Ioana Vasilescu*, Ioana Chitoran, Bianca Vieru, Martine Adda-Decker, Maria Candea, Lori Lamel and Oana Niculescu

Studying variation in Romanian: deletion of the definite article *-l* in continuous speech

https://doi.org/10.1515/lingvan-2017-0046
Received October 12, 2017; accepted February 26, 2019

Abstract: Studies of variation in continuous speech converge towards the conclusion that in everyday speech, words are often produced with reduced variants: some segments are shortened or completely absent. We describe an initiative to automatically exploit spoken corpora, in order to better understand linguistic behavior in spontaneous speech. This study focuses on the reduction of the postposed definite article in Romanian. The Romanian corpora used here cover several speaking styles including both prepared and spontaneous speech, such as broadcast news and debates, elicited dialogues and monologues on suggested topics. Taken together, the data sets contain more than 10 hours of speech produced in a variety of communicative frameworks. The deletion of the definite article *-l*, i.e. L-dropping in continuous speech, is investigated across speaking styles using pronunciation variants aligned with the speech. The main question addressed in the study is the influence of speaking style on the distribution of L-dropping. We examine the role of the surrounding context in L-dropping and L-retention. The results show that, in prepared speech and broadcast news, deletion is triggered by the context as a consequence of the communicative framework and the nature of the following segment (following vowel-initial words favor L-retention, while consonant-initial words favor L-dropping). In spontaneous speech, L-dropping is more frequent and the context is less important in predicting the occurrence of reduced variants than in other speaking styles.

Keywords: Romanian; spontenous speech; phonetic variation; morpho-phonology; definite article; ASR

1 Introduction

Studies of variation in continuous speech converge towards the conclusion that in everyday speech, words are often produced with reduced variants, in which some segments are shortened or completely absent. Reduction is the result of various factors, including speech rate, speaking style, lexical frequency, contextual settings, and the morphological properties (part of speech, morphological structure) of the word within the lexicon of a language. Reduction has been described as a gradient phenomenon (Ernestus 2011). However, some processes are depicted both as categorical and gradient, e.g. schwa deletion in French (Bürki et al. 2005; Bürki et al. 2011) or *t,d* deletion in English (Guy 1980; Temple and Tagliamonte 2005). It is commonly accepted that frequent content words, which are highly predictable from their context are likely to be reduced when produced in spontaneous communication (van Bergem 1993; Adda-Decker 2006). The frequency of the lemma and the relative ease of a word's morphological decomposability seem to be involved in the reduction

***Corresponding author: Ioana Vasilescu,** LIMSI CNRS, Paris Saclay University, Orsay, France,
E-mail: ioana.vasilescu@limsi.fr
Ioana Chitoran: Denis Diderot University, Paris, France
Bianca Vieru: Vocapia Research, Orsay, France
Martine Adda-Decker and Maria Candea: Sorbonne Nouvelle University, Paris, France
Lori Lamel: LIMSI CNRS, Paris Saclay University, Orsay, France
Oana Niculescu: University of Bucharest, Bucharest, Romania

processes which affect morphological affixes (Hanique and Ernestus 2011; Ernestus et al. 2015). These findings appear to be language independent, yet for the time being they rely mainly on the study of a few European languages for which there is a long tradition in both corpus linguistics and speech technology, e.g. French (Adda-Decker 2006), English (Bell et al. 2003), and Dutch (Ernestus 2000).

Analyzing the manner in which reduction occurs can simultaneously benefit linguistics (phonetics and phonology) and speech technology, specifically automatic speech recognition (ASR). ASR research relies on large volume corpora collected in real-life conditions. The integration of such data in linguistic analyses may improve models of speech production and perception with information from non-laboratory speech (Ernestus 2011). In return, ASR can be improved by adding newly documented pronunciation variants (Adda-Decker and Lamel 1999; Bell et al. 2003; Goldwater et al. 2010). Within this common research framework, ASR systems are increasingly used to explore variation in speech data. As data processing tools, ASR systems enable the location of segments in the speech flow through the process called forced alignment. They are also useful to classify word tokens according to their pronunciation variants. Such pronunciation variants are stored in a lexicon which contains both each word's full pronunciation and potentially reduced variants. Generally a probability is also associated with each variant (Lamel and Gauvain 2009). The system will select the most probable variant given the actual acoustic realization. Although they operate categorically and propose only predefined variants, ASR systems offer an alternative method to human perception, which is known to compensate for the missing acoustic information with other available cues (i.e. speech rate, context, and word length) (Mitterer 2011).

This study focuses on Romanian, described in speech technology as a less-resourced or under-resourced language (Trandabăţ et al. 2012). Among the few studies with a speech technology focus, Vasilescu et al. (2014) describe a speech recognition system dedicated to continuous speech in Romanian. The system was used to conduct linguistic investigations based on forced alignment of manual transcriptions with the audio via a lexicon including pronunciation variants. The system allowed for the quantification of the phonemic contrast of Romanian central vowels [ʌ] and [ɨ], through acoustic analysis and the contextual distribution of variants (Renwick et al. 2016).

Studies on reduction and more generally variation phenomena based on continuous speech and 'real-life' corpora are still rare in Romanian. Among the few topics explored, the deletion of the masculine singular marker of the definite article -*l*, henceforth L-dropping (e.g. pom-pomu*l* → pom*u*, 'tree-the tree') in continuous speech has received increasing attention in recent years, e.g. Miret (2017). Most studies are based on field surveys and discuss the potential causes of reduction (Avram 2009). To the best of our knowledge, the only case study which provides a different approach relies on a speech recognition system used to produce a forced alignment of different pronunciation variants involving the deletion of the definite article -*l* on a subset of the training corpus (Vasilescu et al. 2014).

The aim of this study is to examine the impact of speaking style on the distribution of full and reduced pronunciation variants of the definite article -*l*, and to analyze the role of the surrounding context in L-dropping and L-retention.

The data discussed here come from a collaborative industry research and innovation program set in a multilingual context[1]; from efforts within the French research program LabEx EFL (Laboratoire d'Excellence 'Empirical Foundations of Linguistics') to study variation in spoken languages; and also from data elicited for analysis for the PhD dissertation of the seventh author (Niculescu 2018). The data are all manually orthographically transcribed and homogeneously processed by an ASR system. Taken together, these data provide more than 10 hours of speech produced in various communicative frameworks, with up to 100 speakers.

L-dropping in continuous speech is investigated across speaking styles using the aligned speech allowing for pronunciation variants. Two specific questions are addressed: what is the frequency of the L-dropping phenomenon in continuous speech as a function of the communicative setting? And what is the role of right-word context in favoring L-dropping or L-retention?

1 www.quaero.org

The following three hypotheses are tested on the studied corpora:

1. At the current stage of the Romanian language, L-dropping is a reduction phenomenon, characteristic of an informal speaking style (Miret 2017). This hypothesis is supported if we find a significantly higher rate of deletion in our least formal corpora, compared to the more formal ones.
2. L-dropping results from reanalysis. If the rate of deletion is comparable across corpora, regardless of degree of formality, then it cannot be reliably attributed to a difference in style. Instead, such a result would lend more plausibility to the proposal by Avram (2009) that L-dropping is indicative of reanalysis, whereby the function of the masculine definite article has been transferred from -*l* to the preceding desinence vowel -*u*-.
3. L-dropping is primarily a contextual reduction phenomenon, affected by its position in the word and the right-edge segmental context (Avram 2009). This hypothesis is supported if the deletion rate of the article is significantly higher when it immediately precedes a consonant-initial word than a vowel-initial word.

The remainder of this study is organized as follows. In the next section (Section 2) we provide a description of L-dropping based on previous linguistic studies for the Romanian language. Section 3 describes the corpora, the ASR system, and the methodology. Section 4 presents the results and is followed by a discussion in Section 5.

2 L-dropping in Romanian

In Romanian, definite articles are unstressed affixes postposed to nouns:[2] (1) fem. sing. *floare-floare**a*** 'flower-the flower'; (2) masc. sing. *pom-pomu**l*** 'tree-the tree'; (3) fem. pl. *flori-flori**le*** 'flowers-the flowers'; (4) masc.pl. *pomi-pomi**i*** 'trees-the trees'. This study focuses on case (2), where it has been commonly observed that final -*l* is often deleted in continuous speech resulting for exeample in *pomu* as the definite form of *pom* 'tree'. The deletion is interpreted as a consequence of reanalysis by native speakers: the meaning of definiteness is transferred to the desinence vowel -*u*- as the -*l* is increasingly deleted (Avram 2009).

The acquisition of the lateral final definite article is reinforced probably in the first years of primary school, when stem+-*(u)l* pronunciation is emphasized. Native speakers are aware that in everyday speech the definite marker in masculine nouns -*l* is absent, even if prescriptive grammars recommend the use of final -*l* forms. Historically, Romanian -*l* comes from a Latin affix derived as in other Romance languages, from the personal pronoun *illu(m)* (Avram 2009). Alternative forms with and without -*l* are attested in the Middle Ages, as shown by fluctuating orthography. The oldest record of written Romanian dates back to the 16th century (Toma 1984). In this text -*l* occurs as a definite article in nominative-accusative cases for toponymic and patronymic names (e.g. patronym *Neacşu* → *Neacşu**l***). According to Stan (2013) the usage of -*l* as a definite article is likely to be older, although writings in Romanian are not available before the 16th century. The fluctuation between written forms with and without final -*l* lasts until the 19th century when both orthographies are attested and accepted (Chivu et al. 2012). During the 20th century the norm only mentions written -*l* although it is not necessarily pronounced -*l*, and field surveys quantify the phenomenon throughout the country. They reveal a tendency toward L-deletion in the Southern regions (e.g. Muntenia, Dobrogea), whereas Northern parts of Romania (Maramureş, but also Romanian Moldavia and the Republic of Moldova) show a stronger tendency for its maintenance (Rusu 1984; Barbu-Mititelu et al. 2012). Nowadays, modern dictionaries specify that although in the written form -*l* is mandatory, the lateral is pronounced only in formal speech (DOOM2 2005).

Linguistic studies on L-dropping date from the early 20th century (Lombard 1935; Avram 2009). Findings are that L-dropping is frequent in everyday speech, and that the phenomenon does not occur randomly, its

2 Adjectives can carry the definite marker when they precede the noun in a marked word order (*pomu**l** frumos* vs. *frumosu**l** pom*, 'tree-def. beautiful' vs. 'beautiful-def. tree', 'the beautiful tree'), as well as ordinal numbers (*întîiu**l*** 'the first') and undefined pronouns (*unu**l*** 'one', *altu**l*** 'the other one'). Taken together, such contrastive instances are underrepresented in our speech corpora (10%), and we will therefore not discuss them.

deletion being triggered by its immediate right-edge context. For instance, the deletion is more likely to occur before a consonant-initial word (e.g. *o.mul bun* > *o.mu bun* 'the good person'), than before a vowel-initial word which leads to re-syllabification and enables L-retention (e.g. *o.mu**l** a.ce.la* > *o.mu.**l**a.ce.la* 'that person'). Such claims have been supported by results of field surveys. Most recently, Miret (2017) showed that L-dropping is more likely to occur in spontaneous than in read speech.

With regard to the acoustic properties of the definite marker -*l* and of its conditions of occurrence, Romanian /l/ is reported by Marin and Pouplier (2014) to be non-velarized (i.e. clear), based on articulatory (EMA) data analysis. In a cross-language study of the acoustic realizations of /l/ in 23 languages, Recasens (2012) also reports a clear realization of the Romanian /l/. However, the two studies do not specifically address the liquid in the definite article, which always follows the back vowel /u/. It is therefore possible that a certain amount of coarticulation may be present in this specific context, resulting in a tendency for a more velarized -*l*.

A different approach is adopted by Vasilescu et al. (2014), where an ASR system is used to align variants with and without final -*l*. The experiment conducted for that study on a corpus of 3.5 hours of broadcast news shows that on average 20% of -*l* are deleted and that the percentage increases with more spontaneous recordings (debates vs read or prepared news). Chitoran et al. (2014) addressed the acoustic properties of the definite article and suggested that a reduced -*l* with more vowel-like formant structure is less likely to be aligned as a -*l* variant by the ASR system.

The study presented in the following sections builds on the approach described in Vasilescu et al. (2014) by extending the corpora and refining the quantitative analysis.

3 Corpora and methodology

3.1 Corpora

Our data are comprised of different speaking styles that cover a range of degrees of formality: semi-prepared broadcast news known to comprise carefully articulated speech, broadcast debates including radio debates and spontaneous interactions, spontaneous dialogues based on Diapix elicitation (Baker and Hazan 2011), informal reading and free monologues. The spontaneous dialogues and the read corpora were recorded simultaneously in Bucharest by M. Candea, and they involve the same speakers. The free monologues corpus was recorded by O. Niculescu as part of her PhD project (Niculescu 2017, 2018). It is a subset of a larger corpus recorded in Bucharest comprising 9 speakers (4 female, 5 male). Data from 6 of the speakers were manually transcribed by O. Niculescu, and the data from one of them was used in the current experiment. Broadcast data (semi-prepared and debates) come from the radio stations RFI Journal, RRA - Radio România Actualităţi, the news agency Euranet and from the TV station Antena 3. They are designed for a broad audience and are representative of standard Romanian. All speakers are native speakers of the Southern dialect described as standard Romanian.

The broadcast news and the spontaneous dialogues corpora are expected to be representative of two speaking styles: semi-prepared and spontaneous. The reading and monologues corpora serve as controls. The corpus of read material was recorded for comparison with the spontaneous dialogues corpus. Since the same speakers are recorded for both, the comparison serves to establish the degree of spontaneity of the latter corpus. The monologues corpus was selected to estimate the degree of spontaneity of informally elicited monologues and as an item for comparison with the broadcast debates and spontaneous dialogues. Each corpus was manually orthographically transcribed by a native speaker of Romanian and double checked by two co-authors, both native speakers of Romanian (I. Vasilescu and B. Vieru).

Table 1 summarizes the corpora characteristics. The first two columns correspond to total duration and number of speakers per corpus, the third column gives a rough measure of the articulation rate computed as number of aligned phonemes per second after excluding filled and empty pauses. The articulation rate measure is used to verify the adequacy of the speaking style label assigned to each corpus. The measure confirms the relatively fast articulation rate of the broadcast speech compared to spontaneous settings

Table 1: Summary of data sources: total duration, number of speakers (number of female speakers in parentheses) and articulation rate (phon./sec.).

Corpus	Duration (h)	Nbr. of speakers	Phon./sec.
Semi-prepared BN	3.5	79 (29)	15.4
Broadcast debates	3.5	48 (18)	14.3
Spont. dialogues	3.0	29 (20)	12.5
Reading	0.5	29 (20)	13.7
Free monologues	0.5	1 (0)	13.7

Table 2: Word token distribution in corpora: types and total word tokens, consonant (#C) and vowel-initial (#V), raw numbers and percentages.

Corpus	Types	Tokens	#C	#V
Semi-prepared BN	6743	32k	22832 (71%)	9455 (29%)
Broadcast debates	6025	26k	18857 (72%)	7506 (28%)
Spont. dialogues	2154	27k	19138 (73%)	7572 (27%)
Reading	207	5k	4365 (80%)	1111 (20%)
Free monologues	1313	4k	3229 (73%)	1215 (27%)

(Trouvain 2004), and suggests the following formality hierarchy in decreasing order: semi-prepared broadcast news, broadcast debates, spontaneous dialogues. The reading and free monologues corpora have comparable speech rates and fall in the middle suggesting that both are produced rather informally.

Table 2 presents the lexical characteristics of the corpora: the number of types (distinct word tokens) per corpus, the number of word tokens per corpus, the number and percentage of consonant-initial words and vowel-initial words.

The data are classed in the following speaking styles:

- **Semi-prepared broadcast news (BN)**: broadcast data gathered from several Romanian radio and television shows (3.5 hours, 79 different speakers) within the Quaero project.[3] It consists mainly of read and semi-prepared news. The number of speakers varies according to the broadcast channel, ranging from 3 to 24. Broadcasts with significant quantities of overlapping speech, foreign or regional accents, and noisy background were not included.
- **Broadcast debates**: televised debates recorded from the Romanian national channel Antena 3 (3.5 hours, 50 speakers).
- **Spontaneous dialogues**: 38 short dialogues recorded using the method adopted by Baker and Hazan (2011). The method called Diapix is a dialogue elicitation technique that involves pairs of participants cooperatively completing a linguistic task. The Diapix pictures from Baker and Hazan (2011) were adapted by M. Candea to serve in several studies of Romanian. The same 29 speakers as for the reading protocol participated in the experiment.
- **Reading**: a short text of 207 words, neutral in terms of topic (a short story about a couple going to a restaurant and selecting a Romanian menu), read by 29 speakers (9 male and 20 female), aged between 27-68 years. The read corpus was recorded from the same speakers as the spontaneous dialogues. The style is therefore informal in both. The participants were encouraged to read the text at a normal speed. The duration of the entire corpus is 0.5 hours.
- **Free monologues**: This corpus is extracted from a larger one, as described at the beginning of the section. It contains speech from one male speaker, totalling 0.5 hours and containing stories about his childhood and recent holidays.

3 www.quaero.org

3.2 Forced speech alignment

The ASR system described in Vasilescu et al. (2014) and Renwick et al. (2016) is used in forced alignment mode in this study. Since the manual transcription is provided *a priori*, the system only has to select the best matching pronunciation for each word, and the corresponding locations of word and phone boundaries. Due to acoustic modeling constraints, phone segments have a minimum duration of 30 ms (3 frames).

The system takes as input the speech files and their associated manual orthographic transcriptions. These transcriptions also are split into speech segments roughly corresponding to speaker turns. Each word in the transcript is decomposed into phonemes using the pronunciation lexicon that contains all the words present in the transcriptions. Acoustic features are extracted from the audio, and using context-independent acoustic models, each phoneme is aligned with an audio segment. All data are processed using these segments and, for each segment, the corresponding transcripts are aligned with the audio providing time codes at the word and phone levels. The pronunciation lexicon and acoustic models are based on a set of 29 phones, including 20 consonants, 7 vowels and 2 glides. The phonemic diphthongs /o͡a/ and /e͡a/ are encoded as sequences of vowels. The phone sets also includes a special symbol for silence as shown in Table 3. The acoustic models are trained in a semi-supervised manner using approximately 400 hours of unannotated audio (Vasilescu et al. 2014).

3.3 Pronunciation variants

The purpose of this study is to estimate the rate and characteristics of the contexts which trigger L-dropping in Romanian through an objective analysis, as earlier studies are mostly based on listening and on impressionistic decisions of L-dropping or retention. The experimentation described here is based on the selection of pronunciation variants defined in the lexicon and chosen during the alignment process by the ASR system. Prior to this, the system was trained on more than 400h of audio data from different web radio sources. During this phase the system built *-(u)l* models based on many full and weakened realizations of *-l*. Consequently, the high acoustic variability of *-l* does not affect the speech recognition process. However, as word-final *-l* is always orthographically present during the acoustic model training, there is a bias in favor of L-retention. The consequence is a potential overestimation of L-retention variants in the current study that will need further experimentation.

Table 3: Phones used in the Romanian ASR system.

IPA	Ex. Romanian	IPA	Ex. Romanian
p	**p**as	b	**b**an
t	**t**are	d	**d**acă
k	**c**al	g	**g**ol
m	**m**ic	n	**n**or
f	**f**oc	v	**v**al
s	**s**are	z	**z**id
h	**h**orn	t͡s	**ț**ara
r	**r**epede	l	**l**ung
ʃ	**ș**arpe	ʒ	**j**ar
tʃ	**c**er	dʒ	**g**er
a	**a**pa	e	**e**rou
i	**i**nsula	o	**o**ra
u	**u**da	ə	udă
ɨ	înspre		
o͡a	f**oa**rte	j	**ia**pa
e͡a	m**ea**	w	dau
-	silence	-	breath
-	filler		

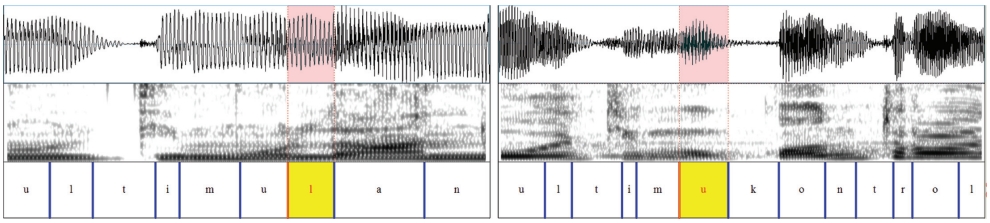

Figure 1: Examples of L-retention (left) vs L-dropping (right) as a function of the right-edge context: #V **ultimul an** [ultimul#an] 'the last year', #C **ultimu∅ control** [ultimu#kontrol] 'the last control'.

Table 4: Masculine nouns with graphemic -*l* marker present in the data classed by right-edge context (total, followed by C-initial words (#C), by V-initial words (#V), by silence (#sil)).

Corpus	#C	#V	#sil	Total
Semi-prepared BN	830	269	29	1128
Broadcast debates	512	139	30	681
Spont. dialogues	217	104	30	351
Reading	74	0	11	85
Free monologues	44	22	9	75
Tokens	1677	534	109	2320
%	72%	23%	4%	100%

In a pilot study (Chitoran et al. 2014), two phoneticians who are native speakers of Romanian inspected 824 contexts extracted from a subset of the two broadcast news corpora for which the ASR system described by Vasilescu et al. (2014) aligned variants with and without -*l*. These contexts consist of masculine nouns at the center of three word sequences. The phoneticians listened to the audio and examined the corresponding spectrograms. The findings showed false positives in favor of L-retention for 33% of the items before a C-initial word and 13% before a V-initial word. For these words -*l* was acoustically and auditorily undetected by the trained phoneticians, but it was detected by the system. False negatives were also present. A pronounced -*l* missed by the system was acoustically observed and auditorily detected in 15.9% of the cases before a C-initial word and 4.1% before a V-initial word.

In the present study, the system is allowed to align /l/ or ∅ (nothing) for lexical -*l*. As underlined by Chitoran et al. (2015), L-dropping can be the result of segment deletion, or of phonetic undershoot. Figure 1 shows two sequences of Romanian words aligned by the ASR system with an -*l* in *ultimul an* ([ultimul#an] 'the last year'), and without -*l* in *ultimu∅control* ([ultimu#kontrol] 'the last control').

The selection of words used in the experiment presented in this study is based on the orthographic -*ul* ending and consists of 2320 occurrences. Among the selected words, 90% are nouns, the remaining 10% of items corresponding to pronouns (*unu**l*** 'one', *altu**l*** 'the other one') and adjectives (*Sfîntu**l** Spiridon* 'Saint Spiridon'), which take the first position in the nominal group and carry the definite marker.

The words meeting the criteria above were extracted from all corpora, and separated according to right-edge context (consonant vs vowel initial word). Table 4 shows the contexts in which the graphemic -*l* is followed by a consonant-initial word (#C), by a vowel-initial word (#V), or by silence (#sil) as detected by the ASR system.

As seen in Table 2, the C-initial context is the most frequent in all corpora. Table 4 also highlights that the C-initial context is the most frequent right-edge context of the words displaying an -*(u)l* ending. It shows that C-initial words occur on average 2.5 times more often than V-initial words.

4 Results

First, the incidence of L-dropping was estimated across all corpora without considering the following context. Figure 2 shows L-dropping vs L-retention rates as a function of speaking style.

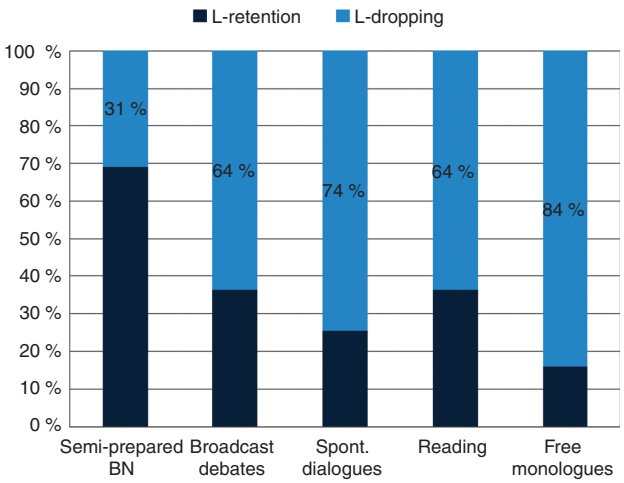

Figure 2: Distribution of L-dropping vs L-retention pronunciation variants across speaking styles.

The overall comparison revealed that the distribution of variants with final *-l* dropped vs retained is not the same across the different corpora ($\chi^2(4, N = 2320) = 341.44, p < .001$) and that deletion rates increase with the degree of spontaneity of the data. There are nearly 2.5 times more deletions in debates and dialogues than in prepared broadcast news. The reading datasets pattern identically with the broadcast debates, and both show around 10% less deletion rates than the spontaneous dialogues. The result is consistent with the articulation rate computed for all datasets and reported in Table 1. It is also somewhat surprising for the reading corpus: we would have expected to obtain deletion rates similar to the ones observed for prepared broadcast, that partly contain read news. This result suggests that the criterion "speaking style" needs to be refined, because in our data the communication condition (informal reading vs formal news, delivered by professional speakers), and the social status and distance between the speakers seem to be more important than the opposition: read vs spontaneous speech. However, the fact that the corpus of monologues has the highest *-l* deletion rate (84%) should be considered with caution, since this corpus consists of data from only one speaker. Therefore, the high deletion rate cannot be reliably attributed to the communicative setting, since it could very well reflect pronunciation habits of this individual speaker. At this time, we treat this result as anecdotal evidence that the communicative context of free monologues may be highly informal, and *-l* can be often deleted. Clearly, the additional data collected by O. Niculescu need to be analyzed in order to draw firm conclusions.

The deletion rate was then examined according to the following context. Previous field surveys or small-scale studies suggest that the context and the speaking style are the main factors which trigger deletion (Lombard 1935; Miret 2017). L-dropping is more likely to occur before a C-initial than a V-initial context as highlighted by Avram (2009). This pattern has been linked to a more general trend to simplify consonantal clusters at word boundaries (Chitoran et al. 2015). In the case of V-initial words, L-dropping might lead to hiatus: a relationship may thus be established between the lower L-dropping rates before V-initial words and hiatus resolution at word boundaries (Niculescu 2018). The comparison of L-dropping vs L-retention as a function of the following context considered all corpora together. It shows that the difference in L-dropping across C-initial and V-initial words is significant ($\chi^2(2, N = 2320) = 17.44, p < .001$). Figure 3 shows that the rate of L-dropping is significantly higher before a C-initial context (53%) compared to V-initial (42%) and pre-pausal (48%) contexts.

Next we aimed to provide a more detailed account of the possible interaction between speaking style and context. Figure 4 shows the L-dropping rates as a function of the following context and displays the rates for broadcast prepared speech, broadcast debates and spontaneous dialogues, for which we have the highest volume of data. These rates highlight the interaction between the context (#C and #V) and the speaking style (more formal vs less formal). The deletion rates are identical in both #C and #V contexts for spontaneous dialogues, showing that in less formal settings, C-initial and V-initial words equally trigger deletion. In more

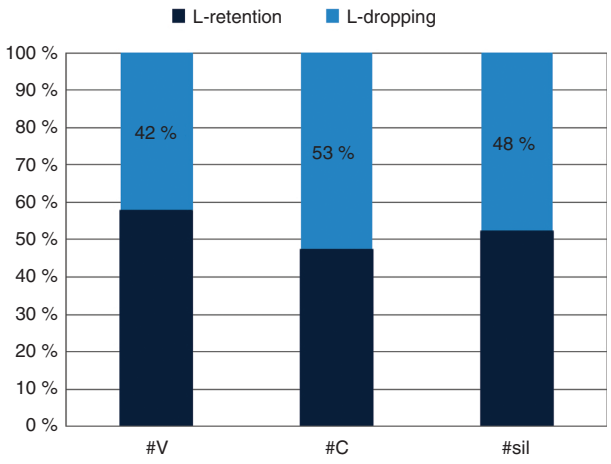

Figure 3: Distribution of L-dropping vs L-retention pronunciation variants as a function of the right-edge context.

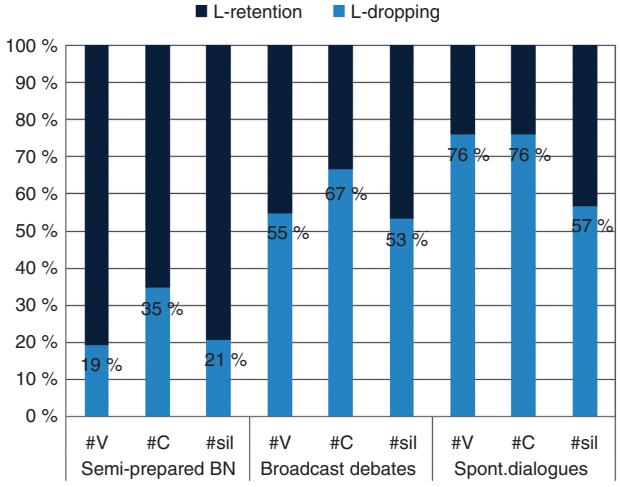

Figure 4: Effect of the right-edge context on L-dropping as a function of the speaking style.

formal contexts such as prepared broadcast news, the factor "context" seems to predominate and the deletion rates are higher before C-initial words.

The differences observed across corpora as a function of the right-edge context motivated a series of additional comparisons to estimate the contribution of each subset of data to the global result. The reading and monologues corpora were excluded from the comparisons for the following reasons: as shown in Table 4, in the reading corpus *-l* is never followed by a vowel, and the monologues corpus does not allow for a reliable comparison because it contains one single speaker.

We then considered the selection of variants with dropping vs retention of final *-l* in different corpora as a function of the right-edge context. A first comparison considered the least spontaneous data, that is the semi-prepared broadcast news whose distribution of L-dropping vs L-retention variants differs from the other data (see Figures 2 and 4). The results support the first hypothesis, showing that in semi-prepared settings, speakers tend to delete the final *-l* to a lesser extent than in informal settings, especially if the following word starts with a vowel. In particular, compared to broadcast debates, dropping is less frequent regardless of the right-edge context ($\chi^2(2, N = 1128) = 24.29, p < .001$).

The pattern is less clear in spontaneous speech, where *-l* deleted variants are most often selected by the system, and to a comparable extent, before both C-initial and V-initial words. In broadcast debates, although the selection of variants does not show a clear trend regarding the following V-initial context, *-l* is more often absent before C-initial words than before V-initial words ($\chi^2(2, N = 681) = 8.13, p < .05$). These findings

suggest that in formal communication contexts such as broadcast prepared speech, L-dropping is limited by the influence of the prepared written material and the social pressure of a prescriptive standard Romanian pronunciation. Despite the high articulation rate measured for this corpus (15.5 phonemes/second), the L-retention rate is also the highest (69%), suggesting that L-dropping cannot be attributed primarily to increased speaking rate. This pressure decreases or disappears in more casual settings, when the speakers discuss freely as in the broadcast debates, the spontaneous dialogues or the free monologues, or when a written support is provided but the context is informal, as suggested by the results for the read corpus. This finding is consistent with previous results reported for Romanian and other languages. Previous analyses of Romanian vowels conducted by Renwick et al. (2016) revealed differences in the acoustic patterns of vowels as a function of the speaking style, and suggested that the communicative setting is at least as important as the local phonetic context. Similarly, in an analysis of the impact of the speaking style on the selection of reduced pronunciation variants in French and English, Adda-Decker and Snoeren (2011) showed that the amount of shortened and potentially reduced segments aligned by ASR systems increases in spontaneous settings.

5 Discussion

This study aimed to describe and quantify the factors that influence the phenomenon of definite marker deletion (L-dropping) in masculine nouns in contemporary Romanian.

We used a collection of corpora containing more than 10 hours of speech to extract occurrences of masculine nouns+-*l* and of other parts of speech with similar behavior (adjectives and undefined pronouns, in initial position in the nominal groups). Such instances were analyzed with an ASR system designed to process continuous speech and parametrized to optionally align pronunciation variants with and without final -*l*. The results were interpreted with respect to three main hypotheses: (1) -*l* is a reduction phenomenon characteristic of spontaneous speech, (2) -*l* is the result of a morphological reanalysis regardless of the speaking style, resulting in the transfer of the function of -*l* as definite marker to the desinence vowel -*u*-, or (3) -*l* is a reduction phenomenon sensitive to the right-edge context.

Our results confirm the first hypothesis and are consistent with previous linguistic assessments showing that the communicative setting strongly influences speech deletion rates (Avram 2009; Miret 2017).

These results also confirm the relevance of the phonetic context (third hypothesis): -*l* is more likely to be dropped before C-initial words in the case of prepared speech (semi-prepared broadcast news). The second hypothesis (Avram 2009), concerning the transfer of the grammatical function of the definite marker to the desinence vowel, is supported by one particular finding of our study: in spontaneous dialogues, where -*l* deleted variants are most frequently selected by the system, they are selected to an equal extent (76%) before both C-initial and V-initial words (see Figure 4). Based on this result, we will not reject the hypothesis that, at least in certain conditions (e.g. unconstrained by spelling) speakers' morphological planning may shift from -*l* to -*u*- with respect to definiteness. This question goes beyond the scope of the present study, but we have shown that it is worth pursuing. In light of our results, we conclude that the notion of 'speaking style' itself needs to be revisited: L-dropping is less favored by a V-initial right-edge context for speech produced in a formal setting, with a written or at least prepared support, as in semi-prepared broadcast news. However, the presence of a written support seems less important in speech read in an informal context. The high percentage of L-dropping in this corpus is more reliably attributed to the casual character of the communicative setting. Further work will focus on testing other possible variants, and in particular the variant /w/. As mentioned earlier, L-dropping before a V-initial word may lead to hiatus, which in turn may be resolved in one of two ways: by retaining or dropping -*l*, or by inserting a glide (/u/ to /w/). At the same time, a high occurrence rate of hiatus by retention of /u/ even before a vowel-initial word (e.g. *om[**u**] acela* 'that person'), could be interpreted as further support for the reanalysis of the desinence vowel as a definite marker.

Further studies will also focus on the acoustic and prosodic characteristics influencing L-dropping. Finally, we will conduct a wider survey of phenomena lying at the interface between phonetics and

morpho-phonology, which may be affecting word final affixes in Romanian, and may provide evidence about the evolution of the language.

Acknowledgment: This work was partially supported by the French Investissements d'Avenir - Labex EFL program (ANR-10-LABX-0083) through a mobility grant awarded to O. Niculescu, and by the Maison des Sciences de l'Homme Paris-Saclay through an 'Emergence' grant awarded to I. Vasilescu. We are also grateful to the editors and two anonymous reviewers for their comments, which improved the quality of the paper.

References

Adda-Decker, M. 2006. De la reconnaissance automatique de la parole à l'analyse linguistique des corpus oraux. Paper presented at Journées d'Étude sur la Parole (JEP 2006), Dinard, France, 12–16 June.

Adda-Decker, M. & L. Lamel. 1999. Pronunciation variants across system configuration, language and speaking style. *Speech Communication - Special issue on modeling pronunciation variation for automatic speech recognition* 29(2–4). 83–98.

Adda-Decker, M. & N. Snoeren. 2011. Quantifying temporal speech reduction in French using forced speech alignment. *Journal of Phonetics* 39(3). 261–270.

Avram, A. 2009. Despre cauzele dispariţiei lui l final - articol hotarât [Concerning the disappearance of final l]. In *Probleme de fonologie a limbii române [Topics regarding the Romanian phonology]*, 171–176. Editura Academiei Române, Bucharest.

Baker, R. & V. Hazan. 2011. DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods* 43(3). 761–770.

Barbu-Mititelu, V., M.-M. Morcov, M. Nevaci, C.-I. Radu, D. Răuţu, & D.-M. Zamfir. 2012. *Atlasul lingvistic român pe regiuni. Sinteză. [Romanian linguistic atlas by regions. Synthesis] Vol. I şi II, coord. Nicolae Saramandu*. Editura Academiei Române, Bucharest.

Bell, A., D. Jurafsky, E. Fosler-Lussier, C. Girand & D. Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113(2). 1001–1024.

Bürki, A., M. Ernestus & U. Frauenfelder. 2005. 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, A., C. Fougeron, C. Gendrot & U. Frauenfelder. 2011. Phonetic reduction versus phonological deletion of French schwa: Some methodological issues. *Journal of Phonetics* 39(3). 279–288.

Chitoran, I., J. Hualde & O. E. Niculescu. 2015. Gestural undershoot and gestural intrusion – from perception errors to historical sound change. In G. Adda, V. B. Mititelu, J. Mariani, D. Tufis & I. Vasilescu (eds.), *Errors by Humans and Machines in Multimedia, Multimodal and Multilingual Data Processing*, 175–192. Editura Academiei Române, Bucharest.

Chitoran, I., I. Vasilescu, B. Vieru & L. Lamel. 2014. Analyzing linguistic variation in a Romanian speech corpus through ASR errors. Paper presented at Laboratory Approaches to Romance Phonology VII (LARP VII), Aix en Provence, France, 3–5 September.

Chivu, G., G. P. Dindelegan, A. Dragomirescu, I. Nedelcu & I. Nicula. 2012. *Studii de istorie a limbii române. Morfosintaxa limbii literare în secolele al XIX-lea şi al XX-lea [Studies about the history of the Romanian language. The morphosyntax of literary Romanian in the 19th and 20th centuries]*. Editura Academiei Române, Bucharest.

DOOM2. 2005. *DOOM 2 – Dicţionarul ortografic, ortoepic şi morfologic al limbii române, Ediţia a II-a revăzută şi adăugită [Orthographic, orthoepic and morphological dictionary of the Romanian language. Revised and extended edition]*. Univers Enciclopedic, Institutul de Lingvistică 'Iorgu Iordan', Bucharest.

Ernestus, M. 2000. *Voice assimilation and segment reduction in casual Dutch. A corpus-based study of the phonology-phonetics interface*. LOT, Utrecht.

Ernestus, M. 2011. An introduction to reduced pronunciation variants. *Journal of Phonetics* 39(3). 253–260.

Ernestus, M., I. Hanique & E. Verboom. 2015. The effect of speech situation on the occurrence of reduced word pronunciation variants. *Journal of Phonetics* 48. 60–75.

Goldwater, S., D. Jurafsky & C. D. Manning. 2010. Which words are hard to recognize? Prosodic, lexical, and disfluency factors that increase speech recognition error rates. *Speech Communication* 52(1). 181–200.

Guy, G. 1980. Variation in the group and the individual: The case of final stop deletion. In W. Labov, (ed.), *Locating language in time and space*, 1–36. New York: Academic Press, Inc.

Hanique, I. & M. Ernestus. 2011. Final /t/ reduction in Dutch past-participles: The role of word predictability and morphological decomposability. In *the 12th Annual Conference of the International Speech Communication Association (INTERSPEECH 2011), Florence, Italy, August 27–31*, 2849–2852.

Lamel, L. & J.-L. Gauvain. 2009. Speech recognition. In R. Mitkov (ed.), *The Oxford handbook on computational linguistics*, 305–322. New York: Oxford University Press Inc.

Lombard, A. 1935. *La prononciation du roumain*. Uppsala: Lundequistska Bokhandeln.

Marin, S. & M. Pouplier. 2014. Articulatory synergies in the temporal organization of liquid clusters in Romanian. *Journal of Phonetics* 42. 24–36.

Miret, F. S. 2017. Elision de la /l/ del articulo en dos hablantes de rumano. In A. Dragomirescu, A. Nicolae, C. Stan, & R. Zafiu (eds.), *Sintaxa ca mod de a fi [Syntax as way of being]*, 361–370. Bucharest: Editura Universitaţii din Bucureşti.

Mitterer, H. 2011. Recognizing reduced forms: Different processing mechanisms for similar reductions. *Journal of Phonetics* 39(3). 298–303.

Niculescu, O. E. 2017. Analiza structurilor vocalice de tip hiat din limba română comtemporană. O anchetă lingvistică [Analyzing vowel sequences in contemporary standard Romanian. A linguistic inquiry]. In R. Zafiu, A. Dragomirescu & A. Nicolae, (eds.), *Diacronie şi sincronie în studiul limbii române. Actele celui de al 13-lea Colocviu internaţional al Departamentului de Lingvistică [Diachrony and synchrony in the study of the Romanian language. Proceedings of the 13th International Conference of the Linguistic Department]*, 229–241. Bucharest: Editura Universităţii din Bucureşti, CNCS, categoria B.

Niculescu, O. E. 2018. *Hiatul intern şi hiatul extern în limba română: O analiza acustică [Internal and external hiatus in contemporary standard romanian. An acoustic analysis]*. Bucharest: University of Bucharest.

Recasens, D. 2012. A cross-language acoustic study of initial and final allophones of /l/. *Speech Communication* 54(3). 368–383.

Renwick, M., I. Vasilescu, C. Dutrey, L. Lamel & B. Vieru. 2016. Marginal contrast among Romanian vowels: Evidence from ASR and functional load. In *The Annual Conference of the International Speech Communication Association (INTERSPEECH 2016), San Francisco, United States, 8–12 September*, 2433–2436.

Rusu, V. 1984. *Tratat de dialectologie românească [Romanian dialectology studies]*. Craiova: Scrisul Românesc.

Stan, C. 2013. *O sintaxă diacronică a limbii române vechi [A diachronic syntax of old Romanian]*. Bucharest: Editura Universităţii din Bucureşti.

Temple, R. & S. Tagliamonte. 2005. New perspectives on an ol' variable: (t,d) in British English. *Language Variation and Change* 17(3). 281–302.

Toma, S. 1984. *Scrisoarea lui Neacsul de la Câmpulung - Crestomaţie de literatură română veche [Neacsul from Câmpulung's letter - Compendium of old literary Romanian]*. Cluj Napoca: Editura Dacia.

Trandabăţ, D., E. Irimia, V. B. Mititelu, D. Cristea & D. Tufiş. 2012. The Romanian language in the digital age. In G. Rehm & H. Uszkoreit, (eds.), *META-NET White Paper Studies*, 1–87. Berlin: Springer.

Trouvain, J. 2004. Tempo variation in speech production: Implications for speech synthesis. *Phonus : Reports in Phonetics* 8. 1–129.

van Bergem, D. R. 1993. Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication* 12(1). 1–23.

Vasilescu, I., B. Vieru & L. Lamel. 2014. Exploring pronunciation variants for Romanian speech-to-text transcription. In *4th Workshop on Spoken Language Technologies for Under-resourced Languages, (SLTU 2014), St. Petersburg, Russia, May 14–16*, 161–168.

Article note: The study is a team work and the different authors participated with data collection, annotation and processing, and/or with phonetic and morpho-phonological analysis of the investigated phenomenon: Ioana Vasilescu, Ioana Chitoran and Oana Niculescu have all a background in Romanian phonetics and phonology. Their contribution concerns the state of art and the analysis of the processed data. Ioana Vasilescu participated in the collection and annotation of the broadcast news data and Oana Niculescu provided the monologue corpus. Martine Adda-Decker, Lori Lamel and Bianca Vieru are computer scientists with background in linguistic analysis of large-scale speech corpora and their contributions concern the development of the speech alignment systems and the automatic processing of the data. Martine Adda-Decker supported the collection of the spontaneous dialogues and reading corpora within the Labex EFL data collection actions. Maria Candea contributed with the design and the collection of the spontaneous dialogues and reading corpora. All authors contributed to the content and scope of the written document.