difficult to assess because residual plots are much harder to interpret than in classical linear regression. In order to assess the quality of the fit, I used the R package `DHARMa` (Hartig 2019), which produces scaled residual plots (normalized to values between 0 and 1) based on simulated samples generated from the fitted model. Additionally, `DHARMa` provides several formal goodness-of-fit tests to check for common misspecification

[17]I tried to consider several additional random slopes but the resulting models failed to converge.

problems, including a test to check the uniformity of the scaled residuals. The residuals of the final model reported in the next section were checked for uniformity in order to ensure that the fit was adequate.

Finally, Nakagawa & Schielzeth's (2013) pseudo-R^2, as implemented in the R package `MuMIn` (Bartoń 2018), was used to estimate the amount of variance explained by the model. This metric provides values for the fixed effects alone (marginal R^2) and for the combination of fixed and random effects (conditional R^2).

4 Results

This section reports on the best fitting model. Since the outcome variable represents the (log) odds of schwa realisation, a positive regression coefficient means that the predictor favours schwa retention, whereas a negative coefficient means that the predictor favours schwa deletion. The fixed effects of the model are reported in Table 5. Likelihood ratio tests showed that this model, which contained a random intercept for both SUBJECT and WORD and a random slope for TASK by SUBJECT, was superior to a model without the random slope ($\chi^2(1) = 65.63, p < 0.0001$), or to a model that only had grouping at the subject level ($\chi^2(1) = 342.51, p < 0.0001$) or at the word level ($\chi^2(1) = 283.07, p < 0.0001$). In addition, a one-sample Kolmogorov-Smirnov test revealed no significant deviation from uniformity in the scaled residuals ($p = 0.164$), which indicates that the fitted model is adequate. The fixed effects of this model accounted for about half of the variance in the data (marginal $R^2 = 0.520$) whereas the full model accounted for about two thirds (conditional $R^2 = 0.661$).

The following variables were found to have no significant effect on the deletion of schwa once the full random effects structure was taken into account: MORPHOLOGY, NEIGHBOURHOOD DENSITY, ATTRACTOR FREQUENCY, CONSONANT PLACE, WORD LENGTH and CONSONANT VOCOIDNESS. In addition, although the random slope for TASK by SUBJECT was significant, there was no main effect of TASK.

Furthermore, we can see that there is no main effect for GENDER, although there is a significant interaction between AGE and GENDER. In order to get a better understanding of this interaction, Figure 3 displays the realisation of schwa as a function of age across both genders: for subjects below 50 (the median age), the rate of realisation tends to be lower for males than for females.

Besides AGE and GENDER, we see that the rate of realisation differs by DIALECT: the Basque and Provençal areas both display a significantly lower rate of realisation than Languedocian (see Figure 4). A post hoc analysis using Tukey contrasts showed that the difference between Basque and Provençal was not significant ($p = 0.779$).

Turning our attention to the phonological predictors, we see a significant effect of the LEFT CONTEXT and RIGHT CONTEXT, as well as a significant interaction between them. In order to better understand how these contexts interact, let us look at the rate of realisation in the four environments they define (see Table 6). Interestingly, the rate of realisation at the end of an IP is virtually the same whether there is a single consonant (78.5%) or a cluster (77.7%) on the left. However, within an IP, the VC_#C environment favours schwa deletion (61.0%), whereas the CC_#C environment clearly favours schwa retention (92.0%).

	β	std. error	z value	p value
(Intercept)	4.60	0.32	14.36	< 0.0001
DIALECT = BASQUE	-3.00	0.33	-9.22	< 0.0001
DIALECT = PROVENÇAL	-2.82	0.35	-8.15	< 0.0001
AGE	0.07	0.01	7.81	< 0.0001
GENDER = FEMALE	-0.14	0.22	-0.64	0.524 (*n.s.*)
AGE × GENDER	-0.05	0.01	-4.35	< 0.0001
TASK = INFORMAL	0.12	0.11	1.04	0.300 (*n.s.*)
TASK = TEXT	0.25	0.18	1.42	0.156 (*n.s.*)
GRAMMATICAL CATEGORY = FUNCTIONAL	-0.53	0.24	-2.26	< 0.05
FREQUENCY	-0.49	0.12	-4.28	< 0.0001
DIALECT = BASQUE × FREQUENCY	0.17	0.11	1.55	0.121 (*n.s.*)
DIALECT = PROVENÇAL × FREQUENCY	0.26	0.12	2.20	< 0.05
LEFT CONTEXT = CC_#	0.39	0.21	1.876	0.061 (*n.s.*)
RIGHT CONTEXT = CONSONANT	-1.30	0.10	-12.67	< 0.0001
LEFT CONTEXT × RIGHT CONTEXT	2.20	0.23	9.39	< 0.0001
CONSONANT VOICING	0.72	0.19	3.81	< 0.001
CONSONANT SONORANCY	-1.00	0.18	-5.62	< 0.0001
CONSONANT CONTINUANCY	-0.71	0.17	-3.94	< 0.0001
DIALECT = BASQUE × CONTINUANCY	-0.33	0.19	-1.73	0.083 (*n.s.*)
DIALECT = PROVENÇAL × CONTINUANCY	-0.51	0.20	-2.47	< 0.05
SONORITY	0.78	0.19	-4.01	< 0.0001
VOWEL NASALITY	-0.52	0.18	-2.86	< 0.01

Table 5. Fixed effects of the best fitting model

Preceding context \ Following context	_IP]	_C
VC_#	1,722/2,195 (78.5%)	2,360/3,867 (61.0%)
CC_#	471/606 (77.7%)	1,029/1,119 (92.0%)

Table 6. Realisation of schwa as a function of the left and right contexts

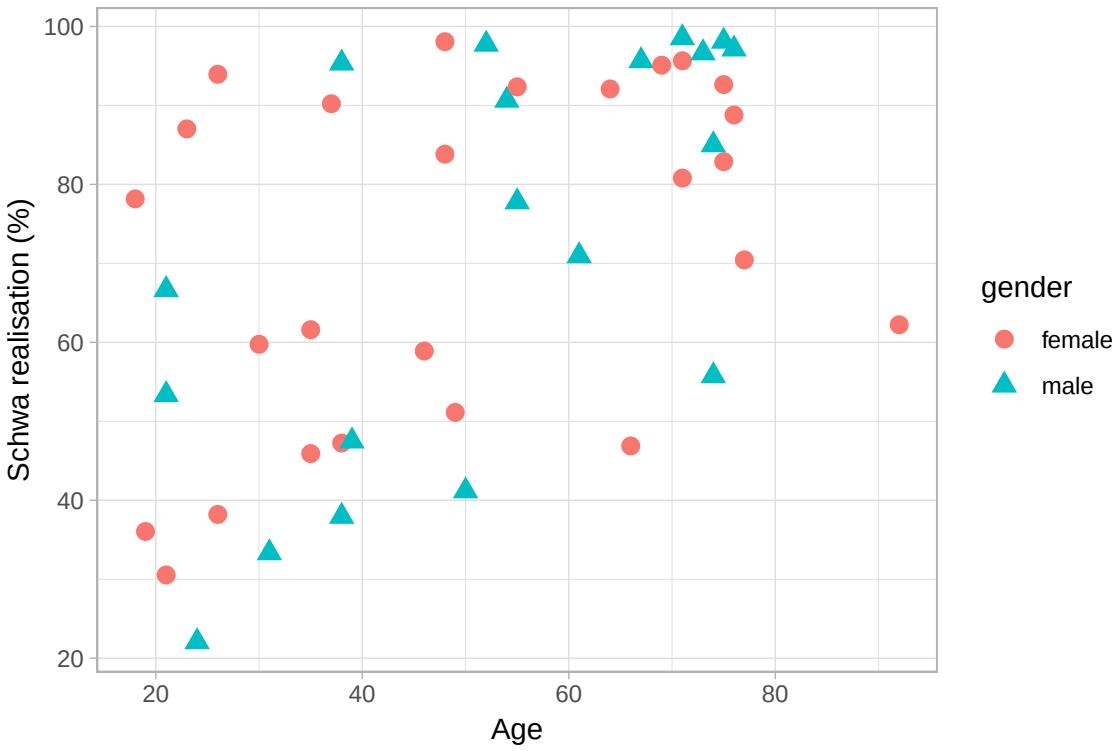

Figure 3. Realisation of schwa by AGE and GENDER (VC_# context)

Investigating the nature of the left context in more detail, it appears that SONORITY is a significant factor in the realisation of schwa: a falling sonority profile favours the absence of the vowel whereas a rising sonority profile promotes its presence. The specification of the consonant preceding schwa also plays an important role: the coefficients for VOICING, CONSONANT CONTINUANCY and CONSONANT SONORANCY altogether show that schwa is more likely to be preserved if the previous consonant is voiced, is an obstruent and/or is a stop. In order to better understand the interplay between these features, let us consider the effect for the six natural classes they define in French: voiced plosives ([-son; -cont; +voice]), voiceless plosives ([-son; -cont; -voice]), voiced fricatives ([-son; +cont; +voice]), voiceless fricatives ([-son; +cont; -voice]), nasals ([+son; -cont; +voice]) and approximants ([+son; +cont; +voice]). As we can see in Table 7, the voiced plosives /b d g/ have the highest rate of schwa realisation, whereas the voiceless fricatives /f s ʃ/ are the consonants which favour schwa deletion the most. Interestingly, the effect of continuancy is not uniform across dialects: it is significantly lower for Provençal than for the other two varieties. The last phonological factor, VOWEL NASALITY, is particularly noteworthy since it is significant, but not in the expected direction: schwa is actually *less* likely to be realised if it is preceded by a nasal vowel than an oral one.

Among the remaining grammatical and lexical factors, only GRAMMATICAL CATEGORY and FREQUENCY turn out to be significant: the rate of schwa retention is higher in content words (4,941/6,742, about 73.3%) than in functional words (641/1,045, about 61.3%) and schwa is more likely to drop in more frequent words than in rarer words.

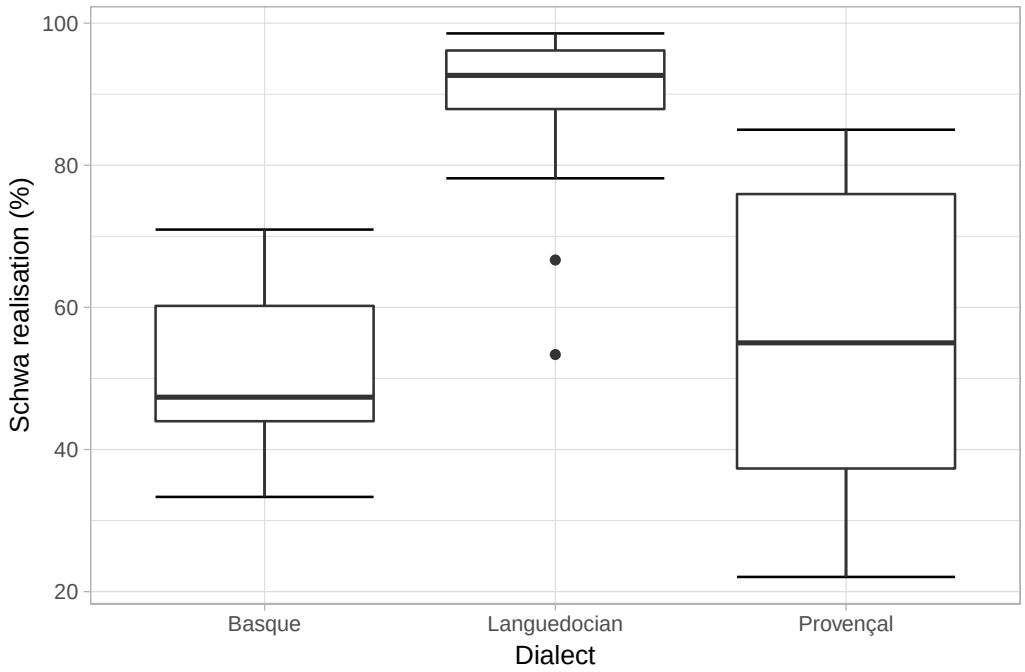

Figure 4. Realisation of schwa across dialects

Class	Schwa realisation	%
voiceless plosives	1,344/1,689	79.6
voiced plosives	225/271	83.0
voiceless fricatives	410/686	59.7
voiced fricatives	528/737	71.6
nasals	1,098/1,525	72.0
approximants	1,977/2,880	68.6

Table 7. Schwa realisation as a function of the natural class of the preceding consonant

Furthermore, the significant interaction between DIALECT and FREQUENCY shows that Provençal is once more singled out: the effect of FREQUENCY is weaker in this variety than in the other two.

5 Discussion

Taken together, the results of this study provide us with a more precise picture of the behaviour of schwa in Southern French. The effect of dialect is clear and confirms previous findings suggesting that the Languedocian area is more resistant to change than the Basque Country and Provence with respect to schwa (Eychenne 2009a,b, Lonnemann & Meisenburg 2009, Coquillon & Turcsan 2012). Although this might indeed be the case, we must remain cautious in our interpretation of this difference since the two Languedocian

surveys represent rural areas, whereas the two cities in the Basque Country and Provence are urban centres. There is some evidence that the rural vs urban distinction might be more important than dialect: Courdès-Murphy (2018) compared, among other things, the realisation of schwa in Marseilles and Toulouse, a large urban centre which is also the main city of Occitanie. She found that the average rate of realisation in both cities was very similar, and that both of them displayed a wide range of inter-speaker variation. Since Toulouse is less than 150 km away from both Douzens and Lacaune (recall that both surveys are part of the région Occitanie), it seems that rurality might be a more important factor than dialectal area. The fact that linguistic innovation tends to first appear in, and spread from (large) urban centers, where speakers are more mobile, is indeed a well established result in sociolinguistics and dialectology (Chambers & Trudgill 1980: 172-178). A large sociolinguistic survey controlling for rurality would therefore be needed to better understand the respective roles of dialectal variation and urbanity in Southern French.

The atypical interaction between age and gender was an unexpected, yet robust finding: although older speakers tend to have a very high rate of schwa realisation irrespective of their gender, younger men displayed a noticeably lower rate of realisation than younger women. This result is at odds with prior work on young Southerners (Armstrong & Unsworth 1999) and more generally with the long-held view that female speakers tend to lead linguistic change and favour standard over non-standard features (Labov 2001). Since the absence of schwa in word-final position is the norm, we would expect (younger) females to display an overall lower rate of realisation of schwa, but this is not what we observe in the data. However, Armstrong & Pooley (2010: §7.3) point to several cases of 'sociolinguistic gender pattern' reversal in French, which might be partly due to social network structures.

The absence of main effect for task is another result that must be emphasised since previous studies have repeatedly found an asymmetry between the reading task and conversation in the PFC data (Eychenne 2006, Lonnemann & Meisenburg 2009, Courdès-Murphy 2018). This discrepancy most likely stems from differences in terms of methodology, since previous analyses used univariate statistics or did not control for as many variables as the present study. Indeed, when we look at the global results, schwa is realised 68.9% of the time in conversation (3,131/4,544), but 75.6% of the time in the reading task (2,451/3,243). A simple chi-square test on these global data *would* find a significant difference ($\chi^2(1) = 6.81, p < 0.01$). The absence of main effect in our data, however, does not mean that TASK is insignificant since including a random slope for TASK by SUBJECT did improve the model. Our results, therefore, offer a more nuanced picture of the role of this factor and suggest that there is a great amount of inter-speaker variability. Part of the explanation might be that some speakers display a different level of sensitivity to the task, and that seems to match my own subjective impression based on a listening of the recordings. Some speakers are equally comfortable in the formal and informal interviews, whereas others have a clear distinction between the two registers: for instance, some speakers quite consistently use the negation clitic *ne* in the formal interview but drop it in the informal interview, as is commonly done in casual conversation. Another possible source of explanation is that some speakers simply almost never delete schwa, and indeed 10 out 45 speakers have a realisation rate greater than 95%. For these speakers, the realisation of schwa is hardly variable at all and there is very little room

for TASK to have any influence.

Let us now turn our attention to lexical factors. The effect of word category shows that schwa is more robust in content words than in functional words. As pointed out in §2.3.3, this asymmetry might simply be the result of the word's prosodic realisation, since functional words tend to be prosodified as proclitics which are dependent on, and usually shorter than, the following lexical head. In order to more precisely determine the respective contribution of word category per se and prosody, one would need a corpus which combines rich syntactic and prosodic annotations, as in the Rhapsodie treebank (Lacheret-Dujour et al. 2019) for instance, with schwa codings such as those from the PFC database. Unfortunately, such a corpus is currently unavailable. The effect of the other significant lexical predictor, namely word frequency, is noteworthy for two reasons: it is the first time that this factor has been demonstrated to play a role in word-final position in Southern French, and it is at odds with recent findings in Northern French. As noted in §2.3.1, Bürki et al. (2011a: 3987) observed that, after controlling for word-level variation in their Northern French data, there was no significant effect of lexical frequency. They concluded that the effect reported in previous studies was probably the result of the behaviour of a number of words which also happen to be frequent. Our results, on the contrary, show that word frequency is significant even after including a random term for WORD. However, it must be emphasised that there are a couple of important methodological differences between the two studies. First, while the current study focused on schwa in word-final position, Bürki et al. (2011b) excluded the word-final position from their analysis since schwa is usually considered as epenthetic in this context in Northern French. It is not clear to me whether this may have had an influence, but this cannot be entirely ruled out. Second, Bürki et al. (2011a) set aside words that were represented by fewer than five tokens in their corpus, whereas I included them all. The downside of Bürki et al.'s approach is that it somewhat distorts the empirical distribution of lexical items: since words follow a power law (Zipf 1949)[18], whereby a small number of items are very frequent (the "fat head") and a large number of items are very rare (the "long tail"), many words end up being discarded. Effectively, the long tail of the distribution ends up being chopped off. This remark notwithstanding, it is easy to determine whether the discrepancy between the two studies could be partly due to their different sampling procedures. Applying Bürki et al.'s approach to the 7,787 tokens in the dataset, I excluded 2,288 items (29.4%). After refitting the model presented in Table 5 to this subset, word frequency still emerged as significant ($p = 0.0007$). A plausible interpretation of this result is that schwa's sensitivity to lexical frequency is indeed different in Southern French and Northern French. This seems to be supported by the fact that the effect of frequency is not uniform across the three dialects: it is weaker in Provence. If sub-dialects of Southern French can differ in their sensitivity to lexical frequency, it might well be the case that some varieties are simply insensitive to it.

Moving on to the phonological predictors that influence the presence/absence of schwa, we can see that the results reported above shed a new light on the complex interaction between the left and right contexts. The highest rate of realisation is found in the CC_C context: as we have seen in §2.2.1, this is the context where schwa is most likely to be realised in Northern French, whether this schwa is lexical (e.g. *par s̲emaine* [parsœmɛn]

[18]See also Laks et al. (2018), who discuss the influence of this distribution on the realisation of French liaison.

'weekly') or epenthetic (e.g. *ours _ brun* [ursœbrɛ̃]). Although epenthetic schwa is rare or inexistent for most varieties of Southern French, the data suggest that Grammont's law is also an important force in this variety: while it usually does not trigger schwa *epenthesis*, it protects schwa from *deletion*. On the other hand, when schwa is preceded by a single consonant, it appears that the IP-final position has a protective role compared to the IP-internal position, where schwa is followed by one or more consonants. This should not come as a surprise: most researchers consider that stress is assigned on the rightmost syllable of the phonological phrase in Northern French (Fougeron & Jun 2006), and the right edge of an IP of course always coincides with the right edge of a phonological phrase. In Southern French, stress is instead assigned on the final trochee, which contains a syllable headed by a full vowel optionally followed by a schwa-headed syllable (Durand 1976, Watbled 1995, Eychenne 2006): since schwa ends up in the dependent position of a stressed trochee, some form of positional faithfulness (Beckman 1997) could protect it from deletion.

The unexpected direction of the effect of vowel nasality is undoubtedly the most puzzling result from this study. However, a closer inspection of the data suggests that most if not all of this effect can be explained away by the behaviour of certain numerals. The words *trente* 'thirty', *quarante* 'forty', *cinquante* 'fifty', *soixante* 'sixty' all end with /-Vᴺtə/: although they were coded as word final, they are frequently encountered in compound numerals such as *mille neuf cent trente-six* '1936', in which case schwa is frequently dropped. These items ($N = 100$) indeed show a very low rate of realisation (30%). After discarding these words and refitting the model, the p-value increased from 0.004 to 0.067, slightly above the standard significance threshold (0.05), which indicates that the observed effect was essentially driven by these items. All things considered, it appears that VNC sequences (e.g. *p<u>en</u>te* /paᴺtə/ 'slope') are different from VCC sequences (e.g. *p<u>ar</u>te* /partə/ 'leave.1SG.PRES.SUBJ') but do not fundamentally differ from VC sequences (e.g. *p<u>â</u>te* /patə/ 'pasta') with respect to schwa deletion. A possible explanation is that schwa and nasal vowels are both undergoing change in Southern French. In a recent study based on 22 speakers from Marseilles (who were not involved in the current study), Courdès-Murphy & Eychenne (ms) reported the existence of a positive correlation between the loss of the nasal glide and the rate of schwa deletion: southern speakers who tend to realise nasal vowels as truly nasalised vowels (instead of oral vowels followed by a nasal glide) also have a lower rate of realisation of schwa. If the same dynamics are at work in our corpus, we indeed expect the effect of nasal vowels not to be significantly different from that of oral vowels.

To finish off, let us look at the role of the consonant (cluster) preceding schwa. Our results concerning the sonority profile of the cluster preceding schwa agree with previous findings: for example, Dell (1977) observed that in Northern French, schwa was more likely to be realized after two obtruents (e.g. *zest(e)* /zɛst/ 'zest') than after a liquid followed by an obstruent (e.g. *farc(e)* /fars/ 'prank'), since the latter type of cluster constitutes a better coda than the former. Regarding the specification of the preceding consonant, it appears that the place of articulation and vocoidness of the consonant have no effect, but sonorancy, continuancy and voicing do. The fact that stops favour schwa retention seems consistent with Côté's (2000: 119) observation, based on Laurentian French data, that stops are preferentially followed by a vowel, since this is the position where their acoustic cues (burst and formant transitions) are perceptually most salient (see also

Côté 2007). Finally, the results for voicing provide empirical support for the laryngeal specification hypothesis presented in §2.2.2. On the other hand, the attractor frequency hypothesis was falsified since attractor frequency turned out to have no explanatory power beyond what is already accounted for by consonant voicing.

6 Conclusion

The main goal of this paper was to offer a better picture of the factors influencing the loss of word-final lexical schwa in Southern French, using a corpus containing 7,787 data points obtained from 45 subjects spread over three dialectal areas, from West to East (Basque, Languedocian and Provençal). The range of variables that were taken into account, as well as the fact that subject-level and word-level variation was controlled for, should make the results generalisable to a large part of Southern France. In addition to clarifying the contribution of sociolinguistic and stylistic factors (age, gender, task), this study has demonstrated the importance of lexical frequency, grammatical category, sonority, as well as the voicing, sonorancy and continuancy of the consonant before schwa.

Although my purpose has been essentially descriptive, the data reported in this work can inform phonological theory in several ways. First, they broaden the range of empirical facts known about French schwa and shed a new light on some well known aspects of the phonology of French. As we have seen, Grammont's law seems to be at work in the CC_#C context in Southern French despite the fact that this variety usually does not have schwa epenthesis: this constraint simply favours schwa retention instead of schwa epenthesis in this environment. Interestingly, the fact that there is no epenthesis does not mean that Southern French always tolerates complex clusters well: a common repair strategy in this variety is consonant deletion. For example, while most northern speakers pronounce the (invented) newspaper name *Ouest Liberté* as [wɛstœlibɛrte] in the PFC text, southern speakers typically delete the final /t/ in the first word, yielding [wɛslibɛrte]. Schwa realisation and consonant deletion are therefore two different strategies to achieve the same output configuration (roughly, no triconsonantal cluster). This is a new illustration, within the same dialect, of what McCarthy (2002) calls HOMOGENEITY OF TARGET/HETEROGENEITY OF PROCESS: the same markedness constraint can be satisfied by two apparently unrelated phonological processes.

Furthermore, the findings reported in this study regarding the effects of word-level variation and word frequency add to an already sizeable body of work which shows the importance of the lexicon in language processing and language change. For example, in the case of Northern French schwa, psycholinguistic research has shown that lexical access latency for word variants with (e.g. *fe̲nêtre* [fœnɛtr] 'window') and without (e.g. [fnɛtr]) a schwa was negatively correlated with the variant's frequency (the more frequent variant is accessed faster): this suggests that both variants are stored in the mental lexicon, along with frequency information. We also know that adults and infants both exploit word-specific information in order to learn phonological categories (Feldman et al. 2013), and there is strong diachronic evidence that the likelihood of losing a phonological contrast in a language is correlated with its functional load in the lexicon (Wedel et al. 2013). Such findings are consistent with the predictions of usage-based and exemplar models which integrate word-specific information (Bybee 2001, Pierrehumbert 2002). At

the same time, the fact that the featural specification of the consonant before schwa affects its loss, irrespective of the consonant's frequency, can be interpreted as pointing to the need for abstract phonological features as are traditionally recognised in formal phonology. What is needed, perhaps, is a hybrid framework that can satisfactorily model word-specific and fine phonetic details as well as abstract phonological primitives, as argued for instance by Nguyen et al. (2009) and Ernestus (2014).

Whether or not one sees the development of such a hybrid model as a desirable (and achievable) goal, the facts discussed in this paper can still inform current formal theories. Over the last few decades, generative phonology has been moving away from the study of an "ideal speaker-listener, in a completely homogeneous speech-community" (Chomsky 1965: 3), and has started recognising the importance of variation and gradience (e.g. Hayes & Wilson 2008, Pater 2009; see Bayles et al. 2016 for a review). Despite the progress that has been accomplished, a lot of work still needs to be done to be able to account for the range of variation found in observational data. The fact that the corpus that was constructed for this study is made available as supplementary material means that theoretical phonologists now have at their disposal another complex, intricate dataset which displays a wide range of intra- and inter-individual variation, and against which they can evaluate competing models of variation.

References

Anderson, Steven (1982). The analysis of French schwa : Or how to get something from nothing. *Lg* 58(3). 534–573.

Andreassen, Helene N. & Julien Eychenne (2013). The French foot revisited. *Language Sciences* 39. 126–140.

Armstrong, Nigel & Tim Pooley (2010). *Social and Linguistic Change in European French.* London: Palgrave Macmillan.

Armstrong, Nigel & Sharon Unsworth (1999). Sociolinguistic variation in southern French schwa. *Linguistics* 37(1). 127–56.

Bartoń, Kamil (2018). *Mumin: Multi-model inference.* `https://CRAN.R-project.org/package=MuMIn`.

Bayles, Andrew, Aaron Kaplan & Abby Kaplan (2016). Inter- and intra-speaker variation in French schwa. *Glossa* 1(1). 19.

Beckman, Jill N (1997). Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 14(1). 1–46.

Biers, Kelly (2017). Vowel Epithesis Variation in French. *Cahiers AFLS* 21(1). 68–101.

Brooks, Mollie E., Kasper Kristensen, Koen J. van Benthem, Arni Magnusson, Casper W. Berg, Anders Nielsen, Hans J. Skaug, Martin Maechler & Benjamin M. Bolker (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal* 9(2). 378–400.

Bürki, Audrey, Mirjam Ernestus & Ulrich Hans Frauenfelder (2010). Is there only one "fenêtre" in the production lexicon? On-line evidence on the nature of phonological representations of pronunciation variants for French schwa words. *Journal of Memory and Language* 62(4). 421–437.

Bürki, Audrey, Mirjam Ernestus, Cédric Gendrot, Cécile Fougeron & Ulrich Hans Frauenfelder (2011a). What affects the presence versus absence of schwa and its duration: A corpus analysis of French connected speech. *JASA* 130(6). 3980–3991.

Bürki, Audrey, Cécile Fougeron, Cedric Gendrot & Ulrich Hans Frauenfelder (2011b). Phonetic reduction versus phonological deletion of French schwa: Some methodological issues. *JPh* 39(3). 279–288.

Bybee, Joan (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Bybee, Joan (2015). Articulatory Processing and Frequency of Use in Sound Change. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford Handbook of Historical Phonology*, 467–484. Oxford: Oxford University Press.

Carignan, Christopher (2017). Covariation of nasalization, tongue height, and breathiness in the realization of F1 of Southern French nasal vowels. *JPh* 63. 87–105.

Chambers, Jack K & Peter Trudgill (1980). *Dialectology*. Cambridge: Cambridge University Press.

Chomsky, Noam (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Clairet, Sandra (2008). Une étude aérodynamique de la nasalité vocalique en français méridional. In *Actes des 27e Journées d'Etudes sur la Parole*, 297–300. Avignon.

Coquillon, Annelise & Gabor Turcsan (2012). An overview of the phonological and phonetic properties of Southern French: Data from two Marseille surveys. In Gess et al. (2012) 105–127.

Côté, Marie-Hélène (2000). *Consonant Cluster Phonotactics: A perceptual approach*: MIT dissertation.

Côté, Marie-Hélène (2007). Rhythmic constraints on the distribution of schwa in French. In José Camacho, Nydia Flores-Ferrán, Liliana Sánchez, Viviane Déprez & María José Cabrera (eds.), *Romance Linguistics 2006*, vol. 287, 79–92. Amsterdam: John Benjamins.

Courdès-Murphy, Léa (2018). *Nivellement et sociophonologie de deux grands centres urbains : Le système vocalique de Toulouse et de Marseille*. Toulouse: Université Jean Jaurès dissertation.

Courdès-Murphy, Léa & Julien Eychenne (ms). Dynamiques à l'œuvre dans le nivellement des voyelles nasales à Marseille. Manuscript.

Dauses, August (1973). *Etudes sur l'e instable dans le français familier*. Tübingen: Niemeyer.

Delattre, Pierre (1966). *Studies in French and Comparative Phonetics*. London: Mouton.

Dell, François (1977). Paramètres syntaxiques et phonologiques qui favorisent l'épenthèse de schwa en français moderne. In Christian Rohrer (ed.), *Actes du Colloque Franco-Allemand de Linguistique Théorique*, 141–153. Tübingen: Niemeyer.

Dell, François (1985). *Les règles et les sons. Introduction à la phonologie générative*. Paris: Hermann 2nd edn.

Detey, Sylvain, Jacques Durand, Bernard Laks & Chantal Lyche (eds.) (2016). *Varieties of Spoken French*. Oxford: Oxford University Press.

Durand, Jacques (1976). Generative phonology, dependency phonology and southern French. *Lingua e Stile* 11(1). 3–23.

Durand, Jacques (1988). Les phénomènes de nasalité en français du Midi: Phonologie de dépendance et sous-spécification. *Recherches Linguistiques* 17. 29–54.

Durand, Jacques (1995). Alternances vocaliques en français du Midi et phonologie du gouvernement. *Lingua* 95. 27–50.

Durand, Jacques, Michel Aurnague & Julien Eychenne (2004). La phonologie du français contemporain au Pays basque. In Maria Pilar Perea (ed.), *Dialectologia i recursos informàtics*, 155–198. Barcelona: Promociones y Publicaciones Universitarias.

Durand, Jacques & Julien Eychenne (2004). Le schwa en français : Pourquoi des corpus ? *Corpus* 3. 311–356.

Durand, Jacques & Bernard Laks (2000). Relire les phonologues du français : Maurice Grammont et la loi des trois consonnes. *Langue française* 126. 29–38.

Durand, Jacques, Bernard Laks & Chantal Lyche (2002). La phonologie du français contemporain: Usages, variétés et structure. In Claus D. Pusch & Wolfgang Raible (eds.), *Romanistische Korpuslinguistik- Korpora und gesprochene Sprache/Romance Corpus Linguistics – Corpora and Spoken Language*, 93–106. Tübingen: Gunter Narr Verlag.

Durand, Jacques, Bernard Laks & Chantal Lyche (eds.) (2009). *Phonologie, variation et accents du français*. Paris: Hermès.

Durand, Jacques, Catherine Slater & Hilary Wise (1987). Observations on schwa in Southern French. *Linguistics* 25(2). 983–1004.

Durand, Jacques & Jean-Michel Tarrier (2016). Variation in a rural village in southern France: Douzens. In Detey et al. (2016) 417–429.

van Eibergen, Joëlle & Rabia Belrhali (1994). Etude statistique du /ə/. In Chantal Lyche (ed.), *French Generative Phonology: Retrospective and Perspectives*, 277–287. Salford: Association for French Language Studies.

Ernestus, Mirjam (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua* 142. 27–41.

Eychenne, Julien (2006). *Aspects de la phonologie du schwa dans le français contemporain: Optimalité, visibilité prosodique, gradience*: Université de Toulouse-Le Mirail dissertation.

Eychenne, Julien (2009a). La prononciation du français au Pays basque. In Durand et al. (2009) 231–258.

Eychenne, Julien (2009b). Une variété de français conservatrice en Languedoc. In Durand et al. (2009) 259–284.

Eychenne, Julien (2014). Schwa and the loi de position in Southern French. *Journal of French Language Studies* 24(2). 223–253.

Eychenne, Julien (2015). Observations on the phonetic realization of opaque schwa in Southern French. *Studies in Phonetics, Phonology and Morphology* 21(3). 457–494.

Eychenne, Julien & Roberto Paternostro (2016). Analyzing transcribed speech with Dolmen. In Detey et al. (2016) D35–D52.

Fagyal, Zsuzsanna (2000). Le retour du e final en français parisien : Changement phonétique conditionné par la prosodie. In *Actes du 12° Congrès International de Linguistique et Philologie Romanes*, Bruxelles.

Feldman, Naomi H., Emily B. Myers, Katherine S. White, Thomas L. Griffiths & James L. Morgan (2013). Word-level information influences phonetic learning in adults and infants. *Cognition* 127(3). 427–438.

Fouché, Pierre (1956). *Traité de prononciation française*. Paris: Klincksieck.

Fouché, Pierre (1969). *Phonétique historique du français : Volume II : Les voyelles*. Paris: Librairie C. Klincksieck.

Fougeron, Cécile & Sun-Ah Jun (2006). Realizations of accentual phrase in French intonation. *Probus* 14(1). 147–172.

Gahl, Susanne (2008). "Time" and "Thyme" Are not Homophones: The Effect of Lemma Frequency on Word Durations in Spontaneous Speech. *Lg* 84(3). 474–496.

Gess, Randall, Chantal Lyche & Trudel Meisenburg (eds.) (2012). *Phonological Variation in French: Illustrations from three continents*. Amsterdam: John Benjamins.

Gouskova, Maria (2003). *Deriving Economy: Syncope in Optimality Theory*: University of Massachusetts at Amherst dissertation.

Grammont, Maurice (1920). *Traité pratique de prononciation française*. Paris: Delagrave.

Gries, Stefan Th (2015). The most underused statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora* 10. 95–125.

Haase, Martin (1981). *Sprachkontakt und Sprachwandel im Baskenland: Die Einflüsse des Gaskognischen und Französischen auf das Baskische*. Hamburg: Buske.

Hambye, Philippe & Anne Catherine Simon (2009). La prononciation du français en Belgique. In Durand et al. (2009) 95–130.

Hansen, Anita Berit (1994). Etude du E caduc — stabilisation en cours et variations lexicales. *Journal of French Language Studies* 4. 25–54.

Hansen, Anita Berit (1997). Le nouveau [ə] prépausal dans le français parlé à Paris. In *Polyphonie pour Iván Fónagy*, 173–198. Paris: L'Harmattan.

Hartig, Florian (2019). *DHARMa: Residual diagnostics for hierarchical (multi-level / mixed) regression models*. `https://CRAN.R-project.org/package=DHARMa`. R package version 0.2.4.

Hayes, Bruce & Colin Wilson (2008). A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *LI* 39(3). 379–440.

Kaplan, Aaron (2011). Variation through markedness suppression. *Phonology* 28(3). 331–370.

Kurisu, Kazutaka (2001). *The phonology of morpheme realization*: University of California, Santa Cruz dissertation.

Labov, William (2001). *Principles of Linguistic Change, Vol. 2: Social Factors*. Oxford & Cambridge: Blackwell Publishers.

Lacheret-Dujour, Anne, Sylvain Kahane & Paola Pietrandrea (eds.) (2019). *Rhapsodie: A prosodic and syntactic treebank for spoken French*. Amsterdam: John Benjamins Publishing Company.

Laks, Bernard, Basilio Calderone & Chiara Celata (2018). French Liaison in the light of corpus phonology: From lexical information to patterns of usage variation. In Diane Brentari & Jackson L. Lee (eds.), *Shaping Phonology*, 278–308. Chicago: University of Chicago Press.

Lenth, Russell V (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software* 69(1). 1–33.

Léon, Pierre (1966). Apparition, maintien et chute du "e" caduc. *La linguistique* 2. 111–122.

Lonnemann, Birgit & Trudel Meisenburg (2009). Une variété française imprégnée d'occitan (Lacaune/Tarn). In Durand et al. (2009) 285–306.

Lucci, Vincent (1983). *Étude phonétique du français contemporain à travers la variation situationnelle*. Publications de l'Université des langues et lettres de Grenoble.

Lyche, Chantal (2016). Approaching variation in PFC: The schwa level. In Detey et al. (2016) 352–362.

Malécot, André (1976). The Effect of Linguistic and Paralinguistic Variables on the Elision of the French Mute-e. *Phonetica* 33(2). 93–112.

McCarthy, John (2002). *A Thematic Guide to Optimality Theory*. Cambridge: Cambridge University Press.

Moreux, Bernard (1985a). La "loi de position" en français du Midi. 1. Synchronie (Béarn). *Cahiers de grammaire* 9. 45–138.

Moreux, Bernard (1985b). La "loi de position" en français du Midi. 2. Diachronie (Béarn). *Cahiers de grammaire* 10. 95–179.

Morin, Yves-Charles (1978). The status of mute "e". *Studies in French Linguistics* 1(2). 79–139.

Morin, Yves-Charles (1983). Quelques observations sur la chute du e muet dans le français régional de Saint-Etienne. *La Linguistique* 19. 71–93.

Morin, Yves-Charles (1987). French data and phonological theory. *Linguistics* 25. 815–843.

Munson, Benjamin & Nancy Pearl Solomon (2004). The Effect of Phonological Neighborhood Density on Vowel Articulation. *Journal of Speech Language and Hearing Research* 47(5). 1048–1058.

Nakagawa, Shinichi & Holger Schielzeth (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4(2). 133–142.

New, Boris, Christophe Pallier, Ludovic Ferrand & Rafael Matos (2001). Une base de données lexicales du français contemporain sur internet : LEXIQUE. *L'année psychologique* 101(3). 447–462.

Nguyen, Noël, Sophie Wauquier & Betty Tuller (2009). The dynamical approach to speech perception: From fine phonetic detail to abstract phonological categories. In François Pellegrino, Egidio Marsico, Ioana Chitoran & Christophe Coupé (eds.), *Approaches to phonological complexity*, 5–31. Berlin & New York: Berlin: Mouton de Gruyter.

van Oostendorp, Marc (2000). *Phonological Projection : A Theory of Feature Content and Prosodic Structure*. Berlin & New York: Mouton De Gruyter.

Pater, Joe (2009). Weighted constraints in generative linguistics. *Cognitive Science* 33(6). 999–1035.

Pierrehumbert, Janet B (2001). Exemplar dynamics: Word frequency, lenition and contrast. In Joan L. Bybee & Paul J. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, vol. 45, 137–158. Amsterdam: John Benjamins Publishing Company.

Pierrehumbert, Janet B (2002). Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology 7*, 101–139. Berlin: Walter de Gruyter.

Plag, Ingo, Julia Homann & Gero Kunter (2017). Homophony and morphology: The acoustics of word-final S in English. *JL* 53(1). 181–216.

Pooley, Tim (2007). Dialect Levelling in Southern France. *Nottingham French Studies* 46(2). 40–63.

Prince, Alan & Paul Smolensky (2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden & Oxford: Blackwell Publishing.

Pustka, Elissa (2007). *Phonologie et variétés en contact. Aveyronnais et Guadeloupéens à Paris*. Tübingen: Narr.

Pustka, Elissa (2009). PFC et la phonologie du français en Vendée. In Durand et al. (2009) 307–335.

R Development Core Team (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. `http://www.R-project.org`.

Racine, Isabelle & François Grosjean (2002). La production du E caduc facultatif est-elle prévisible? Un début de réponse. *Journal of French Language Studies* 12(3). 307–326.

Rice, Curt (2007). Gaps and repairs at the phonology–morphology interface. *JL* 43(1). 197–221.

Rochet, Bernard (1980). The mid-vowels in Bordeaux French. *Orbis* 29. 76–104.

Russell Webb, Eric (2009). Minimalism and French /R/: Phonological representations in phonetically based phonology. *Journal of French Language Studies* 19(1). 87–115.

Scarborough, Rebecca (2004). *Coarticulation and the structure of the lexicon*. Los Angeles, CA: UCLA dissertation.

Scheer, Tobias (2011). Portrait of a phonological cripple: What schwa can(not) do to preceding vowels in French and elsewhere. Paper presented at the International workshop on metrics, phonology and acquisition, Paris 8, 27-28 June.

Selkirk, Elizabeth (1978). The French foot: On the status of "mute" e. *Studies in French Linguistics* 1(2). 141–150.

Storme, Benjamin (2017). The loi de position and the acoustics of French mid vowels. *Glossa* 2(1). 1–25.

Taylor, Jill (1996). *Sound Evidence. Speech Communities and Social Accents in Aix-en-Provence*. Berne: Peter Lang.

Tranel, Bernard (1981). *Concreteness in Generative Phonology. Evidence from French*. Berkeley: University of California Press.

Tranel, Bernard (1999). Optional schwa deletion: On syllable economy in French. In Jean-Marc Authier, Barbar Bullock & Lisa Reed (eds.), *Formal Perspectives on Romance Linguistics*, 271–288. Amsterdam & Philadelphia: John Benjamins.

Watbled, Jean-Philippe (1995). Segmental and suprasegmental structure in Southern French. In John Charles Smith & Martin Maiden (eds.), *Linguistic Theory and the Romance Languages*, vol. 122, 181–200. Amsterdam: John Benjamins.

Wedel, Andrew, Abby Kaplan & Scott Jackson (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128(2). 179–186.

Zipf, George Kingsley (1949). *Human behavior and the principle of least effort: An introduction to human ecology.* Cambridge, Mass: Addison-Wesley.