UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari (DISLL)

SCUOLA DI DOTTORATO DI RICERCA IN: Scienze linguistiche, filologiche e letterarie

INDIRIZZO: Linguistica, lingue classiche e moderne

CICLO: XXVI

THE INVARIANT IN PHONOLOGY
The role of salience and predictability

Direttore della Scuola : Ch.ma Prof.ssa Rosanna Benacchio

Coordinatore d'indirizzo: Ch.mo Prof.ssa Carmen Castillo Peña

Supervisore: Ch.mo Prof. Alberto Maria Mioni

Dottorando: Antonio Baroni

Abstract.

This aim of this thesis is to give a phonological account of acoustic variation and reduction. It is argued that phonological representations are uneven and include information about the relative strength of the segmental and subsegmental units composing them. This unevenness implies a distinction between the *invariant* – the "phonetic essence" of a word, which is practically undeletable – and other units which can be dispensed with under certain circumstances. In the first chapter I compare different theoretical approaches to the problem of acoustic variation, in particular with reference to generative phonology and exemplar-based theories. In the second chapter I propose a model which combines aspects of Optimality Theory, Element Theory and usage-based linguistics. Additionally, I discuss the role of acoustic salience in the formation of the invariant. In chapter three, typological and experimental data are examined in order to establish a salience scale for consonants. In chapter four, the results of the acoustic analysis of four dialogues extracted from a corpus of spoken Italian are presented. As expected, highly salient consonants are preserved to a greater extent than less salient ones. In chapter five I attempt to identify the phonological correlates of acoustic salience and discuss other factors which may favor reduction and deletion, among which predictability. In chapter six I draw some conclusions, deal with some pending issues and suggest future directions for research.

Riassunto.

Lo scopo di questa tesi è di rendere conto della variazione e della riduzione acustica da un punto di vista fonologico. Secondo il modello che propongo, le rappresentazioni fonologiche sono disomogenee e racchiudono informazioni sulla forza relativa delle unità segmentali e subsegmentali che le compongono. Questa disomogeneità implica una distinzione tra l'*invariante*, o "essenza fonetica" di una parola, che è praticamente incancellabile, e altre unità di cui, in certe circostanze, si può fare a meno. Nel primo capitolo vengono confrontati diversi approcci teorici al problema della variazione acustica, facendo riferimento in particolare alla fonologia generativa e alla Teoria degli Esemplari. Nel secondo capitolo, oltre a proporre un modello che combina aspetti della Teoria dell'Ottimalità, della Teoria degli Elementi e della linguistica *usage-based*, si discute anche il ruolo della salienza acustica nella formazione dell'invariante. Nel terzo capitolo vengono esaminati dati tipologici e sperimentali per costruire una scala di salienza delle consonanti. Il quarto capitolo presenta i risultati dell'analisi acustica di quattro dialoghi estratti da un corpus di italiano parlato. Come previsto, le consonanti più salienti vengono conservate più frequentemente di quelle meno salienti. Nel quinto capitolo si tenta di individuare i correlati fonologici della salienza acustica e vengono discussi altri fattori che possono favorire la riduzione e la cancellazione, tra cui la predicibilità. Nel sesto capitolo si traggono alcune conclusioni, vengono trattate alcune questioni irrisolte e si suggeriscono delle future linee di ricerca.

CONTENTS

ACKNOWLEDGMENTS

First of all, I need to express my gratitude towards my supervisor, Alberto Mioni. He has, in fact, been my supervisor for more than just the three years of this PhD, since we worked together for my MA thesis as well. In all honesty, I have been extremely fortunate to work with someone with such an eclectic knowledge of all aspects of linguistic theory, such broad experience and such an incredibly open-mind. Not only is Alberto one of the most important Italian linguists, but he is also wonderfully curious about the newest theories and approaches.

I am also indebted to René Kager, who agreed to be my tutor during my stay in Utrecht as a visiting PhD student. Although René was tremendously busy, given his dedication to his students, he managed to discuss my project with me and advise me about the experiment I was preparing. René also generously offered me his own funds in order to pay the participants to the experiment. Moreover, he introduced me to all aspects of the academic life of the Utrecht Linguistic Institute and made me feel more than welcome.

Among the professors and my colleagues at the University of Padua, thanks are due in particular to Paola Benincà, Rosanna Benacchio, Maria Grazia Busà, Davide Bertocci, Jacopo Garzonio, Vania Masutti, Alessio Muro, Andrea Padovan, Diego Pescarini, Cecilia Poletto, Luca Rognoni, Marija Runić and Laura Vanelli.

Outside Padua, I had the chance to meet some very inspiring linguists who helped broaden my view on phonology and phonetics: Birgit Alber, Pier Marco Bertinetto, Hans Basbøll, Joaquim Brandão de Carvalho, Paul Boersma, Eugeniusz Cyran, Wolfgang U. Dressler, Katarzyna Dziubalska-Kołaczyk, Mirjam Ernestus, Brigitte Garcia, Bruce Hayes, Giovanna Marotta, Mark van Oostendorp, Tobias Scheer, Jolanta Szpyra-Kozłowska, Linda Shockey.

In no particular order, I want to thank the following people for exchanging ideas and/or sharing some drinks with me: Athanasia Asyllogistou, Tommaso Balsemin, Claudia Bianchini, Florian Breit, Nathalie Boll-Avetisyan, Basilio Calderone, Marta Castella, Edoardo Cavirani, Chiara Celata, Stefano Coretta, Bartlomiej Czaplicki, Radek Iwanckiewicz, Adèle Jatteau, Brigitta Keij, Yuni Kim, Tom Lentz, Pedro Martins, Sandrien van Ommen, Diana Passino, Des Ryan, Shanti Ulfsbjorninn.

I also want to thank Iris Mulders, Paola Sessa and Theo Veenker for their technical support for the experiment, Steve Parker and Guillaume Jacques for sharing with me their typological data, Joey Redwood for her invaluable proofreading skills and all the participants in the experiment.

I owe thanks to Federico Albano Leoni, Francesco Cutugno, Renata Savy, Valentina Caniparoli and all the people involved in the CLIPS project for their great work and for making the corpus available to researchers.

This thesis would not have seen light without the collaboration of Marko Simonović, who helped me develop the idea of the *invariant* and conversed with me for endless hours in Utrecht, Padua, Bologna, Wroclaw, Split and Google Drive.

Last but not least, I thank Gianluca, Silvia, Giovanni, Maddalena, Chiara, Irene, Piero, Marco, Matteo, Giovanna, Anna and all the friends who have supported me, my aunt and godmother Rosa, my siblings Carlo, Riccardo, Valentina and Rita and my parents, who always encouraged me to do what I love.

ABBREVIATIONS AND CONVENTIONS

_ (underlined)	Headedness (of an element)
?	The occlusion element, [ʔ]
#	Word boundary
$	Syllable boundary
☞	Winning candidate in a tableau
💣	Wrong winning candidate in a tableau
1SG	First person singular
2SG	Second person singular
3PL	Third person plural
3SG	Third person singular
A	The aperture element, lowness, [a]
AF	Acoustic form
AL	Association line
ArtF	Articulatory form
B&B	Beats-and-Binding
Bold	Having a strong status in the surface form
C	Any consonant, or a consonantal slot in the skeleton
C	The element standing for occlusion/noise
COG	Center of gravity
CR	Correctness rate
CVCV	CVCV or Strict-CV theory
DP	Dependency Phonology
ET	Element Theory

ET-OT	A version of Optimality Theory employing elements instead of features
F	Any fricative
FEM	Feminine
FUT	Future
G	Glide
GP	Government Phonology
h	The noise element, [h]
H	Any non-sibilant fricative
H	The voicelessness element
I	The palatality/frontess element, [ɪ]
IMPER	Imperative
IMPERF	Imperfect
IND	Indicative
Italics	Having a weak status in the surface form
J	Any palatal stop
K	Any dorsal stop
L	Any liquid
L	The voicing element
MASC	Masculine
N	Any nasal, or Nucleus
N	The nasality element
NI	Northern Italian
NOM	Nominative
O	Onset
OT	Optimality Theory

P	Any labial stop
PL	Plural
PRES	Present
R	Any sonorant
R	The rhoticity element
RCVP	Radical CV-Phonology
S	Any sibilant
SF	Surface form
SG	Singular
SPE	*Sound Pattern of English*
T	Any stop
$\mathrm{T}_{\mathrm{cor}}$	Any coronal stop
U	The labiality/velarity element
UF	Underlying form
UG	Universal Grammar
V	Any vowel, or a vocalic slot in the skeleton
V	The element standing for loudness, sonority and/or voicing
VOC	Vocative
X	Any dorsal fricative
θ	Any non-sibilant coronal fricative
Φ	Any labial fricative

0

INTRODUCTION

It is quite surprising that we, as speakers, pronounce sound sequences that we think we are not able to produce, probably because we are sure that we are pronouncing something completely different. Similarly, as listeners, we hear things that are not being pronounced and we do not hear things that are actually being pronounced. Basically, most of our speech interactions are based on hallucinations and distortions. Most of these distortions are due to reduction and deletion processes, which target both consonants and vowels, although to different extents. While reduction and deletion in casual speech are pervasive, they are still poorly understood. The main questions that I aim to address in this thesis are the following: what are the factors which determine what can be reduced/deleted in conversation and what must be preserved? Is it random? Is it utterly dependent on the phonological context? Does it depend on the nature of the single segment and/or feature? Is it word-specific or universal? One of the simplest possible solutions is that reduction (including deletion, which I consider the final stage of reduction) is determined by ease of articulation (or better, by decrease of articulatory gestures) and articulatory undershoot. The preservation of certain units is explained instead by the need to maintain lexical distinctivity.

The thesis is structured as follows: in the first chapter, I compare different phonological frameworks, particularly with respect to their treatment of phonetic variation. While the main problem of generative models is the postulation of one and only one underlying form for each morpheme, exemplar-based models fail to explain the primacy of the citation forms in relationship to reduced forms, even though the latter are more frequent.

The second chapter puts forward the proposal that the generative and the exemplar-based model can somehow be unified by assuming one representation for each lexical entry which contains information about the relative strength of the units composing it. Additionally, the theoretical machinery employed in this thesis is explained in detail, i.e., a version of Optimality Theory using elements instead of features and referring to multilinear surface forms which are mapped onto acoustic forms. Another proposal formulated in chapter 2 is that acoustic salience plays a prominent role in the preservation of certain segments, as greater salience implies a stronger status in the representation.

The third chapter can be divided into two parts. In the first part, it is shown that acoustic salience is able to explain certain phonotactic patterns which are problematic for both markedness- and sonority-based accounts. Subsequently, plateau clusters – clusters of consonants of equal sonority – are analyzed, showing that when sonority does not play a role, the most salient consonant tends to occur in the position furthest from the vowel. A typical example is the existence of /s/C clusters in a number of unrelated languages. The second part of the chapter presents the results of an experiment focusing on the perception of plateau clusters in a series of nonce words. It appears that participants performed better at identifying peripheral consonants (the first consonant in word-initial clusters and the last consonant in word-final clusters) when they were acoustically salient (e.g., /s/). The fourth chapter, after a brief overview of casual speech studies in other languages, focuses on Italian casual speech, in order to test if reductions and deletions in spoken Italian are somehow related to the salience of the segments involved. Four dialogues are extracted from a corpus of spoken Italian (CLIPS, Savy & Cutugno 2009) and the realization of several consonants are analyzed using Praat (Boersma & Weenink 2013). The results yielded by the analysis seem to prove that speakers do indeed tend to preserve highly salient material.

The fifth chapter attempts to establish the phonological correlates of salience, which has been considered so far exclusively on a phonetic level. It is proposed that high-ranked faithfulness constraints protect elements that occur in positions where they are not expected (given their greater informativeness) while markedness constraints militate for consonants to contain consonantal elements (such as **C** = **h**, **?**, **H**) and for vowels to contain vocalic elements (such as **V** = **A**, **L**). Phonological salience, however, does not depend solely on informativeness, but also on inherent salience, determined by headedness. Chapter 5 also aims to unpack MAX-INVARIANT – a constraint proposed in chapter 2 in order to account for the preservation of certain units in casual speech. It is argued that MAX-INVARIANT is an umbrella name for a series of constraints with the following template: MAX-*x*-IN-*y* where *x* can be any element on any tier and *y* is a syllabic position (either C or V). It is also proposed that the preservation of certain segmental or subsegmental units depends simply on positional factors (e.g., strong vs. weak). However, it is concluded that it is not possible to dispense with the concept of a lexically stored invariant altogether, since variation is often word-specific and even almost identical words may reduce to different extents.

Finally, in chapter 6, I attempt to draw some (albeit temporary) conclusions, address remaining issues and propose some topics for further research.

1

REPRESENTATIONS: POOR VS. RICH

Traditionally, generative phonologists have always been interested in the relationship between abstract representations and their phonetic output. As a matter of fact, most phonological theories couched in the generative framework have posited at least two forms, an underlying form and a surface form (henceforth, UF and SF, respectively. See Cole & Hualde 2011 for an overview). This has become the standard after generative phonology's milestone *Sound Pattern of English* (Chomsky & Halle 1968; SPE, henceforth). The idea is that children, during acquisition, are able to perceive all kind of acoustic information about the words they hear (Eimas 1974, Eimas et al. 1971, Jusczyk & Aslin 1995) but they gradually lose this capacity and become sensitive only to what is phonologically distinctive in their own native language. As a consequence, they deprive their overly rich representations from all redundant material and only maintain essential information, i.e., the UF. Subsequently, the passage from the UF to the SF is obtained through mappings, rules, processes, constraints or a combination of those (depending on the phonological theory). However, several problems arise if one considers that there is one UF for each lexical entry. First of all, it is undisputed that human beings are able to store a great deal of non-linguistic phonetic information without any apparent memory-related fatigue, such as speaker's voice, age, gender, mood, etc.

(Goldinger et al. 1991, Goldinger 1996, 1998, Smith & Hawkins 2012). If storing this information does not appear particularly costly for our cognitive system, why should it not be the case for linguistically relevant information? Secondly, it is a well-known fact that the phonetic output can vary greatly, especially for highly frequent lexical items, which can be reduced to the point of being nothing but a blurry "phonetic icon" (Albano Leoni et al. 1999:46). Nevertheless, variation does have a limit and while certain function words can be heavily reduced, different lexical items seem to tolerate different degrees of reduction and, at the melodic level, certain segments and features appear to be preserved more often than others. In order to solve the issues associated to positing an UF, non-generative theories, such as exemplar-based models (Johnson 1997, Pierrehumbert 2001, Gahl & Yu 2006, Hay et al. 1999, Clopper & Pisoni 2004), and the related usage-based phonology (Bybee 2001, 2006a, 2006b, 2010) dispense with the UF altogether and propose instead very rich word clouds where all the tokens of a lexical items are stored. Importantly, not all tokens have the same cognitive status but the higher the frequency rate of a token, the more likely its retrieval. The aim of this first chapter is to give a brief overview of the competing visions about lexical storage in phonology and then propose a unifying approach. The idea is that an abstract representation is eventually constructed by language learners but that this representation is not *homogeneous* like in classical generative accounts, but rather *uneven* (Baroni 2013:38, Baroni & Simonović 2013a, 2013b). Where does this unevenness come from? Mainly from two factors: frequency and acoustic salience. The most frequent tokens of a word are more likely to be stored and preserved, while the most salient acoustic features are more likely to be heard correctly through noise. Very frequent and/or acoustically salient features of a word represent its *invariant.*

1.1. Abstractionist models: less is more

Generative phonology is today represented by three main frameworks: Optimality Theory (Prince & Smolensky 1993/2004), Government Phonology (Kaye 1990) and its offspring, such as CVCV theory (Lowenstamm 1996, Scheer 2004, 2012) and GP 2.0 (Pöchtrager 2006), and the Concordia school (Hale & Reiss 2008). As much as these frameworks differ, they all agree on the following tenets:

- phonology is a module of the Universal Grammar (UG) and there is, therefore, an innate component.
- The lexicon contains abstract phonological structures that are phonetically interpreted. There is one and only representation for each word. Predictable, contextual, low-level phonetic details are dispensed with, since they are implemented outside phonology proper.
- Economy is an advantage. The less is stored in the lexicon, the better.

1.1.1 Optimality Theory

Optimality Theory (OT henceforth) finds its origins in both connectionism (Rumelhart et al. 1986, Smolensky 1987) and generativism (Chomsky 1957, 1965, Chomsky & Halle 1968) and moves the focus from representation to computation. In its classical form (Prince & Smolensky 1993) the grammar is composed of three components: GEN, H-EVAL and CON. Given an input, GEN generates a series of potentially infinite possible outputs that are fed to another component, H-EVAL. The job of H-EVAL consists in comparing and evaluating the outputs generated by GEN and it does that through CON, the constraint set. Constraints are universal but their ranking is language-specific. Similarities between unrelated languages are accounted for by claiming that the constraints at work are the same, while differences are explained by different constraint rankings. OT scholars are not so much interested in the nature of the input or in the structure of GEN, since it is basically H-EVAL which does all the work. One of the fundamental ideas of OT is the Richness of the Base, which claims that the input can potentially be everything – even monsters such as |kdpfsjklp| – but H-EVAL will just impede such monstrous inputs to surface as such. Another important distinction to make concerns the constraint set: most OT constraints can be assigned to two different families: Faithfulness constraints and Markedness constraints. Faithfulness constraints assign a violation every time there is a discrepancy between input and output, e.g., MAX militates against the deletion of input segments and DEP against the insertion of output segments (i.e., epenthesis). Markedness constraints assign a violation every time the output contains (relatively) marked structures. For example, it is quite undisputed that front rounded vowels are crosslinguistically marked. Therefore, if a hypothetical input contained /y/, a markedness constraint, such as *FRONTROUND (cf. Rubach 2000) would assign a violation to every output form containing [y]. However, [y] is the most faithful candidate to underlying /y/ and therefore there would be no violation of any faithfulness constraint. Languages that exhibit in their phonological inventory segments such as /y/ (e.g., French) must have a constraint ranking where faithfulness constraints referring to the height and rounding specification of vowels (e.g., IDENT-V[BACK, ROUND]) are ranked higher than *FROUNTROUND, as in Tableau 1. In languages that do not have phonologically distinctive front rounded vowels, on the contrary, *FROUNTROUND is undominated, as in Tableau 2.

Tableau 1. Ranking: F > M.

Input: /y/	ID-V[BACK, ROUND]	*FROUNTROUND
☞a) [y]		*
b) [i]	*!	

Tableau 2. Ranking: M > F.

Input: /y/	* FROUNTROUND	ID-V[BACK, ROUND]
a) [y]	*!	
☞b) [i]		*

OT has been heavily criticized for a number of reasons, among which phonetic variation raises a series of thorny issues. For example, if there is variation at the phonetic level, i.e, a single input can have several outputs, then one must assume that the ranking is not fixed but that, under certain circumstances, either constraints are re-rankable or some constraints are not crucially ranked with respect to others. However, since languages differ between each other in how they rank the universal constraint set, then the descriptive power of the theory appears to be significantly reduced. As a matter of fact, if both acoustic variation within the same language and crosslinguistic differences are accounted for by different constraint rankings, then the borders between inter- and intra-linguistic variation are blurred. Kager (1999:405-407) deals with the problem of variation in OT and exposes two ways in which scholars have tried to solve the issue. One proposal is to split the grammar in two distinct constraint rankings, i.e., *co-phonologies*. This first approach is advocated by Itô & Mester (1995) and has been applied mainly to Japanese. In their view, Japanese lexicon can be divided into native (Yamato), Sino-Japanese, foreign and unassimilated foreign (Itô & Mester 1995:184). The differences between the four lexical strata are particularly visible in relation with a number of constraints. Those concerning syllabic structure, such as NOCOMPLEXONSET, NOCOMPLEXCODA and CODACOND (only nasals or the first half of a geminate are allowed in the coda) are ranked high in the entire lexicon. NOVOICEGEM, which bans geminate voiced obstruents, is demoted under faithfulness in the unassimilated foreign stratum, e.g., *doggu* from English *dog*. NO-[P], which penalizes every candidate containing a singleton [p], is demoted in both unassimilated foreign and foreign strata, e.g., *sepaado*, from English *shepherd*, and *peepaa*, from *pepper.* POSTNASVOI, which states that post-nasal obstruents must be voiced, is ranked high only in the Yamato stratum, but is demoted in Sino-Japanese, foreign and unassimilated foreign, e.g., *sampo* ‘walk’, *hantai* ‘opposite’, *kompyutaa* ‘computer’, *santa* ‘Santa’. In (1) I show the relevant constraint ranking for each lexical stratum (note how FAITH is gradually promoted from the most native to the least native stratum).

(1)

Yamato: SYLLSTRUC > NOVOICEGEM >NO-[P] > POSTNASVOI > **FAITH**.

Sino-Japanese: SYLLSTRUC > NOVOICEGEM >NO-[P] > **FAITH** > POSTNASVOI.

Foreign: SYLLSTRUC > NOVOICEGEM >**FAITH** > NO-[P] > POSTNASVOI.

Unassimilated Foreign: SYLLSTRUC > **FAITH** > NOVOICEGEM >NO-[P] > POSTNASVOI.

Another possibility is to maintain a single constraint ranking allowing some constraints not be crucially ranked with another. An example of this approach is the Partially Ordered Grammar proposed by Anttila (1997) for Finnish genitive suffix selection. Anttila quotes Prince & Smolensky's claim that "crucial nonranking" is a possibility for the theory, even though they always implicitly assumed that grammar was totally ordered. Conversely, Anttila distinguishes between grammar and tableaux:

(2)

GRAMMAR 1: A > B > C

GRAMMAR 2: A > B, A > C

GRAMMAR 3: A, B, C

Typical OT grammars normally correspond to Grammar 1, where all the constraints are ranked. In grammar 2, A dominates B and C but B and C are not ranked with respect to each other. As a result, grammar 2 consists of two tableaux, as shown in (3).

(3) GRAMMAR 2:

Tableau 3. Ranking: A > B > C

Input	A	B	C
Candidate 1	*	*!	
☞Candidate 2	*		*

Tableau 4. Ranking: A > C > B

Input	A	C	B
☞Candidate 1	*		*
Candidate 2	*	*!	

The two tableaux have two different winners, therefore, variation is predicted. In grammar 3, the possible constraint rankings are six:

(4) GRAMMAR 3:

A > B > C

A > C > B

B > A > C

B > C > A

C > A > B

C > B > A

Grammar 3 contains six tableaux. If candidate (1) violates constraints A and B and candidate (2) violates C, then candidate (2) will win in 2/3 of all tableaux. Put differently, there is variation but candidate (2) is the preferred output and it will occur more often than candidate (1).

A third solution is invoked by Pater (2000, 2010), who proposes the existence of indexed constraints that are selected only by certain lexical items. For example, Pater (2010) shows that in Yine there are suffixes that trigger syncope and suffixes that do not, and since there is no phonological difference between the two types, this difference must be lexically encoded.

(5)

(a) /heta + ya/[1] [hetya]	'see there'
(b) /heta + wa/ [hetawa]	'going to see yet'.

According to Pater, suffixes that behave like *ya* are lexically marked as syncope triggers, whereas those that behave like *wa* are not. Consider now the following constraints:

(6)

ALIGN-SUF(L)-C	The left edge of a suffix (which is lexically marked) coincides with the right edge of a consonant.
MAX	Input segments must have a correspondent in the output.
ALIGN-SUF-C	The left edge of a suffix coincides with the right edge of a consonant.

---

[1] Here I maintain Pater's notation <ya> but the reader should be aware that <y> stands for IPA /j/.

Distinguishing between ALIGN-SUF(L)-C and ALIGN-SUF-C is crucial because the former will be ranked higher than the latter: ALIGN-SUF(L)-C > MAX > ALIGN-SUF-C.

Tableau 5: Lexically-dependent syncope in Yine

Input	Output	ALIGN-SUF(L)-C	MAX	ALIGN-SUF-C
heta + yaL	hetaya	*!		*
	☞hetya		*	
heta +wa	☞hetawa			*
	hetwa		*!	

Boersma (1997, 1998) and Boersma & Hayes (2001) propose a revised version of the theory called Stochastic OT, in which the constraint ranking is continuous and that the distance between different constraints in the hierarchy is not fixed. For example, given a hierarchy such as C1 > C2 > C3 > C4, C2 and C3 might be much closer than, say, C3 and C4, so that, under certain conditions, random noise can affect the position of C2 and C3 and their relative ranking can be reversed. As a result, there might be situations where C2 dominates C3 90% of the time and C3 dominates C2 10% of the time. Stochastic OT is advantageous since it is able to include frequency effects but not directly; speakers do not have to store information on the frequency of forms as such but indirectly, as constraint rankings. However, the UF is still considered to be only one and free of redundant information, as in former generative approaches.

The variation problem in OT is also connected to the problem of abstractness, not only of the input but of the output as well. In other words, variation can be both phonological and phonetic. For example, in Standard Italian, nasals assimilate to the place of articulation of the following stop regardless of morpheme boundaries, e.g., *stanco* |stanko| 'tired' → /staŋko/, *con Carlo* |kon+karlo| 'with Charles' → /koŋkarlo/. However, in hypercorrect, formal speech, speakers can avoid assimilation and pronounce the sequence *con Carlo* as [konkarlo]. I argue that at this level, variation is still phonological. Quite differently, if we examine data from spontaneous speech, here the deal of variation is much greater, and, most importantly, speakers are normally completely unaware of it, e.g., *sinistra* 'left', /sinistra/ → [sinistra, sinistrə, sĩɲistrə, sĩisra, sɲsrə, sə̃ɲsr, sʔ̃sr, …]. This kind of variation is normally described as acoustic reduction (Ernestus 2000) and generative theories tend to ignore it, considering it outside the scope of grammar proper. However, acoustic reduction has important implications for sound change, language acquisition and the relationship between the input and the output. Boersma (2011:34) proposes to consider not only two levels of representation, i.e., underlying and surface, but four, i.e., underlying form (UF), surface form (SF), acoustic form (AF) and articulatory form (ArtF). Merely phonological variation concerns the relation between the UF and the SF, whereas phonetic variation is a matter of relationship between SF and AF/ArtF. Following Boersma's notation, the UF is given in pipes, |a|, the SF in slashes, /a/, and the AF/ArtF forms in brackets, [a]. For example, in Northern Italian (NI), there can be intra- and inter-speaker variation in the pronunciation of intervocalic /s/ when the morphemic boundary has become or is becoming opaque, e.g., a word like *risaltare* /risaltare/ was originally formed by the iterative prefix *ri-* and the verb *saltare* 'to jump', but has now the meaning of 'to stand out, to show up, to be salient' (together with the transparent meaning of 'to jump again'). As a matter of fact, both [risaltare] and [rizaltare] are possible pronunciations, the former denoting that the speaker still perceives a morphemic boundary between *ri-* and *saltare*, the latter signaling that the compound has been reinterpreted as a lexical entry of its own (cf. Baroni 1999, Passino 1999). The variation between the form with [s] and the form with [z] is a matter of ranking of faithfulness and markedness constraints which from the UF |ri(+)saltare| can select either /risaltare/ or /rizaltare/ as output. Here we are still in the domain of grammatical competence. In an OT tableau, the variation can be represented as such.

Tableau 6: Variation in the voicing of |s|

|risaltare|	ID-VOICE	*VsV
a) /risaltare/		*
b) /rizaltare/	*	

In Tableau 6 the faithfulness constraint ID-VOICE and the markedness constraint *VsV are not crucially ranked (as shown by the dotted line in the tableau). ID-VOICE militates for the preservation of input voicing specifications in the output, whereas *VsV bans sequences of two vowels with an intervening /s/. In NI, both outputs are possible and therefore the tableau shows no winner. However, in fast or casual speech styles, a great deal of phonetic variation can occur, for example, unstressed vowels can be realized as schwa or zero, as in [rizltar, rʲzaltarə, rzaltr, rzltarə…]. These pronunciation variants are examples of AF/ArtF, which are the output of the SF /rizaltare/. Faithfulness constraints preserve some of the elements composing the SF, whereas sensorimotor constraints drive towards reduction. Whereas the markedness constraints affecting the SF might actually be part of one's grammar, sensorimotor constraints concerning the AF/ArtF can still be called *constraints* but in a broader sense, i.e., they act more like natural tendencies due to articulatory and perceptual ease, grounded in physical facts. On the contrary, faithfulness must refer to some abstract representation.

Tableau 7: possible casual speech realization of /rizaltare/

/rizaltare/	MAX-C	MAX-STRESSV	*UNSTRESSEDV	MAX
☞a) [rizltarə]			**	*
b) [iaae]	*!****		***	*****
c) [rzltr]		*!		****
d) [rizaltare]			***!	

Going from the SF to the AF/ArtF, faithfulness constraints as MAX-C and MAX-STRESSV militate for the preservation of the input consonants and of the stressed vowel, whereas MAX wants to preserve all the segments present in the SF. However, the constraint *UNSTRESSEDV aims for articulatory ease and wants to reduce the movements necessary for uttering the word. Unstressed vowels might then be deleted altogether. Under this hypothetical ranking, MAX-C > MAX-STRESSV > *UNSTRESSEDV > MAX, the winning candidate is (a) because both consonants and the stressed vowels are pronounced. (b) is ruled out because no consonants are preserved and (c) is out of the game because the stressed vowel is deleted. The most faithful candidate, (d), loses against (a) because, maintaining all input vowels, it crucially violates *UNSTRESSEDV one time more than (a). Similarly, in Standard Italian, the prefix *ri-* before a stem beginning with a vowel is normally realized with a vowel and not with a glide, as a consequence of the morpheme boundary: |ri + andare| → /riandare/ → [riandare] 'to go again'. However, in casual speech, there is free variation between [ria…] and [rja…], where the faster the tempo the more likely the vocoid will be realized as a glide (regardless of the crosslinguistic avoidance for /r+j/ sequences, cf. Hall & Hamann 2010, Lyche 1979). In an unpublished experiment carried out by the author (Baroni 2012b), six Italian native speakers were recorded while pronouncing a list of words containing sequences of a consonant followed by a high vocoid and a vowel. The recordings were subsequently analyzed using Praat (Boersma & Weenink 2013) in order to determine whether the speakers realized the vocoid as a vowel or a glide. The phonetic correlate of "glideness" was considered to be the amplitude of the transition, i.e., the absolute difference between the F2 values of the onset and the offset of the vocoid. The greater the difference, the more abrupt the transition and the more likely the "glideness". With regards to the high front vocoid, the results showed that the average F2 onset-offset difference was greater after obstruents and nasals than after liquids and was particularly small after [r], as shown in (7).

(7)

Average |F2 onset – F2 offset|: n > obstruents > m > l > r

ANOVA *p* value = .007.

Therefore, morphological information and articulatory ease might conspire to block [rj] to surface in slow tempo (remember that participants to the experiment were reading the word list), but in spontaneous speech both factors are potentially obscured by the urge to speak faster.

While an OT account might at first look satisfactory, there is one major problem with the analysis proposed in tableaux 6 and 7: if the output only depends on the constraint ranking and there is no other element to take into account, then the model is too powerful and predicts too many unattested, if not impossible, grammars. For example, to my knowledge, there are no languages where a SF such as /rizaltare/ would be phonetically interpreted as [iaae]. May this fact suggest that the ranking MAX-C >> MAX-V is universal? Hardly so, since there are many examples of consonant lenition and deletion in the world's languages. One of the aims of the following chapters will then be to explain why [rizltarə] is a possible pronunciation of /rizaltare/ and [iaae] is arguably not.

1.1.2 Government Phonology

Government Phonology (GP henceforth) finds its origins in autosegmental representations, which arose independently in the works of many scholars during the 1970s (Leben 1973, Goldsmith 1974, Kahn 1976, Liberman 1975, McCarthy 1979, Clements 1977). Unlike OT, it is a representation-oriented theory (Scheer 2011:441) and works mainly on the representation of word and syllable structure (Kaye et al. 1990, Charette 1991, Harris 1994, Kaye 2005). Its offspring, CVCV or strict CV (Lowenstamm 1996, Scheer 2004, Szigetvári 2002, Cyran 2003) sees syllable structure as a concatenation of non-braching onsets and non-branching nuclei (i.e., C and V), between which forces such as Proper Government and Licensing apply every time they can. Both GP and CVCV dispense with binary features of the SPE-type (such as [+high, –round]) and use elements instead (Kaye et al. 1985, Harris 1990, Backley 2011). Consequently, /a/ does not consist of a bundle of features such as [+low, –high, –round] but simply of the element **A**, standing for lowness/aperture. The fact that some consonants have shown to have a lowering effect on the adjacent vowels is thus explained as **A**-spreading (cf. English *fit* [fɪt] vs. *fir* [fɜː], where historically, [r] lowered the preceding vowel – therefore [r] contains **A**). The main focus of these theories is the abstract representation and representational economy is highly valued. Elements are preferred to features because they are fewer and less redundant (even though they might overgenerate as well, see Breit 2013) and in CVCV surface phenomena such as the emergence of consonant clusters are regarded as phonetic interpretation of something that underlyingly is always a sequence of C- and V-slots. Little or no attention is paid to low-level phonetic phenomena or to acoustic variation and reduction. Elements are abstract cognitive entities that are phonetically interpreted as sounds and surface variation does not affect the representation and neither directly derives from it. Everything that happens after phonetic interpretation is not part of phonology anymore and nothing is to be said about it. Crucially, Scheer (2013) clearly states that phonology and phonetics are separate domains and that in order to communicate, they need a *translator,* as much as phonology and morphosyntax. Therefore, even if normally phonological patterns and rules are not *crazy* (e.g., underlying high vowels are typically realized as [i, u]-like sounds), a phonology-phonetics mismatch can arise through diachronic evolution and in this light phonologists should not rely excessively on phonetics but observe the phonological behavior of segments. For example, even though in French and German the rhotic is realized as an uvular fricative or trill [ʁ, χ, ʀ], it still behaves as a sonorant, i.e., it is allowed as the second segment in tautosyllabic branching onsets.

1.1.3 Concordia School

The Concordia School is the name given by Blaho (2008) to the phonological approach advocated by Hale & Reiss (2000, 2003, 2008) and Hale et al. (2007). They strongly oppose OT, especially with respect to the use of constraints vs. rules and the groundedness of the constraints themselves. They argue for a rule-based phonology where rules are completely arbitrary and void of any naturalness whatsoever. However, it is also argued that phonetics plays quite an important role, since unless two sounds have exactly the same phonetic interpretation, they must have different phonological representations. The question is: which phonetic interpretation? At which stage are we now? If underlying /a/ is realized as [a] in certain tokens of the same word and as [ə] in other tokens, must this have reflections on its phonological representation? Or should we just consider the citation form and ignore all the pronunciation variants diverging from the standard? In OT, whether /a/ is pronounced as [a] or [ə] does not impact its representation, since it is H-EVAL the component responsible for its realization, but for a theory that allows phonology to be interpreted directly by phonetics, acoustic variation appears more problematic.

1.2 Exemplar-based models: the richer, the merrier

Models based on the storage of several tokens of the same word are not that popular in formal linguistic theory but have been around for a while, especially in psycholinguistics. Ernestus (2013) gives an account of their strengths and weaknesses. The main assumption is that the lexicon contains many exemplars of every word and these exemplars together form a word cloud. Basically, whenever we hear or pronounce a word, we store it with all its details. These include: acoustic and articulatory characteristics, speaker-related idiosyncrasies, pragmatic context, speech rate, etc. Several experimental studies seem to prove that speakers are indeed able to store fully specified lexical representations; the participants to an experiment carried out by Goldinger (1998) had the tendency to imitate the pronunciation of the words they just heard and Cole et al. (1974) proved that it is easier to decide whether two words are identical or not if they are pronounced by the same person. These results lead to believe that humans are able to access and store fine phonetic detail. If speaker-specific characteristics can be stored and consequently recognized, why should recognition of reduced variants of a word be any different? If English *yesterday* is uttered as [jɛʃeɪ], Italian *allora* ‘by the way’ as [alr] or Dutch *afspraak* ‘appointment’ as [ɑspʋaː], then these forms are simply stored in the lexicon and associated, respectively, to unreduced forms such as [jɛstədeɪ, alːora, ɑfspraːk], together with all the other pronunciation variants. Associated forms will cluster together in word clouds. In exemplar-based models, comprehension consists of the mapping between the perceived acoustic forms and the exemplars forming the world cloud in the mental lexicon. With regards to production, there is some evidence that more than one pronunciation variant is stored in the lexicon, since reduction appears to be word-specific and sensitive both to structural factors (types of vowels and syllables composing the word) and frequency (see Bürki et al. 2010, Hinskens 2011). However, as Ernestus (2013) points out, getting rid of an abstract representation altogether may raise some issues. It is well known that speakers, even trained phoneticians, are generally unaware of acoustic reductions, both in perception and production. Therefore, it seems likely that listeners are able to reconstruct the citation form of a word from its reduced variant and do so unconsciously. In order to test this hypothesis, Kemps et al. (2004) presented Dutch native speakers with excerpts of spontaneous speech containing tokens of the suffix /lək/. The suffix appeared either as full [lək] or reduced to [k]. Participants were asked to press a button every time they heard the sound [l]. The results showed that participants pressed the button even when exposed to the reduced variant [k], i.e., when [l] was not physically present but just reconstructed. Other experiments (Mitterer & Ernestus 2006, Janse et al. 2007) showed that listeners are able to reconstruct the full form of a reduced variant if this occurs in the right context, but perform quite poorly when context does not help disambiguate. This is a problem for exemplar-based models, since, if every token is stored in the lexicon, listeners should not face particular challenges in recognizing them, with or without context. Exemplar-based models are not able to explain the primacy of unreduced forms either. It is quite undisputed that, when it comes to frequent words, the hyperarticulated, more citation-like form is not the most frequent one; we probably hear more often [jɛʃeɪ, alr, ɑspʊaː] rather than [jɛstədeɪ, alːora, ɑfsprɑːk]. If storage and lexical retrieval depend on frequency and recency effects, then, we would expect reduced forms to be more easily accessible than unreduced ones, but quite obviously, we know that this is not the case. How the citation forms maintains its primacy in relation with other variants remains somewhat unexplained without positing an abstract UF.

1.2.1 Usage-based linguistics

The work of Joan Bybee (2001, 2006a, 2006b) is one of the most prominent examples of application of the exemplar-based approach to linguistics. Bybee argues that sound changes are the result of natural phonetic processes that apply first to frequent items and extend subsequently to less frequent ones. For example, in English, frequent *camera* and *every* have lost their internal schwa but relatively rare *mammary* and *homily* have retained it. Similarly, *don't* appears heavily reduced when it occurs between *I* and *know, mean, think* but less so when preceded by a less frequent pronoun and followed by less frequent verbs. Again, irregular verbs are less likely to undergo analogical leveling if highly frequent, cf. *kept* vs. *wept* → *weeped* (Silverman 2011:380)[2]. Concepts such as phonotactic competence, which is considered to be part of grammatical knowledge in generative phonology, is claimed to be emergent and gradient (Frisch & Zawaydeh 2001, Frisch 2004, Bybee 2001). Repeated exposure to a sensory event strengthens its representational status and the likeliness of its categorization. In phonological terms, so-called allophones can be described as exemplars clustering together, organized around an exemplar which is the most prototypical of that particular category, traditionally

---

[2] Analogical levelling of infrequent forms has the side effect of regularizing morphology, dispensing with lexical exceptions.

called phoneme (for prototypicality effects in phonology, see also Jaeger 1980). However, usage-based phonology does not need the phoneme, since sensory events can be recorded as such, without the need to decompose it in smaller units (Silverman 2011:382). Usage-based linguistics is typically carried out using corpora and applying statistical analyses. For example, Bybee (2006a:297) reports that in a corpus consisting of 3 hours and 45 minutes of spontaneous conversation tape-recorded by Scheibman, she found 138 tokens of *don't*, which she grouped in four categories, presented in (8).

(8)
Group 1: tokens with a full stop and a full vowel, such as [dõ], as in *we* [dõ] *see him all winter.*
Group 2: tokens with an oral or nasal flap, such as [ɾõt, ɾõ, ɾ̃õ], as in *they* [ɾõ] *know who did it.*
Group 3: tokens with a reduced consonant and a reduced vowel, such as [ɾə, ɾə̃], as in *I* [ɾə̃] *know if I could do that.*
Group 4: tokens with just a reduced vowel, such as [ə, ə̃], as in *I* [ə̃] *know anything about guns*.

Bybee explains the data arguing that *don't* will be more reduced when used in the context where it appears more often. Of 37 tokens of *I don't know* in the corpus, only eight contain a full-vowel variant of *don't*. The sequence *I don't know,* moreover, can be employed both in a lexical sense and with a pragmatic function, like a conversational filler. In the latter case, reduction to schwa is much more likely. Apparently, a generative theory of phonology cannot predict in an elegant way such a one-to-many correspondence between *don't* and [dõ, ɾõt, ɾõ, ɾ̃õ, ɾə, ə, ə̃], together with their contexts of occurrences, frequency rates and extraphonological conditions. Still, it is not clear how a hyperarticulated form such as [dəʊnt/doʊnt] might become the most prototypical token of its exemplar-cloud.

1.3. Hybrid models

The main problem of generative models is their inability to account for frequency-related phenomena, whereas exemplar-based models cannot explain the privileged status of unreduced variants. A possibility is that for each variant that is stored, also information about the context where it occurs is recorded. It might as well be possible that when no top-down information is available (e.g., the proper context), lexical retrieval is based on the phonetic distance between the encountered token and the unreduced variant. The latter possibility opens the way to hybrid models, where both exemplars and abstract representations are needed. Several studies (McLennan & Luce 2005, Mattys & Liss 2008) have proved that indexical information is employed mainly when conditions are made somehow difficult, e.g., when in the experiment very long pauses are introduced or words are articulated very slowly. Pierrehumbert (2002) proposes that speakers mostly use exemplars in perception and both exemplars and abstract representations in production. McLennan & al. (2003), as well as Goldinger (2007), assume abstract representations and exemplars both in production and perception, with the matching of the acoustic input with the abstract representation occurring before the activation of the exemplars. Finally, Polysp (Polysystemic Speech Perception), a model proposed by Hawkins & Smith (2001), sees phonological analysis as unnecessary and highly dependent on the situation. A listener may or may not analyze the acoustic signals in phonemes or other units and several factors, such as visual information, the speaker's mood, articulatory gestures, etc. are considered to play a role. A less detailed yet interesting proposal comes from van Oostendorp (2013:4-5), who argues for the co-existence of an abstract cognitive system, which may even be *substance-free*, with sociolinguistic knowledge, which contains a great deal of detail about the actual pronunciation of each sound and its occurrence. In the next chapter I will sketch my own model, which preserves most of the OT apparatus but assigns a greater role to phonetic variation and frequency.

2

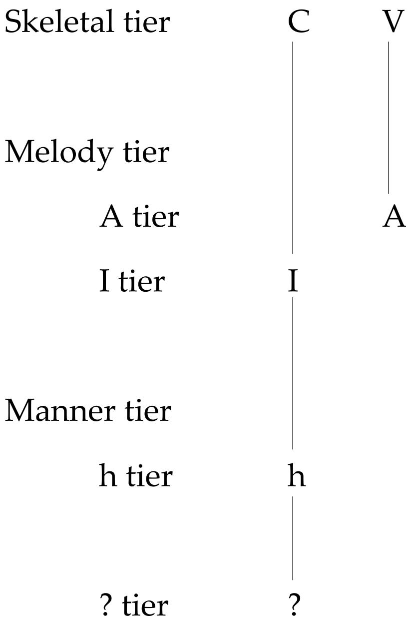
REPRESENTATIONS ARE UNEVEN

In this thesis I aim to reconcile two opposing views, i.e, abstract generative phonology and exemplar-based models, proposing a hybrid model of phonology where frequency and phonetics play a role in shaping abstract representations and OT-like constraint rankings but are not part of the grammar. In my proposal, speaker-specific indexical information is certainly stored somewhere, but is not considered linguistically relevant, i.e., does not impact on either the UF/SF or the constraint ranking. Fine phonetic detail, on the contrary, affects phonology, and does so in different ways. Acoustic information is stored but the acoustic elements that are more frequent in the input acquire a stronger status in the representation. This is a frequency effect. Similarly, if a certain sound is particularly salient, e.g., [s], it will have a stronger impact than, say, [ə]. Put differently, all acoustic information is stored but only frequent and/or salient elements pass from short-term memory to the mental lexicon. Pace Chomsky’s poverty of the stimulus, phonological input is very rich indeed but attention is limited (Carr 2005:27) and children only focus on salient aspects of the input (Vihman 1996). If a sound is momentarily stored and its representation is not soon corroborated by another occurrence of the same sound, it will decay. Therefore, there is but one lexical representation, containing information about all the possible pronunciation variants, e.g., if in all the pronunciation variants of *yesterday* there is an initial [j], /j/ will have a very strong status in the representation, whereas the status of word-internal schwa will probably be much weaker. This unevenness in the representation of the different elements correctly predicts that, under conditions that favor reduction, word-internal schwa will likely be omitted, while it will be practically impossible not to pronounce initial [j]. Bybee (2004:11-12) notes that token frequency has two seemingly contradictory effects: on the one hand, highly frequent items display greater variation, on the other hand, they are more resistant to analogical change. Similarly, I argue that while highly frequent words are allowed to vary to a great extent, the units (segmental or subsegmental) that occur most frequently in their pronunciation variants are more resistant to acoustic reduction. Why certain pronunciation variants occur in certain contexts is undoubtedly an interesting question but it is not the job of phonology to answer it. Among the few things phonology is certain about is what has to stay and what may go. My approach differs from the one proposed by Pierrehumbert (2001) especially when it comes to production. According to exemplar-based models, such as Pierrehumbert's, speakers select a certain exemplar from the word cloud contained in their lexicon. Not only would this imply that exemplars are stored but also that they are retrieved in production. For example, a heavily reduced pronunciation variant, once stored, might be selected as the target in production when the social situation allows for it. On the contrary, I argue that there is no need to see the production process as a selection among stored tokens. Johnson et al. (1993) have brought experimental evidence that phonetic targets are hyperarticulated, i.e., speakers always aim at pronouncing the unreduced variant of a word and if they fail, it is because of factors such as articulatory undershoot, effort reduction, etc. (Johnson et al. 1993:506). This idea is consistent with the view expressed by Jakobson & Halle (1956) and Hockett (1955), namely that the citation form is the most relevant to phonological analysis, because it is the most information-rich. Note that here by citation form or unreduced variant I do not mean a hyperarticulated, unnatural pronunciation, since this would be the result of fortition processes (Donegan & Stampe 1979:142). Fortition processes, however, in order to apply, must manipulate phonological information that is already available. In sum, even if fine phonetic details and several pronunciation variants may be relevant in perception, in production they are not necessary to speakers, since they aim at the citation form, grossly correspondent to what is traditionally called the UF. Reduction, lenition, and the like do occur, but they do so under the pressure of cognitive, physical and pragmatic constraints, which are in turn conditioned by faithfulness constraints.

2.1 A theory of representation

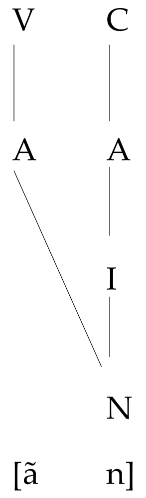
Quite unorthodoxly, in this thesis I stress the importance of both representation and computation. As noticed by Scheer (2004:380), in OT "computation is king" and little importance is given to representations. It can be said that OT is somehow incomplete, since most of its theoretical machinery must be "stolen" from other theories, i.e., the use and the type of features, morae and prosodic units employed heavily depend on the author's choice. However, as stated in the previous chapter, I aim to switch the burden from computation to representation, in that I assume that information about the strength of phonological units is stored. In my view, surface representations are multilinear and composed of autosegments, which are distributed on different tiers, as in classical autosegmental theory (Goldsmith 1976, 1990, McCarthy 1981). I distinguish at least three main tiers, i.e., the skeletal tier, which is composed of C and V units (consonants and vowels), the melody tier, where units corresponding to place features are disposed, and the manner tier, where units corresponding to manner features are arranged. The melody tier and the manner tier can themselves be broken down into further tiers, i.e., one tier for each unit. For instance, the manner tier may consist of a **?**-tier and an **h**-tier, where **?** stands for total occlusion of the vocal tract and **h** stands for partial occlusion or frication.

(9) Representation of the syllable /ta/



Elements can be shared by two skeletal positions, as shown in (10), where the nasalization of a vowel preceding a nasal consonant is represented.

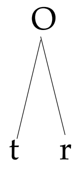
(10)



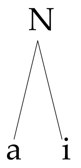
Classical autosegmental theories, such as Government Phonology (Kaye 1990) assume the possibility for onsets and nuclei in the skeleton to be branching. From this perspective, a branching onset such as /tr/ would be represented as in (11a) and a branching nucleus such as /ai/ would be represented as in (11b).

(11)

(a)



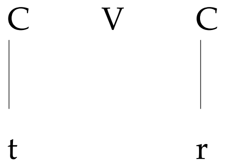
(b)



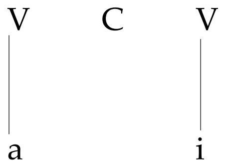
On the contrary, strict CV or CVCV theory (Lowenstamm 1996, Scheer 1999, 2004, Szigetvári 1999) denies the existence of underlying branching onsets and nuclei and assumes that the skeleton consists of a series of C and V units. /tr/ and /ai/ would therefore be represented as in (12a) and (12b), respectively.

(12)

(a)



(b)



Because of this fundamental assumption, CVCV is forced to postulate the existence of a number of empty categories (i.e., empty onsets and empty nuclei). Since the structure of consonant and vowel clusters is not relevant for the analysis that I aim to carry out in the following chapters, I will remain agnostic in this regard and the representations I will present will be as close as possible to the phonetic reality, i.e., a sequence of consonants on the surface will be represented as a sequence of consonants underlyingly. Another question that I will not deal with is the representation of laryngeal features (except occasionally). Since in the current study the focus is on consonants and, specifically, on sonorants (which are inherently voiced) and voiceless obstruents, I will not employ any element or feature indicating their laryngeal specification, such as **L**, **H** or [voice]. Intervocalic voicing will be interpreted as a kind of assimilatory process (see 4.3.3.4), but I will not be specific about which element or feature is actually transmitted from the vowels to the consonant. Finally, it is important to distinguish between the UF, the SF and the ArtF/AF. Following Boersma (2011), the UF is regarded as "a sequence of pieces of phonological material copied from the lexicon, with discernible morpheme structure, for example |an + pa|, where ' + ' is a morpheme boundary" whereas the SF is "a treelike structure of abstract phonological elements such as features, segments, syllables, feet" (Boersma 2011:3). The SF is then phonetically realized, yielding an AF, perceivable by the listener, and an ArtF, produced by the speaker. The AF and the ArtF are subjected to sensorimotor, perceptual and articulatory constraints, while faithfulness constraints refer to the relationship between the SF and the UF and markedness ("structural", in Boersma's words) constraints are well-formedness requirements on the SF. Given that this work revolves around casual speech phenomena, the main focus will be on the relationship between the SF and the ArtF/AF, therefore, where not otherwise specified, multilinear representations stand for SFs. For instance, let us consider the initial sequence of *impossibile* 'impossible' in

Italian, composed of the negative prefix |in-| and of the adjective |posːibile| 'possible'. Its UF is |in + posːibile|, its SF /imposːibile/ (with place assimilation of the nasal consonant) and its AF/ArtF will probably be [ĩmposːibile] in normal speech, with vowel nasalization. Representations are given in (13a-c). Only the relevant part of the word (i.e., the prefix followed by the first syllable of the stem) is shown.

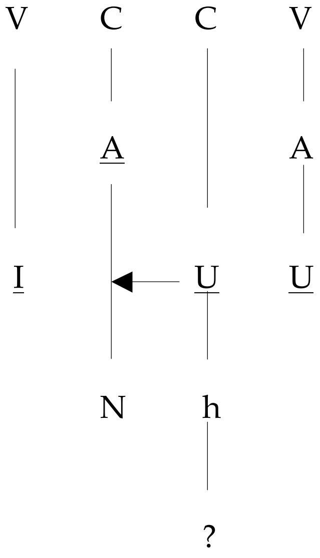
(13)

(a) *impossibile* UF

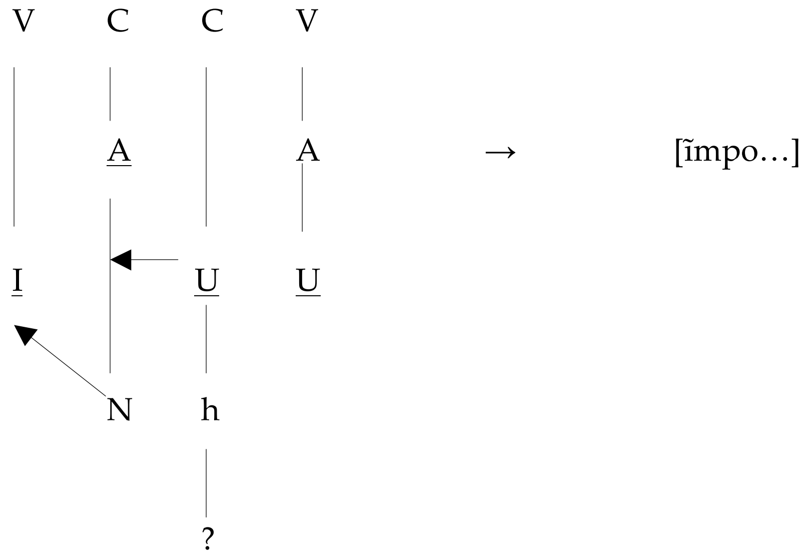
|in|+ |po…|

(b) From UF to SF

|in|+ |po…| →



(c) from SF to ArtF/AF



It is important to point out that while rule-based phonology and OT assume distinct representation levels that are somehow connected by rules, processes, derivations, constraints and the like, autosegmental theories do not distinguish between UF and SF in the same sense. My approach, while employing representations in the style of autosegmental phonology, maintains the distinction between levels, and in particular, insists on the difference between the UF, SF and AF/ArtF. Since in other generative phonological theories the UF generally corresponds to our SF and the SF corresponds to our AF/ArtF, readers should not mistake them. The SF employed in this thesis, albeit being called "surface", is still abstract, whereas the actual phonetic implementation takes place at the AF/ArtF level.

2.2 Elements

Another important point of my proposal is that phonemes or segments are epiphenomenal and that SFs contain monovalent features corresponding to primes, known as elements, articulatory prosodies (Niebuhr & Kohler 2011) or melody in other theories. I will use the term elements, henceforth. For the moment, the elements I choose to consider are basically those proposed by standard Element Theory (Harris 1990, Scheer 2004, Backley 2011) with some modifications. The choice of using elements (e.g., **A**, **I**, **U**) instead of binary features (e.g., [+low, +high, +back]) is not arbitrary but rather consistent with a precise vision of how linguistically relevant sounds are perceived and stored by listeners. More specifically, in accordance with exemplar-based models, words are not seen as a sequence of discrete units such as phonemes but as a more complex co-occurrence of "prosodies", in a Firthian sense (cf. Ogden & Local 1994). Firth (1957) called prosodies phonetic features such as [h] or aspiration, [j] or palatalization, [w] or labialization, etc. In his view, the assignment of these prosodies to the segmental or to prosodic level was a language-specific matter, e.g., in English glottal aspiration is interpreted as a realization of the phoneme /h/ whereas in Ancient Greek it was deemed to be a suprasegmental feature. It appears then as no coincidence that the respective orthographies note the same sound as <h> in the former case and as <'> in the latter. In Firth's words, "any phonetic features (…) can (…) profitably be stated as prosodies of the sentence or word" (Firth 1957:253). A Firthian approach is invoked by Kohler (1999) and Niebuhr & Kohler (2011). In their study of highly reduced German words they demonstrate that speakers are able to recognize lexical items thanks to the preservation of "articulatory prosodies" (e.g., palatality, nasality, etc.). They report the example of the German word *eigentlich* 'actually', whose citation form is [aɪgəntlɪç] and which can be reduced to [aɪ̃]. Participants to a perception experiment were able to identify [aɪ̃] as a token of *eigentlich* thanks to nasality and palatality that were spread across the word. Their theoretical stance is that these prosodies are indices of the "phonetic essence" of a word and that a phoneme-based analysis inadequately explains the recognition of highly reduced words. Another interesting finding of Niebuhr & Kohler's experiment is that word recognition need not be always dependent on the semantic/syntactic context. Participants could correctly distinguish between [aɪ̃] as result of the reduction of *eine* in the phrase *eine rote* 'a single red one' and [aɪ̃] as result of the reduction of *eigentlich 'ne* in *eigentlich 'ne rote* 'a red one, really'. They were able to do so thanks to the longer duration of the palatal prosody in the second token, as well as to the duration of the preceding [a]. How such fine phonetic details are employed in lexical retrieval is hard to explain with a phoneme-centered approach, whereas it does not pose any challenge to an element-based phonological theory. Therefore, I argue that considering elements instead of features and phonemes is advantageous for describing both careful and casual speech, as well as the relationship between the two. The possibility of describing segments as composed by directly interpretable elements goes back to Kaye et al. (1985), who propose that elements be monovalent, i.e., they are either present or not, capable of spreading and directly pronounceable. For example, the element **A**, standing for aperture/lowness, in isolation is interpreted phonetically as an [a]-like vowel. The vocalic elements they present are five: **A**, **I**, **U**, **Ɨ**, **v**.

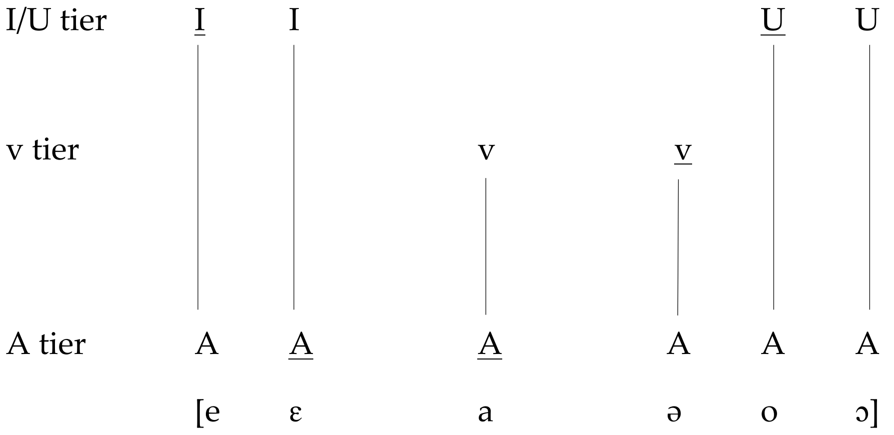
(14) Correspondence between elements and bundles of features:
A = [–round, +back, –high, –atr, +low]
I = [–round, –back, +high, –atr, –low]
U = [+round, +back, +high, –atr, –low]
Ɨ = [–round, +back, +high, +atr, –low]
v = [–round, +back, +high, –atr, –low].

The first three elements grossly correspond to aperture/lowness, frontness/palatality and backness/labiality, respectively. **Ɨ** is called the ATR element (Kaye et al. 1985:312) and participates, for example, in ATR harmony in languages that exhibit this phenomenon, e.g., Maasai. **v** is the cold element and basically indicates the neutral position of the mouth. In addition, **v** distinguishes buccal sounds, such as [t], from non-buccal sounds, such as [ʔ][3]. Elements can combine giving rise to several segments. A segment, at this point, can be defined as the association between one or more elements and a skeletal unit. In each segment one element plays the role of the head and the other is the operator. Headedness is represented by underlining.

---

[3] **v** distinguishes between [t] and [ʔ] in Scheer (2004), where coronal stops are described as having no melody, but other authors posit a different structure for [t, d]. For instance, they contain **I** in Botma (2004) and **A** in Kaye (2000). In this point I agree with Botma's analysis.

(15)



Since its inception, and depending on the author, ET has undergone a great deal of modification and variation. One of the most interesting achievements of the theory is the use of the same element set to define both consonants and vowels (e.g., **I** stands for frontness in vowels and palatality/brightness in consonants) and an alternative definition of sonority that finally escapes circularity. Previously, a segment was said to be more sonorous than another because it behaved as such, e.g., /r/ is more sonorous than /t/ because it occurs closer to the vocalic nucleus word-initially and word-finally, since /#trV/ and /Vrt#/ syllables are more frequent/less marked crosslinguistically than /#rtV/ and /#Vtr/. Harris (1990) proposes to consider sonority as lack of complexity: the more elements a segment is made of, the more complex it is, and therefore, the less sonorous. As a matter of fact, in Harris' view, consonants, unlike sonorants and vowels, contain manner elements, such as **H**, **L**, **?**, **h**, and therefore they differ more from vowels than sonorants. Scheer (1999, 2004) maintains the relationship between complexity and sonority but with a reverse logic: complexity must count only place definers, i.e., melodic elements. In this light, sonorants prove to be more complex than obstruents. I will follow, with some modifications, the elements and the consonantal structures proposed by Scheer (1999):

(16)

Place/Melodic elements

I	palatality/frontness	[ɪ]
U	velarity	[ɯ]
B	labiality/roundness	[ʉ]
A	aperture/RTR	[a]

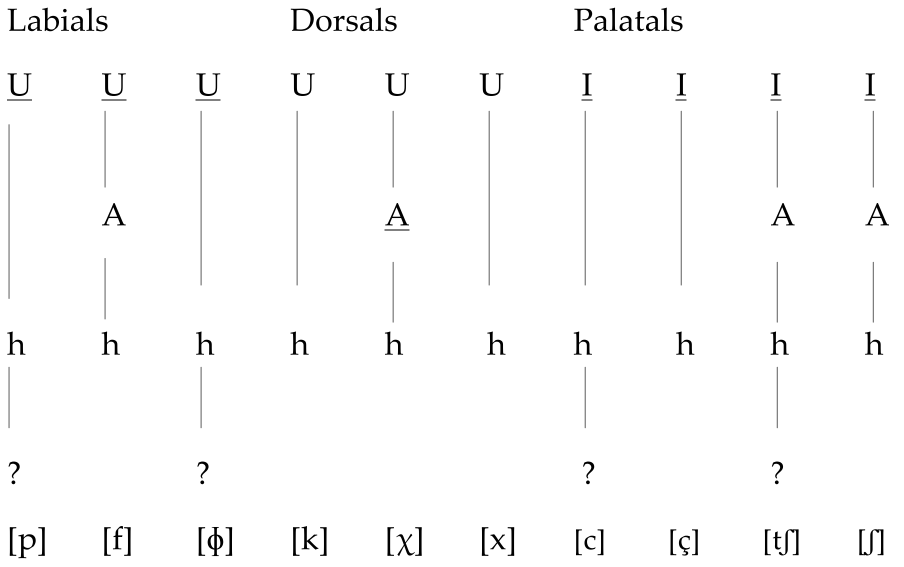
Manner elements

?	constriction	[ʔ]
h	noise	[h]
N	nasality	[˜]
L	lax vocal cords	
H	stiff vocal cords	
T	trill	

Many authors reduce the number of elements, for the sake of economy. The underlying logic is: the fewer the elements, the lower the possibility of overgeneration. However, for my analysis, the element set must be rich enough to account for all relevant phonetic features. Therefore, even if Botma (2009) and Botma et al. (2011) adduce sound reasons to dispense with **N** and to consider **L** the element interpreted as both voicing and nasality, in casual speech they behave differently (e.g., nasality is much more likely to spread than voicing) and it is therefore more descriptively advantageous to keep them apart. Backley (2011) unifies labiality and velarity under the label **U** and distinguishes labials and velars through headedness: labials are **U**-headed, velars contain **U** but are headless. The unity of labiality and velarity had already been noticed by Jakobson (1962), who characterized both sound classes with the feature [+grave]. The element **T** might seem rather ad hoc and, as a matter of fact, is advocated only by Scheer. I wish to replace it with **R**[4], standing for rhoticity, i.e., the phonetic correlates characterizing both rhotic sonorants and rhoticized vowels (see Spreafico & Vietti 2013 for an overview). I argue that **R** is not present in uvular fricatives such as [ʁ], although they may behave as rhotics in certain languages (e.g., French, German).

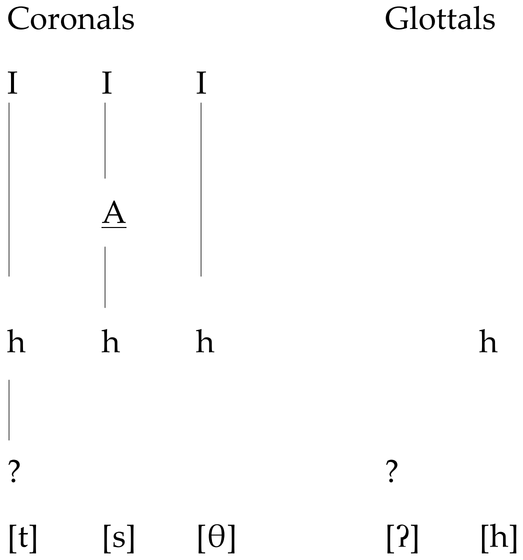
(17)

(a) Obstruents structure (no distinction between voiceless and voiced):

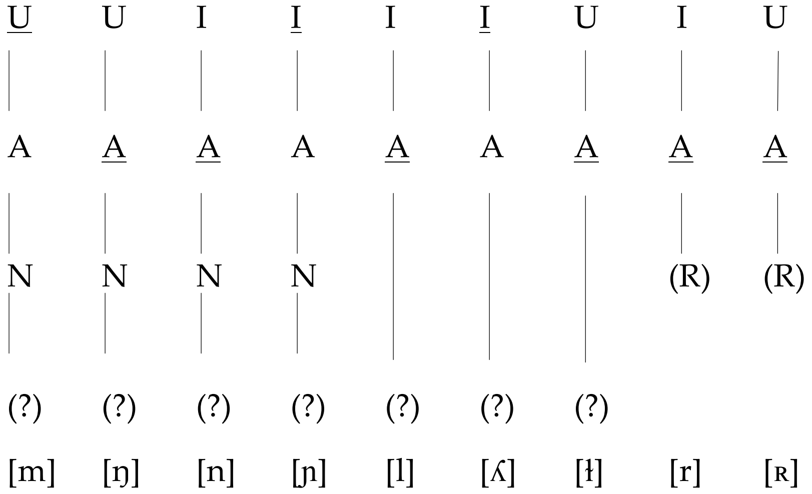


---

[4] The element **R** proposed here must not be confused with Kaye et al.'s (1990) and Harris' (1990) coronal element (later replaced by **A**).



(b) Sonorants structure



Every structure can be justified on diachronic, synchronic or both grounds. Velars and coronal non-strident sounds ([t, θ]) are headless, which is reflected in their phonological weakness, e.g., in Tuscan Italian, stops are lenited to fricatives intervocalically but [k], unlike [t, p], can be reduced to zero (especially in the Pisan variety, see Marotta 2008). In American English, coronal stops are subjected to intervocalic flapping, which does not affect either labial or velar stops. However, note that in some dialects of Polish only labial sounds can be deleted intervocalically (Jaskuła 2013). The fact that labials contain **U** appears pretty obvious. Historically, English [ʊ] switched to [ʌ] but certain words escaped this process, notably those in which [ʊ] was preceded by a labial consonant[5], cf. *nut* [nʌt], *dull* [dʌɫ] vs. *put* [pʊt], *butcher* [bʊtʃə], *full* [fʊɫ]. The structure of velars is more controversial and many have proposed that they are melodically void. Nonetheless, the close relationship they seem to have with labials still needs to be explained. In most languages, velar (back) vowels are rounded (labial), e.g., [u, o] are more frequent than [ɯ, ɤ]. In Czech (Scheer 2004:49) velars are able to spread the element **U** on the following vowel:

(18)
kůň 'horse-NOM' [kuuɲ] → VOC [kɔɲ-i], [ɲ] spreads **I**
hoch 'boy-NOM' [hɔx] → VOC [hɔx-u], [x] spreads **U**
pes 'dog-NOM' [pɛs] → VOC [ps-ɛ], [ɛ] occurs elsewhere

Moreover, in many languages there have been switches from labial to velar and the other way around.

(19)
Latin *noctem, lucta, pectus* → Roumanian *noapte, lupta, piept* 'night, fight, chest'.
Middle English *laugh* [x] → Modern English *laugh* [f].

---

[5] With the notable exception of *but* [bʌt].

Dutch *kracht, lucht, zacht* (where <ch> → [x]) vs. English *craft, loft, soft*, German *Kraft, Lunft, sanft* (Huber 2009:153).

All coronal sonorants are said to contain **A** because of their apparent lowering power on the adjacent vowels:

(20)

English *fit* [fɪt] vs. *fir* [fɜː], *put* [pʊt] vs. *purr* [pɜː].
Veneto Italian *albero* [albɐɾo] 'tree', cf. Standard Italian [albero].
Latin *per* [per] 'for' → French *par* [paʁ].
Proto-Germanic **sterron* → English *star* [stɑː].
High German *sunne, sumer, kumen* → Modern German *Sonne* [zɔnə] 'sun', *Sommer* [zɔmɐ] 'Summer', *kommen* [kɔmən] 'to come'.

Moreover, in many Germanic languages, final [r] vocalizes as a low vowel, such as [ɐ], e.g., German *Feuer* [fɔjɐ] 'fire', Danish *er* [æɐ̯] 'to be'. In the varieties of English with linking and intrusive R[6], [r] can be considered an **A**-glide (that is, the realization of **A** in a C position), as much as [j] is an **I**-glide and [w] an **U**-glide, e.g., *tuner* [ʧunə], *tuner amp* [ʧunəɹæmp], *I saw* [ɹ] *a film today, oh boy!* (from Beatles' song *A day in the life*).
[r, l, n] also contain **I**. In German the phoneme /x/ has two allophones, [ç] and [χ], the former occurs after front vowels and the latter in the other

---

[6] Linking *r* is a sandhi-like phenomenon characterizing some English dialects. Historically, it consists of the conservation in the pronunciation of a word-final rhotic, elsewhere disappeared, before a word beginning with a vowel. In synchrony, it can be described as the insertion, after a non-high vowel, of an epenthetic rhotic segment to avoid hiatus. Intrusive *r* is the same phenomenon, but it owes its name to the fact that it is not historically motivated (there used to be no rhotic) and is therefore unsanctioned by official spelling.

contexts, cf. *Ich* ‘I’ [ɪç], *Dach* ‘roof’ [daχ]. Interestingly, [ç] appears after coronal sonorants as well, see *Dolch* ‘dagger’ [dɔlç], *manch* ‘many’ [manç], *durch* ‘by, through’ [dʊɐ̯ç]. In some variants of Dutch and in Caribbean Spanish both /r/ and /l/ are realized as [j] in coda position, see Dutch [kaːrt] vs. [kaːjt] ‘to card’, [stɔːrt] vs. [stɔːjt] ‘to disturb’, Caribbean Spanish [revojvej] instead of [revolver] ‘revolver’, [kajta] instead of [karta] ‘paper’ (Scheer 2004:57, Harris 1983, Harris 1997).

It turns out that [r, l, n], melodically speaking, are basically equivalent, i.e., **A-I**. Further evidence for their identity comes from Yakoma, Kirundi and other Bantu languages where they are allophones of the archiphoneme /L/ (although loanwords have introduced non-native distinctions between the three phones – Mioni, p.c.). Alternations in Romance languages show that [n] can be the result of the fortition of [l] word-initially, cf. Italian *livello* vs. French *niveau* and Spanish *nivel* ‘level’, from Latin *libella* ‘balance’. In Italian, [l] is replaced with [r] to avoid a sequence of two syllables both starting with [l] when the suffix *-al-* is attached to a root ending with [l], cf. *speciale* ‘special’ (from |spetʃ + al + e|) vs. *alare* (from |al + al + e|) ‘wing-related’. In Veneto Italian, [l] alternates with [e] and zero, e.g., *scola* ‘school’ [skola ~ skoe̯a ~ skoa] and since [e] is made up of **A** and **I**, [l] must be too. Dark /l/ is of course different, since it is composed of **A-U**, rather than **A-I**, cf. the Italian pronunciation of English words with syllabic dark /l/, e.g., *little, middle* [litːol, midːol]. Dark /l/ vocalized in [u̯] in Brazilian Portuguese, e.g., *Brasil* [bɾaziu̯], backed and labialized the preceding vowel in English, e.g., *talk* [tɔːk], probably originally *[tałk], and became [o] in Serbo-Croatian, e.g., *misao* [misao] ‘thought’, from the verb *misliti* ‘to think’, cf. Polish *myśl* [mɨʃl].

The relation between coronal sonorants and [s] is reflected by their melodic structure. Historically, [s], in intervocalic position, was first lenited to [z] and subsequently to [r], e.g., Proto-Latin **auzoza* > Latin *aurora* 'dawn', cf. Ancient Greek *eos*, from Proto-Indoeuropean **ausus*. Similarly, Modern English *was* and *were*, both stemming from the Proto-Indoeuropean root **wes-* 'to remain, abide, dwell', underwent rhotacization. Jacques (2013) reports an unprecedented sound change, **s- > n-*, in Arapaho, a language belonging to the Algonquian family. Goddard (2001) proposes that the sound change took place by steps, firstly as a form of rhotacism, and then as a passage from a rhotic to a nasal. Word-initial rhotacism is attested in only a handful of languages, among which Vietnamese (Ferlus 1982), but is not impossible. Whatever the explanation, this sound change would bring further evidence that [s] and [n] share the same melody. The similarity between [s] and coronal sonorants is still apparent synchronically in certain languages, e.g., in Italian only [s, l, n, r] may occur in the coda ([m] only occurs before another labial sound and other segments only appear as the first part of a geminate).

The use of elements instead of features allows to describe Chinese vowel allophony in a particularly elegant way. In Chinese, there are five distinctive vowel phonemes, /i, y, u, e, a/ but each vowel has multiple allophones, depending on the preceding consonant, as shown in (21).

(21) based on Mioni (in preparation):

Phoneme	*Context*	*Allophone*
/i/	T_, N_, #_	[i]
	_n, _ŋ	[ɪ]

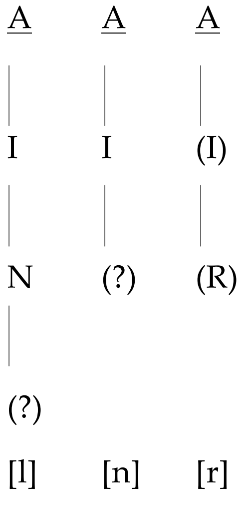
	S_	[ɯ]
	C$_{\text{[retroflex]}}$_	[ɻ̩]
/y/	Default	[y]
	_n	[ʏ]
/u/	_$	[u]
	_n, _ŋ	[ʊ]
/e/	Default	[e/ɛ]
	P_, _ʊ̯	[ɔ]
	T$_{\text{cor}}$_, S_,	
	C$_{\text{[retroflex]}}$_,	
	$_$	[ɤ]
	_n	[ə]
	K_, _ŋ, _r	[ʌ]
/a/	Default	[a]
	_ŋ, _ʊ̯	[ɑ]
	j_n, ɥ_n	[æ]

(21) shows that vowels are regularly lowered when followed by consonants containing **A**, such as nasals, labialized when preceded by consonants containing **U̲**, velarized when followed by consonants containing **U** and fronted when adjacent to **I**-sounds, such as [j]. It is not clear what the actual structure of retroflex consonants might be. Tentatively, given their crosslinguistic markedness, they might have a complex melody composed of **A**, **I** and **U**. As a matter of fact, retroflex consonants in Chinese are either sibilants or sonorants, so they arguably

contain **A** and **I**, but they also change /e/ into the back vowel [ɤ], suggesting the presence of **U**.

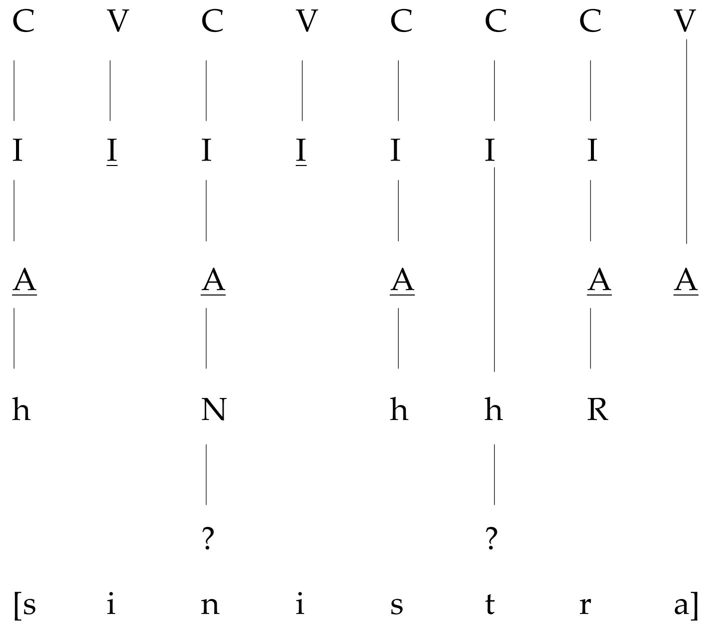
In (22) I show a representation of sonorants with manner elements within parentheses. This is because in some languages nasals and laterals seem to pattern with obstruents, i.e., they may contain a **?** element, whereas in others they pattern with sonorants. One of the axioms of GP is that there is no underspecification, therefore, all elements receive the same phonetic interpretation in every language. In spite of this, I take a different stance here, since I do not wish to work in orthodox GP but simply use GP-like elements instead of features. On a phonetic level, it might result descriptively useful to include manner elements, such as **?** and **R**, in the representation of sonorants because processes of acoustic reduction might leave just **?-N** as a trace for nasals and **R** (lowering of F3 values, r-coloring) as a trace for rhotics.

(22)



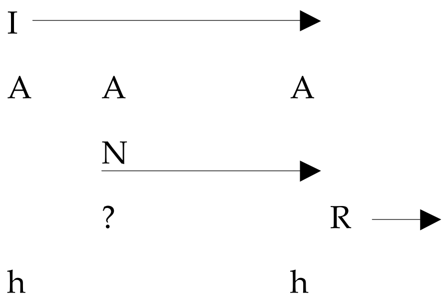
After having presented the internal structure of vowels and consonants, I will expound how ET can be implemented in order to explain acoustic variation. The SF of a word such as *sinistra* 'left' in Italian is presented in (23).

(23)



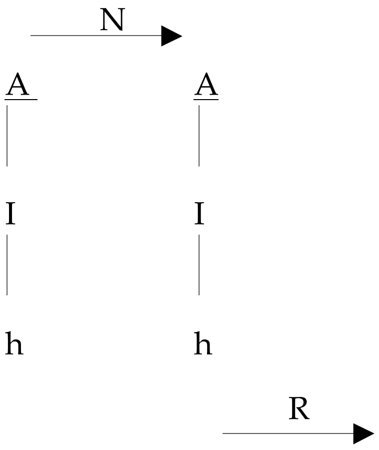
Consider then a series of actual pronunciation variants extracted from a corpus of spontaneous speech (CLIPS – Savy & Cutugno 2009): [sinistɝ, sɲisra, sinisra, si˜str, səsðr, sɲisr, sənʲsrə, sʲəjistr, siɾ̃istra, s˜sr, səʔisra, siʔisra…]. Moving the focus from the citation form to the actual tokens of the word, it appears that a better representation of *sinistra* would have to take into account the persistence of certain elements, i.e., their distribution over a time lapse greater than a single C or V slot. Put another way, elements such as **I, A**, **N** and **R** can be described as being spread over a portion of the word that is definitely bigger than a segment and in certain tokens, such as [sɲisr], **I** can be said to characterize the word as a whole.

(24)



(24) is a more realistic representation of the AF [sɲisr], where it is shown that palatality affects the whole word, nasality is present in the central part, which would correspond to the sequence /-ini-/ and that rhoticity may not be just borne by the last consonantal slot but may characterize the vocalic slot as well. The elements that form part of *sinistra*, depending on the speech style, speed, speaker-specific idiosyncrasies, etc., are mapped differently onto the syllabic skeleton (C and V slots). Looking at the most reduced variant of the corpus, i.e., [s˜sr], it is evident that the invariant consists of a sequence of two stridents interrupted by a nasal prosody and followed by a rhotic prosody. I argue that stridency is the product of the conjunction of the loudest melodic element, **A̲** (headed **A**), with **I** and the noise element **h**. Therefore, the invariant of *sinistra* might correspond to the representation given in (25).

(25)



The idea that certain sounds composing a word may also affect adjacent sounds is not new, and common phenomena such as assimilation and coarticulation are at the core of most phonological descriptions. Even long-distance relationships, such as vowel and consonant harmony, have been the object of much work (for an overview, see van der Hulst & van der Weijer 1995, Rose & Walker 2004). However, recent studies have shown that the mutual influence of co-occurring elements is much more pervasive than has ever been thought before. Hawkins & Smith (2001:113) discuss long-domain segmental information. They quote a study (Kelly & Local 1986) which proves that the resonances of /l/ and /r/ in English are able to affect not only the syllable they belong to, but also the word as whole. Since /l/ in the onset is brighter than /r/, the [i] of *Henry* is darker than the [i] of *Henley*. Even more interestingly, comparing two sentences such as *We heard that it could be a mirror* and *We heard that it could be a miller,* it turns out that the almost identical utterances display darker resonances before *mirror* than before *miller* (Hawkins & Slater 1994, Tunley 1999, West 1999). Quite surprisingly, these anticipatory resonance effects sometimes "skipped" over some syllables, e.g., in the sentence *We heard that it could be a mirror*, *it* showed resonance effects when *could* was stressed, even though *could* did not show any [r]-resonance effects (Heid & Hawkins 2000). Hawkins & Smith (2001:115) conclude that "[e]very phonetic segment is probably cued to one extent or another by both long- and short-domain acoustic properties" but generally short-domain (i.e., local) properties tend to be more informative. Short-domain cues are typically those associated with elements that cannot spread but that are better suited to localize syllabic positions, e.g., the burst of stops, represented by the occlusion element **?**. Long-domain cues range from information on the place of articulation of coda consonants (Warren & Marslen-Wilson 1987), nasality, palatality, labiality, etc., that is, elements that are typically able to spread.

(26)
Potential long-domain elements: **A**, **I**, **U**, **N**.
Short-domain elements: **?**, **h**.

2.3 Elements and constraints

In this section I present a version of OT which employs elements instead of features. This approach is not completely new. Blaho (2008), in her substance-free phonological grammar, abandones binary features and assumes unary features, which can be argued to correspond to elements. Polgárdi (1998) gives an account of vowel harmony combining GP and OT, maintaining concepts such as Proper Government, Licensing and elements on the one hand, and ordered constraint rankings on the other hand. Van der Torre (2003) explains the behavior of Dutch sonorants claiming that both their internal structure and constraint interaction play a role. However, his analysis differs from mine in several points. For instance, he argues that **A** defines both low vowels and velar consonants and that /l/ contains both **I** and **U**. While these choices are somewhat justified for theory-internal reasons, there does not seem to be evidence either for the presence of **A** in velars or for the presence of **U** in /l/, since the Dutch lateral is not dark.

In the current thesis the constraint templates do not differ much from those proposed in classical OT, except that they are element-based instead of feature-based. Faithfulness constraints may refer to elements (e.g., Max-A, Max-I, i.e., preserve the element **A**, preserve the element **I**), syllabic positions, skeletal units (e.g., C-slot or V-slot) and association lines (henceforth ALs), i.e., the occurrence of a particular element in a particular position. It is important to stress the fact that this kind of faithfulness constraints do not militate for the correspondence between the UF and the SF, but between the SF and the AF/ArtF, since only the SF has a multilinear structure.

In (27) I present a hypothetical SF where the elements **A̲**, **U** and **I̲** are associated with V and C slots. **U**, in particular, is associated with the manner elements **?** and **h** and belongs to a C slot. Faithfulness constraints of the type Max protect input elements from undergoing deletion, but a markedness constraint like *VʔV selects against a stop to occur between two vowels. Henceforth, this version of OT using elements instead of features will be denominated ET-OT. The evaluation is given in Tableau 8.

(27)

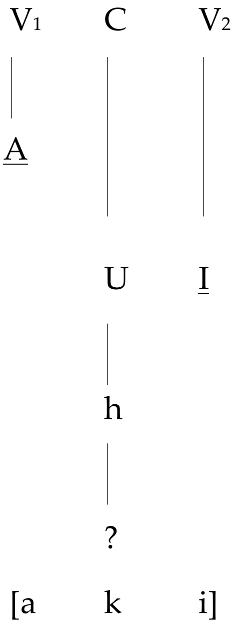
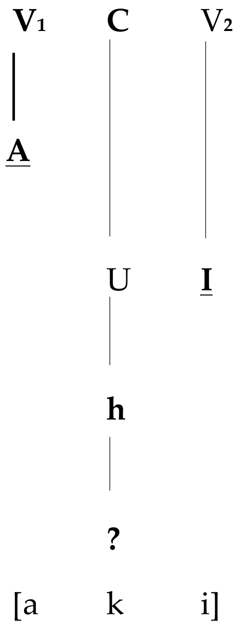


Tableau 8: Evaluation in ET-OT

/aki/	MAX-A	MAX-U	MAX-I	*VʔV	MAX-?
a) [aki]				*!	
b) [aci]		*!		*	
c) [aʔi]		*!		*	
d) [aku]			*!	*	
e) [uki]	*!			*	
☞f) [axi]					*

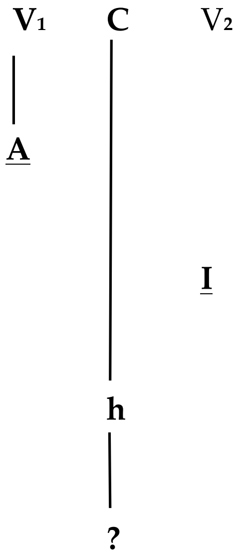
Candidate (a) would be the most faithful candidate, having the phonetic interpretation [aki], but it violates the markedness constraint *VʔV. (b, c, d, e) are all ruled out because they fail to preserve some elements. In (b) **U** is deleted and the C slot is occupied by **I̲**, in (c) **U** is deleted leaving a placeless consonant in the C slot, in (d) **I̲** is deleted and **U** spreads to the following V slot and in (e) **A̲** is deleted and **U** spreads to the preceding V slot. (f) is the winner, since all the melodic elements **A̲**, **I̲** and **U** are preserved and the markedness constraint that dominates Max-? is not violated. The output is therefore [axi]. The representation in (27) is an example of an even representation. An uneven one would assign different degrees of strength to the elements composing it, as in (28).

(28)



In (28) **A**, V_1 and the line associating them are in bold, as well as **?, h** and C. This means that the invariant consists of V_1-**A̲**, C-**h**-**?** and **I̲**, as shown in (29).

(29)



All the rest can be dismissed, e.g., **U** and V_2 can possibly be deleted. It is noteworthy that **A** is preserved but is not free: since its AL with V_1 is strong, **A** cannot spread, whereas **I** can, since its link to V_2 is weak, as well as the V_2 slot itself. Now consider the following constraints:

(30)

Max-Invariant	Preserve the invariant (= assign a violation every time **A**, **?**, **I**, V_1, C are not present in the output and every time **A** is not associated to V_1).
Max	Deletion is not allowed.
Spread-I	The element **I** must spread to the adjacent slots.
*Weak	Elements, ALs and slots that are not part of the invariant must not have correspondents in the output.

(31)

Ranking: MAX-INVARIANT > SPREAD-I > *WEAK > MAX.

Tableau 9: interaction between markedness and MAX-INVARIANT

/aki/	MAX-INVARIANT	SPREAD-I	*WEAK	MAX
a) [aki]		*!	**	
b) [aci]			*!	*
☞c) [ac]				**

(a) is not a possible winner because **I** does not spread. (b) loses the competition because it maintains the weak association between **I** and V_2. (c) appears to be the winning candidate because not only does **I** spread to the C slot, but also because **U** and V_2, which are weak, are deleted.

Two of the constraints proposed here appear as particularly problematic, i.e., MAX-INVARIANT and *WEAK. They are unorthodox, from a classical OT perspective, because they are shortcuts that stand for groups of constraints militating for the preservation or the deletion of certain parts of the input. So MAX-INVARIANT could be translated into MAX-A, MAX-I, MAX-?, etc. and *WEAK could be rewritten as *U, *V_2, etc. Moreover, these two constraints are basically empty or contentless, i.e., they represent a sort of "reserved place" in the ranking, which is bound to receive information from the SF about the content to refer to. Since acoustic reduction is word-specific, I assume that every word is stored in the lexicon with specifications about which elements are strong (i.e., form the invariant) and which ones are not. MAX-INVARIANT and *WEAK basically mean, respectively: what is strong must stay, everything else must go, where strong means both frequent and acoustically salient (the need for these two constraints will be further justified in section 5.5.4). The example in tableau 9 is of course oversimplified. Elements, ALs and syllabic slots may have different degrees of strength/weakness and everything is not simply black or white.

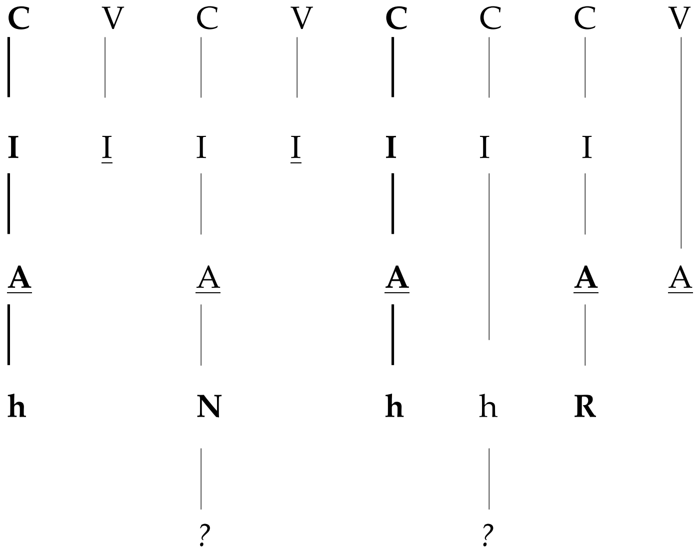
2.4 Acquisition of uneven representations

During acquisition, children are exposed to a plethora of different pronunciations of the same word. One of their main tasks is to develop categories and group tokens of the same word together. If a child hears [sinistɝ, sɲisra, sinisra, si˜str, səsðr, sɲisr, sənʲsrə, sʲəjistr, siɦistra, s˜sr, səʔisra, siʔisra…], she will be able, relatively soon, to abstract from the inherent variation of the acoustic signal and (a) identify the invariant – given in (25) – as well as (b), constructing a detailed UF corresponding more or less to the citation form |sinistra|, whose surface form will be /sinistra/ and whose acoustic/articulatory realizations may vary depending on a series of linguistic and extralinguistic factors. How does the invariant affect the SF /sinistra/? I argue that the invariant creates unevenness, or, to use a metaphor, an anti-democratic situation. In other words, not all the information contained in the SF is assigned an equal status. The elements that form part of the invariant are protected in a special way from any sort of phonetic erosion that could arise in running speech, whereas everything that has not received this special protection is likely to be modified, reduced or deleted under certain circumstances.

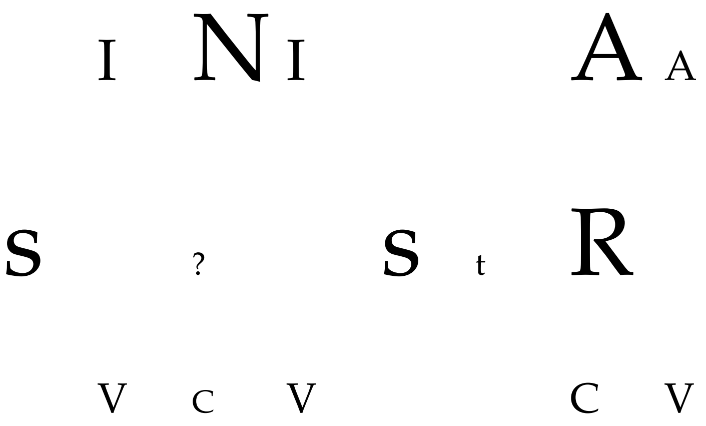
In (32) everything in the structure of /s/ is represented in bold, meaning that all the elements composing it, as well as the AL linking them to the

C-slot form part of the invariant and are, therefore, undeletable. In the structure of /n/, only **N** is bold, and in /r/ only **A** and **R**, but not, for example, their syllabic position and neither their **I** element. The occlusion element present in /n/ and /t/ is represented in italics since it is the weakest, i.e., the first to be dispensed with in running speech. Another way to represent the unevenness of the SF is given in (33).

(32) SF of *sinistra*



(33)



/s/ is a shortcut for **A-I-h**-C and /t/ for **A-h-?-**C. This uneven SF already predicts that from fully articulated [sinistra], reduction will proceed towards [sinisra, sinĩsra] (loss of occlusion), [s̃ᶮsr] (loss of syllabic affiliation for palatality and nasality) and [s˜sr] (loss of palatality). This is of course a necessary idealization, but the underlying idea is that the SF contains information about the strength of its components. This strength can be interpreted both as resistance to deletion and modification and as the activation level of components in lexical retrieval, i.e., hearing [s˜sr] facilitates the lexical retrieval of |sinistra| much more than, say, hearing [iita]. The presence or absence of certain components in the invariant can be explained on different grounds. For example, it is to be expected that onset consonants will be less likely to be lenited than coda consonants, therefore they will appear in a larger number of pronunciation variants and they will have higher chances of being stored as part of the invariant. The strength of onset consonants has both a phonetic and a phonological explanation. Phonetically, consonants rely on both external and internal cues for their recognition. Internal cues consist, among others, of the release burst for stops and the partially obstructed airstream for fricatives, whereas the most salient external cues are the formant transitions on the following vowel. Phonologically, according to the Coda Mirror Theory (Scheer & Ségéral 2001), non-intervocalic onset consonants are strong because they are licensed by the following vowel but ungoverned[7]

---

[7] This generalization is valid only for certain languages, among which Italian and English. Ségéral & Scheer (2001) distinguish between languages where fortition takes place word-initially and post-consonantally and the only allowed word-initial clusters rise in sonority, and on the other hand, languages where, at the left edge of the word, everything is permissable. The first group of languages are called CV-initial, because they assume that the morphological boundary standing for "beginning of the word" is

(licensing being a corroborating force that strengthens the structure of the consonant and government being a destructive force that spoils the nature of its target). Intervocalic consonants, however, even if they are in onset position, are heavily affected by the surrounding vowels and tend to lose some of their consonantality, e.g., typically, stops lose occlusion and fricativize (e.g., /k/ → [x]), voiceless obstruents become voiced (e.g., /t/ → [d]), fricatives lose manner elements and become approximants (e.g., /ɣ/ → [ɰ]), etc. Therefore, it is not surprising that the initial /s/ of *sinistra* is particularly resistant whereas intervocalic /n/ tends to become a nasal flap or a placeless nasal. What is more striking, perhaps, is the preservation of word-internal /s/, since it is both in a coda position and precedes a consonant. However, it is well known that stridents, and especially /s/, are acoustically highly salient and their internal cues are rich enough to be distinguished correctly even in adverse environments. Another force that conspires against the preservation of segments is the reduction of articulatory effort (Ernestus 2000, Kirchner 2004). In running speech, speakers aim to optimize their performance, reducing the number of articulatory gestures in order to communicate faster and with a lesser effort. Accordingly, segments that require a greater deal of precision for their articulation are particularly likely to be simplified or omitted. Whilst the articulation of stridents certainly requires much greater precision than the articulation of vowels, speakers, oddly enough, readily dispense with vowels, but do not get rid of stridents as easily[8]. Thus, among the many pronunciation variants of *sinistra* we find [snstr] but not *[initra]. I argue

---

translated into phonology as an empty CV sequence (Lowenstamm 1999).

[8] However, there are well-known cases of /s/ being lenited to [h], e.g., Ancient Greek (even word-initially), American Spanish.

that the reason for this lies both in the acoustic salience of /s/ and in the predictability of vowels in comparison to consonants. Even if salience is perception-based, speakers are somehow aware that omitting an acoustically salient segment would lower the possibilities of word-recognition in a significant way for the listener. On the other hand, as much as vowels are among the loudest segments, they normally contribute to lexical distinctiveness to a considerably lesser extent than consonants. As a matter of fact, "[t]here is an inherent property of speech that seems at first sight paradoxical: most of the sound energy is concentrated in vowels, while most of the linguistically significant information is concentrated in consonants" (Harris 2006:1491). One piece of evidence for this comes from the very existence of consonantal writing systems (such as Semitic abjads, see Daniels & Bright 1996) vs. the absence of attested vocalic writing systems. Moreover, it is a quite common practice to transcribe only consonants and to ignore vowels in order to speed up the writing process, e.g., English <mmt> for <moment>, Spanish <tmb> for <también> 'also', Italian <cmq> for <comunque> 'anyway' (Baroni 2011:148, Baroni 2013a:43). Generally speaking, languages tend to have more consonants than vowels and the discrimination of consonants is more categorical than that of vowels (Maddieson 2011). In sum, it appears much easier to understand a sentence in which vowels are missing (both in written and spoken language) rather than the other way around, cf. <th bk s n th tbl> vs. <e oo i o e ae> for <the book is on the table> or [ð bk z n ð tbł] vs. [ə ʊ ɪ ɒ ə eɪ] for [ðə bʊk ɪz ɒn ðə teɪbł]. Speakers tend to reduce articulatory effort but at the same time aim to maintain lexical distinctivity. Casual speech phenomena in many languages display the tendency to preserve consonants substantially more often than vowels (see Dalby 1986, Shockey 2003 for English, Gnerre 1976 for Italian, Siptár 1991 for Hungarian, Ernestus 2000 for Dutch). The very existence of languages whose phonotactics allow for sequences of many consonants may be interpreted as the phonologization of fast speech phenomena over time (Blevins 2004:126). One of the most typical examples is the disappearance of *yers* in Slavic languages that gave rise to a great number of consonant clusters. Comparatively, languages exhibiting sequences of more than two vowels are very much rarer. Blevins (2004:213) reports only Gilbertese, a Micronesian language, which allows tautosyllabic VVVV sequences, e.g., the augmentative suffix *–kaaei.* However, one must be cautious here since not all languages display the same pattern. Spoken Danish, for example, has cases of extreme reduction where consonants are heavily lenited, to the point that they become approximants, syllabic approximants, glides, vowels or disappear. Two examples taken from Basbøll (2005:295) are presented in (34a-b).

(34)

(a)

mave ‘stomach’

Lentissimo form (artificial, spelling pronunciation): [ˈmæːvə]

Lento form (rare): [ˈmæːʊ̯ə]

Spontaneous speech: [ˈmæːʊ]

Allegro form: [ˈmæːʊ̯ʊ]

(b)

givet “given”

Lentissimo: [ˈɡ̊iːvəð]

Lento: [ˈɡ̊iːvð]

Moderate: [ˈɡiːəð]

Allegro: [ˈɡ̊iːð̩]

Allegrissimo: [ˈɡiːðð̩]

The spontaneous pronunciation of some words often implies sequences of vowels that may be realized as extra-long vowels, e.g., *vare* ‘goods’ [vɑːɑ], *hårdere* ‘harder’ [hɒːɒɒ]. Cases of extreme reduction can give rise to vowels of impressive length, e.g., *hårdere at åre(lade)* ‘harder to bleed’, careful (artificial) pronunciation: [ˈhɒːɒɒɒˈɒɒˌlæːðˀð̩], reduced: [ˈɒːɒɒɒˈɒːɒ(ˌlæːðˀð̩)].

2.5 Perception, production and unevenness

The unevenness of the SF is caused by variation in perception: the input is variable and therefore the components of the SF have different degrees of representational strength. The first time we hear [sinistra] all the elements are stored and probably all have the same status, even though it is likely that the stressed vowel [i] and the stridents are already prominent with respect to other components. Subsequently, every time we hear a new token of *sinistra,* the strength of some element will increase or decrease, since we are not able to remember sensory events for a long time if we are not exposed to them continuously. At a first stage, the strength of initial [s] might be, say, 1, and after 1000 tokens, all containing acoustic correlates of [s], its strength will be 1000, and so on. A similarly strong status will be assigned to nasality and rhoticity, whereas palatality will be slightly weaker and other elements will be even less strong.

(35)

Stage 1: **s** i n **i s** t r a

Stage 2: s i n i s t r a

Stage 3: s i n i s t r a

Stage 4: s i n i s t r a, etc.

We should not expect random speech errors or slips of the tongue to affect the SF significantly, since, as I already stated, memory decays. Imagine that we hear someone utters [finistra] or [sinihtra]; these tokens will be stored temporarily in our short term memory but will soon decay as it is very unlikely that we will hear such pronunciations again anytime soon. But how does unevenness affect production? Quite simply, unlike other OT models accounting for variation, I propose to switch the burden from constraint re-ranking to the SF. The uneven SF already provides information for the constraint ranking, since there is a fixed, universal hierarchy consisting of MAX-INVARIANT > MARKEDNESS (where markedness here stands for sensorimotor, perceptual and articulatory constraints). Both MAX-INVARIANT and MARKEDNESS are emergent from the experience with the real world, the former from the invariant and the latter through physical phenomena such as ease of articulation and perceptual factors. MAX-INVARIANT is a peculiar type of constraint, since its position in the hierarchy is always fixed and undominated. Unlike other constraints, like MAX-I, MAX-U, *CC, etc., it is word-specific. It refers to the strength values stored in the SF rather than to melody or syllabic structure in general. Given a ranking such as MAX-INVARIANT > NOCODA >

FAITH, we would expect a language to contain no words with coda consonants, since NOCODA dominates all faithfulness constraints. However, NOCODA – or any other constraint – will never be able to dominate MAX-INVARIANT. Therefore, if a coda consonant, by virtue of its inherent characteristics, such as acoustic salience, has a very strong status in the SF and forms part of the invariant, it will surface even though NOCODA is ranked so high in the hierarchy[9]. Grammars are not faithful in a completely arbitrary way, they are faithful to the strong, and what is strong is such either because it is frequent or because it is salient, and, the majority of the time, what is frequent is such because of salience – though assimilatory processes, ease of articulation, etc. also have a part to play. It is noteworthy that here by frequency I do not mean absolute frequency, since that would make wrong predictions (e.g., it would predict coronal stops to be particularly resistant, since they are very frequent, which is not the case). Frequency here is intended as the frequency of a given phonological unit within the range of pronunciation variants of a single word. Put differently, if the UF of a word contains the segments *x* and *y* and *x* is expressed phonetically in the AF/ArtF more often than *y*, the status of *x* in the SF will be stronger than the one of *y*, and consequently, *x* will be more resistant to lenition, deletion and assimilation.

---

[9] If in this hypothetical language NOCODA dominates DEP, then the solution will simply consist in the insertion of an epenthetic vowel after the consonant.

3

DEFINING SALIENCE

So far I have used the term "acoustic salience" to refer to the perceptibility of certain sounds but I have not yet defined it properly. Quite simply, I employ the term salience in place of what Henke et al. (2012:72-73) suggest calling Cue Robustness and Cue Precision. They define the former "as the degree to which the presence of a segment, and that segment's contrastive information, is likely to be apprehended by a listener under normal listening conditions" and the latter "as the degree to which the cue narrows the field of segmental contenders". Cue Robustness is basically the overall audibility of a segment considered in isolation, e.g., the cues of [s][10] are more robust than those of [t], whereas Cue Precision depends on the number of the segmental contenders. In a language with many sibilants[11] in its phonological inventory, such as Polish, [s] might be more difficult to distinguish from [ʂ] and [ɕ], than [t] from [p] and [k]. There is, therefore, an absolute salience, based on spectral acoustic characteristics, and a language-specific salience, depending on the number of similar segments in a phonological inventory. In addition, salience is also

---

[10] In this chapter acoustic salience will be dealt with. Consequently, when referring to acoustic properties, phones (not phonemes) will be the main object of study (hence the use of brackets instead of slashes).

[11] Unless otherwise stated, I will use the terms "sibilant" and "strident" interchangeably referring to the class of /s/-like sounds, which comprises [s, z, ʃ, ʒ, ʂ, ʐ ɕ, ʑ, ɬ, ɮ] and their variants.

contextual, depending on modulation, i.e., the amplitude and the spectral change triggered by a combination of sounds (Ohala & Kawasaki-Fukumori 1997). [s] might be highly salient in Spanish, where it is the only sibilant and less salient in Polish, which has three types of sibilants; very salient before a vowel, a sonorant or a stop but less audible before another fricative. Consonantal salience is particularly interesting to consider because it is able to explain certain synchronic phonotactic patterns that are likely to have diachronic causes. If casual speech phenomena can become phonologized over time and give rise to synchronic phonological processes or rules, such as nasal place assimilation, then consonant clusters resulting from the disappearance of unstressed vowels in running speech were probably preserved only under certain circumstances. I propose that, given a #C_1C_2V sequence, where C_1 and C_2 are of equal sonority, C_1 must be more salient than C_2 in order to be preserved, since C_2, being adjacent to a vowel, can rely on formant transitions for its recognizability, whereas C_1 cannot. Similarly, in a VC_1C_2# sequence, the consonant that is further from the vowel must be the more salient one. By salient here I mean absolutely salient, not language-specifically or contextually salient, although I expect contextual effects, such as the place and manner of articulation of C_1 and C_2, to affect the chances of survival of a given cluster. In this chapter I will deal in particular with plateau clusters, i.e., sequences of consonants of equal (or almost equal) sonority, because their well-formedness cannot be determined on the basis of sonority. It is well known that sonority-rising clusters are preferred word-initially (#OR) and sonority-falling clusters are preferred word-finally (RO#). However, little has been said about obstruent clusters (OO, including FT, TF, TT, FF), nasal clusters (NN) and liquid clusters (LL). It has been suggested that OCP effects, such as the avoidance of a sequence of consonants with the same place or manner of articulation, may partially explain certain patterns, e.g., the relative rarity of /fp, xk, kx, etc./. Markedness has also been invoked, for example, by Morelli (1999). In her study on obstruent clusters, she argues that /s/O is the preferred initial cluster cross-linguistically because an unmarked sequence must display an alternation of [+continuant] and [–continuant] segments (as much as an optimal syllable would be a sequence of stop, which is [–continuant], and a vowel, which is [+continuant]) and because /s/ is the unmarked continuant obstruent, given that is coronal and [–anterior]. Another possibility is to refine the Sonority Hierarchy in such a way that consonants with the same manner of articulation can be distinguished in sonority by their place of articulation. For example, both Zwicky (1972b) and Tsunoda (2008) propose the following sonority scale for nasals: /n > m > ŋ/, based on data from, respectively, English casual speech and Warrongo phonotactics. I intend to show, however, that markedness, sonority and modulation are not able to account for all plateau clusters and that taking salience into account can help shed some light on the matter.

3.1 Markedness-driven account

Markedness is rather difficult to define. It has been used with several different meanings, depending on the author and the theory. Nevertheless, there are some basic ideas shared by most scholars: (a), markedness is relative, i.e., a segment is not absolutely marked but is relatively more or less marked in comparison with another segment; (b), coronal is the unmarked place of articulation for consonants (Paradis & Prunet 1991), although some dispute this, suggesting it is in fact velar (Brandão de Carvalho & Trifit 2013), e.g., in standard Element Theory (Kaye et al. 1985), velar consonants are empty – they contain no melody whatsoever. On the other hand, other phonological theories, especially those based on OT, argue that dorsals and labials are more marked than coronals and dorsals are more marked than labials (Mohanan 1993, Jun 1995, De Lacy 2006); (c), segments that are less marked assimilate their place of articulation to those that are more marked, e.g., in Dutch and English, /n + k, p/ → [ŋk, mp] but /m, ŋ + t/ → [mt, ŋt]. When it comes to neutralization in weak positions, such as the coda, the predictions made by markedness are more controversial, since typically consonants, if reduced, become glottal and not coronal, e.g., British English *pick, keep, meet* [pʰɪʔ, kʰiːʔ, miːʔ], American Spanish *estar* 'to be, to stay', *mes* 'month' [ehtar, meh]. However, there are languages in which [k → t] and [p → t] processes are attested: Taiwanese secret language (Li 1985), Basque (Hualde 1991), Cantonese secret language (Yip 1982). It is reported that in some varieties of Italian (e.g., Veneto, Mioni 2000) as well as in Huallaga Quechua (Weber 1989), Kagoshima Japanese (Haraguchi 1984), Seri (Marlett 1981:20), etc., nasals neutralize to velar [ŋ], but as De Lacy (2006:39) argues, it is in fact a placeless nasal [N], which is perceptually very similar to [ŋ]. The constraint hierarchy proposed by De Lacy (2006:2) is therefore: *DORSAL > *LABIAL > *CORONAL > *GLOTTAL. As a matter of fact, the only consonants attested as epenthetic[12] segments are either

---

[12] The epenthetic segments considered here are independent from the phonological

coronal or glottal, with the notable exception of [w][13] (De Lacy 2006:79). As to whether markedness is just a diachronic product (a view advocated by, among others, Blevins 2006) or part of the synchronic competence of speakers is still being debated. When it comes to plateau clusters, markedness seems to be unexplanatory. It is undisputed that, word-initially TR clusters are less marked than TT, RT, RR, and that, word-finally, RT clusters are less marked than TR, TT, RR. However, if one focuses on TT, RR etc., markedness effects are hard to isolate. The most common obstruent clusters, /s/T sequences, are composed of a fricative and a stop, where the fricative is coronal. Tentatively, one could say that, in a word-initial plateau cluster, the first segment must be more marked than the second mannerwise but unmarked placewise. As unlikely as it sounds, this claim could still stand if there were not further data to confute it. However, I conducted some limited research on languages allowing word-initial plateau clusters, starting with the data collected by Morelli (1999), Parker (2012) and adding more material. 39 languages were examined, belonging to different language families: Abau (Austronesian, Lock 2007), Albanian (Indo-European, Newmark 1998), Ancient Greek (Indo-European, Steriade 1982), Cambodian (Mon Khmer, Nacaskul 1978), Classical Tibetan (Sino-Tibetan, Jacques 2004), Dakota (Siouan, Boas & Deloria 1972, Ullrich 2008), Dutch (Indo-European, De Schutter 1994), Eggon (Benue-Congo, Ladefoged & Maddieson 1986), English (Indoeuropean, Kenstowicz 1994), Georgian (Caucasian, Chitoran 1994,

---

context in which they occur. Cases like English *Thom*[p]*son*, French *cham*[b]*re* or substandard Italian *Is*[d]*raele* for *Israele* have nothing to do with epenthesis, they are better described as "emergent stops" (Ohala 1997).

[13] According to De Lacy, an epenthetic [w] can still be considered less marked than an epenthetic liquid, albeit being labiovelar, because glides are less marked than liquids.

Deprez 1988, Vogt 1971, Butskhrikidze 2002), German (Indo-European, Hall 1992), Hebrew (Semitic, Morelli 1999), Hindi (Indo-European, Nagamma Reddy 1987), Italian (Indo-European, Nestor 1993, Baroni 2012a), Khasi (Mon Khmer, Henderson 1976), Lithuanian (Indo-European, Tankeviciute & Strimaitiene 1990), Isthmus Zapotec (Zapotec, Marlett & Pickett 1987), Haida (isolate, Swanton 1910, Lawrence 1977), Havasupai-Walapai-Yavapai (Yuman, Redden 1966), Klamath (Plateau Penutian, Barker 1964:42-48), Margi (Chadic, Hoffman 1963, Ladefoged 1964), Mawo (Qiang, Hongkai 1986, Namkung 1996), Misantla Totonac (isolate, MacKay 1994), Mitla Zapotec (Zapotec, Greenberg 1978) Nisgha (Tsimshianic, Tarpent 1989), Pame (Otomi, Gibson 1956), Pashto (Indo-European, Penzl 1995), Polish (Indo-European – Gussman 2007), Russian (Indo-European – Halle 1959), Santa María Quiegolani Zapotec (Zapotec, Regnier 1993), Serbo-Croatian (Indo-European, Hodge 1946), Seri (Hokan, Marlett 1988), Sobei (Austronesian, Sterner 1975), Telugu (Dravidian, Nagamma Reddy 1987), Teribe (Chibchan, Oakes 2001), Tsou (Austronesian, Wright 1996), Wichita (Caddoan, Rood 1975), Yatee Zapotec (Zapotec, Jager & Van Valin 1982), Yuchi (isolate, Wolff 1948, Crawford 1973).

Unfortunately, it is quite hard to draw decisive conclusions about which consonant sequence is more or less marked word-initially. I could only formulate the following implicational correlations:

(36)

O = any obstruent, T = any stop, P = any labial obstruent, K = any dorsal obstruent, F = any fricative, S = any strident, H= any non-strident fricative, N = any nasal, L = any liquid, G = any glide.

(a) if a language allows OO, then it allows SO. Exception: possibly Margi.
(b) if a language allows TT, FF, HT, then it allows SO. Exception: possibly Margi.
(c) If a language allows HT, then it allows TT. Exceptions: Mawo, Pashto, Walapai.
(d) If a language allows, TH, then it allows HT. Exception: Modern Hebrew.
(e) If a language allows FF, then it allows TT. Exceptions: English, Pashto, Walapai.
(f) If a language allows NN, LL and/or GG, then it allows OO. Exception: none.
(g) If a language allows OP, then it allows KO. Exception: Wichita.
(h) If a language allows /nm/, then it allows /mn/. Exception: none.

(36a, b) seem to confirm the conclusions drawn by Morelli (1999). Margi may represent an exception to this tendency, but it actually depends on the phonological analysis. As a matter of fact, Ladefoged (1964) describes Margi as having a series of doubly articulated velar-labial stops but there is no agreement on whether /ps, pɬ, pç, p͡tʃ, p͡ts, bz, bɮ, b͜dz, b͜dʒ, ʔb, ʔbʷ, ʔb͜d, ʔd/ are doubly articulated segments or not and on whether /ts, tʃ, dz, dʒ/ are affricates or clusters. In the former case, Margi would simply be a language that does not allow clusters, whereas in the latter case, it would be the only language (attested so far) to allow O/s/ but not /s/O. (36c) states that a sequence of stops is less marked than a sequence of a non-strident fricative and a stop. Three languages contradict this tendency:

Mawo has /xp, xts, xtʂ, xtʃ, xtɕ, χp, χts, xtʃ, χtɕ/, Pashto has /xp/ and Walapai has /θp, θt, θk/ but none of them allow stop clusters. (36d) points out that the order non-strident fricative + stop is preferable to the reverse order. Nevertheless, note that Hebrew allows non-strident /x/ to occur as the second member of an initial cluster but not as the first. (36e) claims that a sequence of fricatives (stridents or non-stridents) is more marked than a sequence of two stops. Again, Pashto and Walapai go against this tendency, as well as English, which has /sf/ (although marginally) but no stop clusters. (36f) predicts that any sonorant cluster (nasal, liquid or glide cluster) is more marked than an obstruent clusters. As a matter of fact, I could find no exception to this point, although data on sonorant clusters are still very scarce and out of 39 languages in the survey, only 15 appear to allow such sequences. (36g) is one of the most interesting findings, because it gives insight into a preference of distribution based on place of articulation. Basically, it states that it is more marked for a labial consonant to be the second member of a cluster than it is for a dorsal to be the first. As a matter of fact, out of 39 languages in the survey, only 10 have C_1C_2 sequences where C_2 is labial while 24 allow C_1 to be dorsal. It is important to note that labial+dorsal and dorsal+labial clusters ought not to be confused with doubly articulated stops, i.e., /k͡p, g͡b/ are different from /kb, gb/. The only exception I could find is Wichita, which, however, does not have any labial obstruent. This preference seems very likely to be related with the tendency of child language described by Fikkert & Levelt (2008), according to which Dutch infants, while acquiring their phonology, pass through a stage where, in C_1VC_2 words, C_1 is always a labial and C_2 always a dorsal. As a result, a word like *kip* 'chicken' [kɪp], at this stage, is realized as [pɪk]. It has been noted that, in early child language, the first combinations of consonants with different places of articulation are labial + coronal (McNeilage & Davis 2000). This preference is explained by the fact that the sequencing of consonants goes from front to back across the word (Ingram 1974). Children would therefore prefer to begin a word with a labial because it only requires one jaw movement, without moving the tongue. Other studies (e.g., Fikkert et al. 2004, Davis et al. 2002) confirm that the preference for word-initial labials leaves a trace even in adult grammar. The complementary tendency, that of dorsals to occur finally, is less evident, at least regarding obstruents, but there are many languages in which velar [ŋ] is restricted to occurring only syllable-finally (Anderson 2011). The preference for words to begin with a labial (rather than with a dorsal) and to end with a dorsal (rather than with a labial) could also be related to the order of acquisition of these sounds. As a matter of fact, labial and coronal consonants are normally produced by children earlier than dorsal consonants. In my survey of languages allowing plateau clusters, only 14 allow labial-dorsal or dorsal-labial sequences. Of these, three (Georgian, Teribe and Yatee Zapotec) only have labial+dorsal clusters, three (Dakota, Khasi, Mawo) only have dorsal + labial clusters and the remainder (Cambodian, Eggon, Hebrew, Klamath, Nisgha, Santa María Quiegolani Zapotec, Seri and Tsou) have both. Albanian has sequences of labial + palatal, /fc, vɟ/. Eggon is notable for having clusters of complex velar-labial consonants followed by a velar, as /g͡bg, k͡pk/, e.g., *ō g͡bgā* ‘grind-3sg’ (Maddieson 1981:90). As the data at hand remain scarce, it is difficult to draw conclusions on these clusters. In general, clusters involving two non-coronal consonants are highly marked, but a certain sequential order does not seem to imply the opposite, i.e., if a language allows labials and dorsals to occur within the same cluster, it generally allows both /#kp/ and /#pk/ (and their variants)[14]. (36h) is a very tentative generalization on nasal clusters, basically arguing that /mn/ is less marked than /nm/, but I could find only one language allowing /nm/ word-initially, i.e., Tsou, therefore further data are needed in order to test this implicational relationship. (36h), if true, would be closely related to (36g), being basically an effect of the preference of labials to occur initially. Regarding liquid and glide clusters, these appear to be extremely marked, so much so that they are attested only in a handful of languages. Both Khasi (Henderson 1976) and Santa María Quiegolani Zapotec (Regnier 1993) have /rl/, the former has /jw/ and the latter /wj/. Klamath (Barker 1964:42-48) does not allow liquid clusters but word-initially /w/ can precede /j̊/ and /j’/ (a voiceless and a glottalized version of the palatal glide, respectively). My data on Mitla Zapotec (Greenberg 1978) and Classical Tibetan (Jacques 2004) are incomplete, but both are reported to allow /rl/ word-initially, although the liquids are said to belong to separate morphemes. I could not find any language with word-initial /lr/.

3.2 *Sonority-driven account*

The failure of sonority-based analyses of consonant clusters is so evident that it is the main reason why scholars have sought other ways to explain this kind of sequences. The most frequent consonant cluster, /s/C, is an

---

[14] In Pengo (Burrow & Bhattacharya 1970:82f) a metathetic process changes KP sequences into PK. However, Blevins & Garrett (2004:136) report that in Mokilese the reverse pattern is attested, e.g., /apkas/ → [akpas] (cf. Buckley 2011).

obvious violation of every sonority hierarchy proposed in the literature. As a matter of fact, "[sonority] becomes particularly unreliable in the case of consonant clusters involving fine-grained local differences in intensity. This especially true when strident fricatives are involved, since these can display higher intensity values than supposedly higher-sonority segments such as nasals" (Harris 2006:1485-1486). According to Selkirk (1984), fricatives are more sonorous than stops and therefore, they should follow, and not precede stops before a syllabic nucleus. This statement predicts that, cross-linguistically, C/s/ should be preferred to /s/C, but the situation is quite the opposite. Other analyses (Morelli 1999, Baroni 2012a) consider stops and fricatives to have the same sonority value, i.e., they form the class of obstruents, but still, even if in this view /s/C does not represent a sonority reversal, it is a sonority plateau, and the preference for /s/O over O/s/, O/f/, TT, etc. is not explained. A revised version of Sonority Scale able to accommodate /s/ in a consistent way with its phonotactic distribution is proposed by Basbøll (2005:173-201, 2012):

(37)
non-lateral non-stop sonorant > consonantal sonorant > voiced obstruent > voiceless obstruent with non-spread glottis > obstruent with spread glottis.

In Basbøll's view, stridents such as /s/ would bear the feature [+spread glottis], whereas non-strident fricatives (e.g., /f/) and non-aspirated stops would not. Applying this model to English phonotactics, it is possible to explain why word-initial voiceless stops are aspirated, e.g., *tap, cap, pad* [tʰæp, kʰæp, pʰæd], but they are not if preceded by /s/, e.g., *step, skip, speed* [stɛp, skɪp, spiːd]. Having a sequence of two [+spread glottis] segments word-initially would imply a sonority violation. One problem is that the presence of the feature [+spread glottis] in stridents is still a debated topic, especially because, despite being a very uncommon phenomenon, some languages possess aspirated stridents, such as /sʰ/ (e.g., Shuiluo Pumi, a Sino-Tibetan language, see Jacques 2011). Moreover, although in stop clusters aspirated segments tend to appear first, as in Cambodian (e.g., pʰt, pʰk, pʰc, tʰp, tʰk, cʰp, cʰk, kʰp, kʰt, kʰc), /s/ can precede an aspirated stop, as in /stʰ/ (Nacaskul 1978). The same occurred in Ancient Greek, which allowed /spʰ, skʰ/ and even /pʰtʰ, ktʰ/ (Steriade 1982).

Regarding other types of clusters, such as TT, FF, NN, etc., the effects of sonority are not that visible. In stop clusters, if we limit our analysis to the four main places of articulation (P = labial stop, T = coronal stop, J = palatal stop, K = dorsal stop), the preferred sequence seems to be PT (15 out of 39), followed by TK (11/39), PK and KT (9/39).

(38)

PT (38.5%)> TK (31%) > PK, KT (23%)> PJ, KP (20.5%)> TP, KJ (12.8%)> JK (10.2%) > JT (5.1%)> JP (2.5%).

These results appear to be consistent with the hypothesis that labial would be the somehow most consonantal place and dorsal the most vocalic one, with coronal in the middle (cf. Botma 2006). As a matter of fact, it is preferable for P to occur before T and K and for K to occur after P and T, even though the sonority distance between K and T must not be as significant as that between P and K, T. Nothing decisive can be said about

J, except that it occurs much more often after P (21%) than before (2.6%). Looking at the languages of the sample for which data were reliable, I obtained the following results:

(39)
Φ= labial fricative, X = dorsal fricative, θ = non-sibilant coronal fricative
SK (93.75%) > ST (90.6%) > SP (87.5%) > PT (59.4%) > PS (53.1%) > KS (50%) >TK, KT (34.4%) > PJ (28.1%) > PK (25%) > SΦ, SJ, KP (22%) >TS, XP (18.75%) > TP (15.7%) > other clusters (less than 12.5%).

The numbers tell us that most languages with word-initial obstruent clusters allow SO sequences. The most common obstruent clusters not beginning with a sibilant begin with a labial stop. Among SO sequences, SK is the commonest, followed by ST and SP. Tentatively, SK might be the most favored SO cluster because it starts with a sibilant and it does not violate neither OCP[coronal], as ST does, or the constraint militating for P to be initial, as SP does. It is possible to formalize this tendency in OT considering the following markedness constraints and ignoring all faithfulness constraints except LINEARITY :

(40)

LABIALFIRST	Consonants bearing the feature [labial] (i.e., containing the element **U**) must be syllable-initial.

OCP[PLACE]	Assign a violation for every sequence of segments that share a place feature (i.e., characterized by the same element).
LINEARITY	No metathesis.

Tableau 10: SK vs. ST, SP

/spV/	LINEARITY	LABIALFIRST	OCP[PLACE]
a) [spV]		*!	
b) [psV]	*!		
c) [stV]			*!
☞d) [skV]			
/stV/	LINEARITY	LABIALFIRST	OCP
a) [spV]		*!	
b) [psV]	*!		
c) [stV]			*!
☞d) [skV]			
/skV/	LINEARITY	LABIALFIRST	OCP
a) [spV]		*!	
b) [psV]	*!		
c) [stV]			*!
☞d) [skV]			

Tableau 10 shows that the hypothetical constraint ranking LINEARITY > LABIALFIRST > OCP would not let any SO clusters other than /sk/ surface, since /sp/ fatally violates LABIALFIRST, and the solution of inverting the order of the segment is not viable, since LINEARITY is undominated. /st/ is ruled out because it violates OCP. Therefore, there seem to be good reasons for the higher frequency of SK with respect to ST and SP.

As for nasal stop clusters, it has already been pointed out that the few languages that allow them prefer /mn/, which is consistent with the behavior of oral stops, i.e., /mn/ is an example of nasal PT. Liquid clusters are so rare that it is highly speculative to say anything about them. However, the absence of /lr/ as opposed to /rl/ is problematic if one assumes that rhotics are always more sonorous than laterals. According to Pons Moll (2008), rhotics occupy different positions on the sonority scale depending on whether they are taps, flaps or trills, with trills being less sonorous than laterals. Assuming that in Khasi, Classical Tibetan and Zapotec languages /r/ represents a trill, then, the sonority scale remains unviolated[15]. Pons Moll's proposal is challenged by the behavior of liquids in Italian, in which /r/ is undisputedly a trill in the standard but is evidently more sonorous than /l/[16]. For example, /Vr.lV/ is a good syllable contact but */Vl.rV/ is not. The latter goes against Vennemann's Contact Law according to which "[a] syllable contact A$B is the more preferred, the less the Consonantal Strength of the offset A and the greater the Consonantal Strength of the onset B" (Vennemann 1988:40). In other words, the onset consonant of the second syllable must be less sonorous than the coda consonant of the first syllable.

---

[15] Jacques (p. c.) claims that in Classical Tibetan /r/ must have been a trill. He also points out that in modern Rgyalrong languages word-initial /#rl/ exists, but /r/ is phonetically realized as [ʐ], sometimes with a trilled component that makes it similar to Czech *ř*.

[16] A possibility is to posit that Italian /r/ is realized as a trill in the onset and when geminate and as a tap intervocalically and in the coda. However, Bertinetto & Loporcato (2005:133) claim that before a consonant /r/ usually consists of a "double [linguopalatal] contact".

3.3. Beats-and-Binding Phonology and the Net Auditory Distance

In order to refine some aspects of the Sonority Hierarchy considered so far to be unsatisfactory, Dziubalska-Kołaczyk (2002, in press) proposes Beats-and-Binding Phonology (B&B henceforth), a syllable-less theory couched in the framework of Natural Phonology (Stampe 1979, Donegan 1978, Donegan & Stampe 1979, Dressler 2009). In B&B, what is traditionally called nucleus corresponds to the *beat* (B) and everything else is just a *non-beat* (n). Relationships between beats and non-beats are called *bindings.* Phonotactics is governed by the NAD (Net Auditory Distance), which involves three factors: Manner of Articulation (MOA), Place of Articulation (POA) and voicing (Lx). In its original form, NAD is defined in the following way: |MOA| + |POA| + |Lx|, "where |MOA|, |POA| and |Lx| are the absolute values of difference in the Manner of Articulation, Place of Articulation and Voicing of the neighboring sounds, respectively" (Dziubalska-Kołaczyk 2009:56). B&B makes finer predictions than traditional Sonority Hierarchy-based theories, e.g., it shows that /brV/, /grV/ are better formed than, say, /drV/, because the NAD of the former is greater than the NAD of the latter:

(41) C1C2V is well-formed iff NADC1C2 ≥ NADC2V

The way NAD is calculated relies on the fact that, in general, it is better for neighboring sounds to differ maximally in MOA, POA and Voicing[17].

---

[17] In the latest version of the theory (Dziubalska-Kołaczyk, in press), Voicing is dispensed with, since it is redundant for sonorants.

(42) Example of calculation of NAD of the sequence /trV/, where V = any vowel (Dziubalska-Kołaczyk 2009:60-61, see Table 1 for the numeric values).

C1 = (MOA1, POA1, Lx1) = /t/

C2 = (MOA2, POA2, Lx2) = /r/

V = (MOA3, Lx3)

/t/ = (4, 2, 0), /r/ = (1, 2, 1), V = (0, 0, 1)

NAD(C1, C2) = ǀ4 -1ǀ + ǀ2 – 2ǀ + ǀ0 – 1ǀ = 3 + 0 + 1 = **4**

NAD(C2,V) = ǀ1 – 0ǀ + ǀ1 – 1ǀ = **1**

4 > 1 = OK.

Table 1: MOA and POA values (Dziubalska-Kołaczyk 2009).

4	3	2	1	0	
obstruent		sonorant			
stop	fricative	sonorant stop	approximant	V	
	affricate		glide		1
p b	ɸ β f v	m ɱ	w	Labial	2
t̪ d̪ t d ʈ ɖ	θ ð s z	n̪ n	r l	Coronal	3
k ɡ c ɟ	ɕ ʑ x ɣ	ɲ ŋ	j	Dorsal	4
				Radical	5
ʔ	h			Glottal	6

Table 2: MOA and POA values (Baroni 2012a).

4			3	2		1	0	**MOA**		
Obstruents			**Sonorant**							
T	Affri–cate	F	N	L		G	V		**POA**	
				Lat	Rh					
p b	pɸ bβ	ɸ β	m			β̞	ü	Bila–bial	**Lab**	3
	pf bv	f v	ɱ			ʋ		Labio–dental		
t d	ts dz tθ dð	s z θ ð	n	l	r		a[18]	Dental/ Alveo–lar	**Cor**	0
c ɟ	tʃ dʒ cç ɟʝ	ʃ ʒ ç ʝ	ɲ	ʎ		j	i	Post-alveolar/Palatal	**Dors**	1
k g	kx gɣ	x ɣ	ŋ			ɰ	ɯ	Velar		2
						w	u	Labio–velar		2.5

In Baroni (2012a) the model is modified in order to enhance its predictive power (see Table 2). Obstruents are assigned the same MOA value (= 4) and POA values are changed in order to account for some coarticulation phenomena. The following numeric values are assigned: 0 to dental/alveolar, 1 to palatal and 2 to velar. These values are able to predict that palatal sounds are disfavored both after dental/alveolar and velar

---

[18] The place of articulation of /a/ is not dental/alveolar, but perception-wise, [a]-like vowels show a significant interaction with many coronal consonants.

sounds since 1 is equidistant from 0 and 2. Dental, alveolar and velar sounds can be palatalized, so it would be somehow problematic to group palatal sounds together with either dental/alveolar (as coronal) or velar (as dorsal). Labial sounds are assigned 3. This might seem to be in contradiction with the actual shape of the vocal tract (the lips are closer to the teeth than to the velum) but a series of universal facts (velar vowels tend to be rounded, labiovelar co-articulations are the most frequent, labial sounds are less likely to undergo assimilation, etc.) justify this choice. Labiovelar sounds are assigned a value between 2 and 3, i.e., 2.5. However, both versions of the model (Dziubalska-Kołaczyk 2002 and Baroni 2012a) treat all plateau clusters as equally ill-formed. Therefore, the concept or relative salience of a segment within its natural class is introduced in order to explain the behavior of these clusters (Baroni 2012a:55). Plateau clusters, in order to survive the overwhelming tendency towards CV sequences, have to meet certain requirements: one of the two consonants must be the most salient within its natural class (that is, a sibilant among obstruents, a labial among nasals or a rhotic among liquids), the least salient consonant is preferably the one adjacent to the vowel, and configurations where two consonants are similar in their degree of salience are avoided (e.g., /st/, where /s/ is maximally salient and /t/ is minimally salient, is predicted to be better-formed than /pt/, where both segments are poorly salient and than /fk/, where both segments are relatively salient). The calculation of NAD can be seen as the combination of the Sonority Hierarchy with the OCP principle (Leben 1973) applied to consonant clusters. However, it is necessary to distinguish between coronal and non-coronal here. As a matter of fact, NAD values predict correctly that, for instance, /fp/ and /xk/ are ill-formed (NAD = 0) but treats sequences such as /st/ in the same way (NAD = 0), although it is evident from our data that /st/ is a very common cluster whilst /fp/ and /xk/ are very rare (to be more precise, /fp/ is unattested in my survey and /xk/ only occurs in Seri, see Marlett 1988).

3.4 The other side of the syllable: coda clusters

In the phonological literature on consonant clusters, very little attention has been paid to coda clusters if compared with onset clusters, let alone final plateau clusters. Morelli (1999) examines the phonotactic of word-initial obstruent clusters, but not of final ones. Intuitively, it appears that differing in place of articulation may be a good thing word-initially, where it is easier to maintain contrast, and less so word-finally, where there is a tendency towards place neutralization to coronal or glottal. We owe one of the few works dealing with both initial and final clusters to Greenberg (1978), who makes the following statements about final clusters:

(43)
a) TT# implies FT# (p. 254).
b) FF# implies FT# or TF# (p. 255).
c) /rl/ is the only possible liquid sequence both word-initially and word-finally (p. 257).
d) NL# and NN# imply LN# (p. 262).
e) NN# implies NO# (p. 267).
f) CC# implies CO_{cor}# (p. 268).

I will now briefly comment on (43). (43a, d, e) can be considered examples of sonority effects: it is better to have a sonority fall word-finally (FT#, LN#, NO#) than a sonority plateau (TT#, NN#) or a sonority reversal (NL#). (43b) points out that a sequence of fricatives is highly marked or at least more marked than a sequence of obstruents differing for continuancy. (43c) is hard to explain, since we would expect the initial and final positions to display opposite preferences, whereas /rl/ is the only attested order in both positions. The absence of /lr/ might have an articulatory basis (e.g., difficulty in pronouncing [r] after [l]) or a perceptual one (e.g., likeliness to hallucinate a stop between [l] and [r], such as [d][19]). (43f) states that the most unmarked segment to appear as the last member of a word-final cluster is a coronal obstruent. Pizzo (2009) is the only work, to my knowledge, that describes directly and specifically the typology of word-final clusters from a cross-linguistic perspective. According to Pizzo's survey, the most unmarked word-final sequence is composed of a sonorant followed by an obstruent, which is unsurprising. Therefore, if a language allows other kinds of clusters, then it will surely allow L/N+O clusters. The second least marked sequence is OO#, followed by sonorant clusters (e.g., NL#, LN#, NN#, LL#) and by O+N/L clusters. Put otherwise, a sonority fall is better than a sonority plateau, which in turn is better than a sonority reversal. Among plateau clusters, obstruent clusters are preferred to sonorant clusters. So far, word-final clusters do not seem to differ significantly from word-initial ones, except that sonority is expected to decrease rather than increase. In order to compare word-final clusters with their word-initial counterpart and to test Greenberg's

---

[19] In fact, the results of the experiment which will be presented in the following paragraphs will show that listeners tend to hear [ldr] when exposed to input [lr].

(1978) predictions, I focused on 15 languages allowing final plateau clusters: Armenian (Indo-European, Vaux 1997, 2003), Basque (isolate, Hualde 1991), Czech (Indo-European, Slavic, Bičan 2011), Dutch (Indo-European, Germanic, De Schutter 1994), English (Indo-European, Germanic, Kenstowicz 1994), French (Indo-European, Romance, Klausenburger 2012), Hungarian (Uralic, Finno-Ugric, Törkenczy 2006), Klamath (Plateau Penutian, Barker 1964:42-48), Lebanese Arabic (Semitic, Kenstowicz 1994), Lezgian (Nakh-Daghestanian – Haspelmath 1993), Ojibwe (Algic, Artuso 1998), Pame (Oto-Manguean, Gibson 1956), Serbo-Croatian (Uzelac 1971, Simonović 2013), Seri (isolate, Marlett 1988), Turkish (Turkic, Clements & Sezer 1982:245). The sample is of course too small to formulate absolute generalizations, yet some interesting tendencies emerge.

(44)
a) sonorant clusters imply obstruent clusters, i.e., there is no language allowing word-final NN#, LL# that does not allow OO# as well.
b) If TH#, then HT#, i.e., if a language allows a sequence of a stop followed by a non-strident fricative, then it allows the reverse order as well.
c) If OP#, then OK#, i.e., if a language allows a sequence of obstruents whose last member is a labial, then it allows a sequence of obstruents whose last member is a dorsal.
d) If TH# or HT#, then SO# or OS#, i.e., if a language allows final clusters containing a stop and a non-strident fricative, in any order, then it allows a sequence of a strident and an obstruent, in any order.

e) If FF#, then TT#, i.e., if a language has a word-final fricative cluster, then it has word-final stop clusters. One exception is Lezgian (Haspelmath 1993) that has /fs#, xʃ#/ but no stop clusters.

f) If OO#, then OS# or OT_{cor}#, i.e., if a language has word-final obstruent clusters, then it has obstruent clusters where the last segment is a strident or a coronal stop. One exception might be Ojibwe (Artuso 1998), that is said to allow only /ʃk#/.

g) If NN#, then N/n/#, i.e., if a language has word-final nasal clusters, then the second nasal is a coronal. This generalization is based on the only three languages of the sample allowing final nasal clusters, i.e., Pame, French and Armenian. Pame has /ŋn#/, French and Armenian /mn#/, although in the latter language /mn#/ is realized phonetically as [mən] (Vaux 2003).

h) If LL#, then /rl/, i.e., the only possible word-final liquid cluster is a sequence of a rhotic followed by a lateral. Languages allowing this type of complex coda are English, French, Hungarian and Czech.

In general, Greenberg's (1978) findings appear to be confirmed. Unsurprisingly, (44c) brings further evidence of the universal preference for dorsals to occur word-finally and for labials to avoid this position. All the languages of the survey either have OK# or OT_{cor}# clusters, whereas only eight allow OP# and most of them are actually SP# (only Seri has final OP# clusters where O is not a strident). Generalizations about word-final plateau clusters are formalized in (45) and (46).

(45)

The preferred word-final plateau cluster consists either of a strident followed by a stop or by a stop followed by a strident. If the first segment is not a strident, then the preferred sequence is an obstruent followed either by a coronal or a dorsal stop.

(46)

SO#, OS# > OT_{cor}#, OK# > all other OO# clusters > sonorant clusters.

It is noteworthy that if word-initially the preferred obstruent cluster is #SO, word-finally the expected mirror sequence OS# is not always allowed. Most languages with final obstruent clusters allow both OS# and SO#, but Basque and Turkish only permit OS# and Ojibwe and Pame only have SO#. Serbo-Croatian (Uzelac 1969) does not have OS# in the native vocabulary, but allows /ps#/ in borrowings.

With regards to stop clusters, Turkish and Ojibwe do not allow them at all. The other languages of the survey display the preferences shown in (47-48). A comparison between the preferences of word-initial and word-final clusters is presented in table 3.

(47)

KT (10/15) > PT (9/15) > TK (5/15) > PK, JK, TP (2/15) > KP (1/15).

Considering all the clusters, the results are the following:

(48)

ST (11/15) > SK, KS, KT (11/10) > PS (9/15) > PT (8/15), TS, ΦT (7/15), SP (6/15) > other clusters (less than 5/15).

Table 3: Comparison between word-initial and word-final plateau clusters.

Word-initial	*Word-final*	*Comments*
OO > NN, LL, GG	OO > NN, LL, GG	OO is the preferred sequence in both positions.
SO > OS, TT, FF, HT	SO, OS > TT, FF, HT	Word-initially, OS implies SO (perhaps with the exception of Margi), but word-finally there is no preference between the two.
HT > TH	HT > TH	HT is preferred to TH in both positions (with the exception of Hebrew). Note that in English, sequences like O/θ/ are always bimorphemic, as in *dep-th, eight-th*, etc.
TT > HT	TT > HT	TT is preferred to HT in both positions (with the exception of Lezgian, Mawo, Pashto, Walapai).
TT > FF	TT > FF	TT is preferred to FF in both positions (with the exception of Lezgian).
PT > TK > PK, KT	KT > PT > TK	In word-initial stop clusters, the first segment is preferably labial, the second one either coronal or dorsal. In word-final stop clusters, the first segment is preferably dorsal, the last one either dorsal or coronal. In general dorsals and labials prefer not to belong to the same cluster and labials prefer

		not to follow any stops.
SK > ST > SP > PT > PS > KS	ST > SK, KS, KT > PS > PT	In both word-initial and word-final obstruent clusters, the first segment is preferably a sibilant. The preferred clusters are SK word-initially and ST word-finally. Again, labials do not like to occur in final position, whereas coronals and dorsals do.
$N^1N^2_{cor}$ > $N^1_{cor}N^2_{non\text{-}cor}$	$N^1N^2_{cor}$ > $N^1_{cor}N^2_{non\text{-}cor}$	Both word-initially and word-finally, the second member of a nasal cluster is preferably coronal.
LL = /rl/	LL = /rl/	/lr/ is unattested in either position.

3.5 A salience-based account of consonant clusters[20]

Markedness, sonority and NAD are all able to account for some aspects of consonant phonotactics but none of them alone offers an entirely satisfactory explanation. I argue that if we consider absolute salience, i.e, the inherent context-free Cue Robustness of a consonant, we can expect that the more salient a segment, the more likely it will persist in preconsonantal position (word-initially) and post-consonantal position (word-finally) without the support of an adjacent vowel. Expressed differently,

---

[20] Most of the contents of this section, as well as the experiment results, are presented in Baroni (in press).

(49)

x and y being two consonants of equal sonority not immediately adjacent to a vowel, x is more salient than y if the manner and place of articulation of x are more easily discriminated than those of y in the same position.

I propose the following salience scales, one for obstruents, one for nasals and one for liquids:

(50)
a) $S > H > K > P > T_{cor}$, where H = Φ, Χ, θ
b) m > n > ŋ
c) r > l

(50a) states that, among obstruents, fricatives have more robust cues than stops, since discrimination of stops is based mainly on the formant transitions of the following vowels, whereas fricatives possess their own internal cues. Among fricatives, stridents are louder than non-stridents and among stops, dorsals appear to more audible than labials, which in turn are more easily discriminated than coronals. This salience scale based on the place of articulation, however, seems to apply only to oral stops, since in nasal stops, as shown in (50b), the labial nasal is more salient than the coronal and dorsal, and here dorsal is the one with the least robust cues. Finally, (50c) proposes that rhotics have greater acoustic cues than laterals. These scales are based on experimental data coming from various sources, which will be discussed in the following sections.

3.5.1 Obstruents

A study conducted by Harris (1958) shows that, in English, [s, ʃ] are more easily discriminated than [f, θ]. Moreover, results from another experiment (Cutler et al. 2004) indicate that both American English and Dutch speakers not only tend to identify correctly [s, ʃ] more frequently than [f, θ], but also that [θ] is easily confused with [f] and vice versa. /s/ is also traditionally considered the unmarked fricative, since if a language has only one fricative, it is normally some type of /s/ (Maddieson 1984). As for dorsal fricatives, their inherent salience has received little attention so far. Traditionally, velar /x/ is classified as non-strident, and should therefore pattern with /f, θ/ as low-intensity fricatives, whereas uvular /χ/ is assigned a [+strident] feature (Zsiga 2013:266), like sibilants. However, to my knowledge, no substantial evidence is available for this distinction, and the distribution of uvular fricatives does not seem to coincide with that of coronal stridents. Diachronic neutralization processes could help explain the position of dorsal fricatives in the scale. For example, many Spanish dialects have neutralized the opposition between /s/ and /θ/ dispensing with the latter, probably because of the greater acoustic salience of the former. In Cockney English, [θ], which is probably the least salient English sound (Fletcher 1972:82-86), is fronted to [f]. In Old English, the word-final fricative spelled as <gh> in words like *laugh, enough* used to correspond to /x/, but today this phoneme has been replaced by /f/, although one would expect a dorsal segment to be preferred word-finally as opposed to a labial one. There could be several reasons for this substitution, e.g., the phoneme /x/ might have been lost independently or was simply replaced by a less marked segment.

However, another possibility is that [f] is more salient than [x]. Nevertheless, in Dutch, stems ending with /f/ turn it into [x] before another consonant, as in *kloof – klucht* 'gap – farce', or compare Dutch *kracht, lucht, zacht* with English *craft, loft, soft* and German *Kraft, Lunft, sanft* 'power, air, soft' (Huber 2009:153). Further evidence that dorsal fricatives might pattern with stridents comes from a study on German fricatives (Kemp 2011:18), that showed that [ʃ] and [x] are perceptually closer than [s] and [ʃ], [s] and [ç] and [ʃ] and [ç], respectively. A cross-linguistic study conducted by Gordon et al. (2002) compared the fricatives of several languages in terms of length, center of gravity, spectra and formant transitions. It became apparent that center of gravity (COG, henceforth) is a good predictor for intensity and noise at higher frequencies, and therefore, acoustic salience. It was also revealed that front fricatives (except [f]) generally have higher centers of gravity than back fricatives. The languages considered are Chickasaw, Western Apache, Gaelic, Western Aleut, Montana Salish, Hupa and Toda.

(51) COG values for fricatives
Chickasaw: s > ɬ > ʃ > f
Western Apache: s > ʃ > ɬ > x
Gaelic: s > f^j > ç > f > ʃ > x
Western Aleut: s > ç > ɬ > x > χ
Montana Salish: s > ɬ > ʃ > χ > χ^w > x^w
Hupa: s > ɬ > ʃ > x > x̹^{w} > x^w
Toda: s̺> ʃ > ʂ > s > f > x > ɬ̜> ɬ > θ.

In general, sibilants always display a higher COG than other fricatives, and [s] always shows the greatest intensity. In Gaelic and Toda, which possess both labial and dorsal fricatives, [f] has a higher COG than [x], and in Montana Salish velar and uvular fricatives are poorly differentiated from each other. Given the scarcity of studies on non-sibilant fricatives, therefore, for the time being it seems wise to group them together as a class, since none of them appears to be significantly more salient than the others.

Another experiment (Hume et al. 1999) on Korean and American speakers demonstrated that, among the three stops [k, p, t], the dorsal and labial ones are easier to identify correctly than the coronal one, with [k] being slightly more salient than [p], but the difference in salience between [k] and [p] appears to be much smaller than the difference between [k, p] and [t]. Jun (1995) justifies the assimilation of Korean stops on phonetic grounds, arguing that the phonological assimilation of [t] to [k, p] and of [p] to [k] is due to the fact that [k] and [p] have stronger acoustic cues than [t]. Unfortunately, phonologists and phoneticians do not agree on the stop salience scale. Hume (1998) argues for the "perceptual vulnerability" of labials. Studies of Miller & Nicely (1955), Malécot (1958) and Wang & Bilger (1973) are inconclusive as well. In Miller & Nicely's experiment, participants found coronal to be the most salient place of articulation, whereas according to Wang & Bilger's findings, coronals and labials were both more salient than dorsals in the onset and coronals were more salient than labials in the coda, with both being more salient than dorsals. Again Malécot's results were different, with labials being more salient than coronals and dorsals, which in turn did not differ significantly from each other. Winters (2000) points out that none of these approaches have been successful so far because of the difficulties connected with top-down effects, such as speakers' expectations and linguistic knowledge. To avoid such effects, Winters devised an experiment where the stimuli were both visual and acoustic. Participants had to identify a series of nonce words, basing their decision sometimes solely on an acoustic input and sometimes on an acoustic input accompanied by a visual stimulus, i.e., a video of a person pronouncing the word. The results indicate that labials are more salient than dorsals, which are in turn more salient than coronals. The greater prominence of labials, interestingly, is not only caused by their visual cues, since the correctness rate for labials was still the highest even in the audio-only stimuli. However, without visual stimuli, the difference between labials and dorsals was not that significant. Moreover, adding visual stimuli dramatically increased the salience of labials, whereas the addition of acoustic stimuli significantly increased the prominence of coronals and dorsals but not that of labials.

3.5.2 Nasals

Nasal stops behave differently from their oral counterpart. Nasals display an asymmetry due to the place of articulation in a more prominent way than oral stops, since there are languages where /m/ patterns with obstruents and /n/ with sonorants. For example, Dutch has two agentive suffix, *-er* and *–aar*, the former is selected by stems ending with an obstruent and the latter by stems ending with a sonorant, e.g., *tennissen* 'to play tennis', *tennisser* 'tennis player', *tekenen* 'to draw', *tekenaar* 'drawer', but *ademen* 'to breathe', *ademer* 'breather' (**ademaar*) (Botma 2004:310). This difference could be due to the fact that labial is the most consonant-like place, whereas dorsal would be the most vowel-like and coronal something in between. According to Halle et al.'s Feature Geometry (2000), all vowels are primarily dorsal. Labiality in vowels is normally a secondary feature (roundness) and limited to back vowels in most languages, where vocalic roundness is seldom distinctive (non-back rounded vowels, such as /y/, are cross-linguistically marked). Labial consonants, on the contrary, are among the first to be learned by children. In particular, /m/, compared to other nasals, has the most salient anti-formant, /ŋ/ the least. /m/ has more *nasality* than /n/, and /n/ has more *nasality* than /ŋ/ (Greenlee & Ohala 1980:290). /ŋ/ is very likely to be misperceived for a nasalized vowel and vice versa, see for example the integration of French loanwords ending with a nasalized vowel in Swedish:

(52)

French *restaurant* [ɑ̃] 'restaurant' → Swedish *restaurang* [ɑŋ]

French *béton* [ɔ̃] 'concrete' → Swedish *betong* [oŋ]

French *balcon* [ɔ̃] 'balcony' → Swedish *balkong* [oŋ]

The distribution of /ŋ/ is limited in many languages, e.g., in English, German and Dutch it is banned from the onset position. No such limitation is known for /m, n/. A perceptual study conducted by Narayan (2006) on Filipino adults and infants shows that both groups readily distinguish between word-initial [m] and [n], but the difference between [n] and [ŋ] is less salient, although in Tagalog /m, n, ŋ/ are all distinctive and can appear word-initially (see also Narayan et al. 2010). Therefore word-final [ŋ] is likely to be mistaken for a vowel and word-initially for [n]. This could explain why /m/ and /n/ are quasi-universal whereas /ŋ/ occurs in only 50% of world's languages (Maddieson 1984). If [m] is acoustically the most salient and the most consonant-like of the three, it comes as no surprise that in plateau clusters it tends to occupy the first position. Botma (2004:316) points out that in Polish /m/ is the only nasal that can appear as the first member of a cluster word-initially (as in *mgiełka* 'mist-dim.', *mnożyć* 'to multiply', *mleko* 'milk', *młody* 'young', *mrugać* 'to wink'). In Khasi, a Mon-Khmer language where almost everything goes word-initially (even obstruent clusters differing in voicing and liquid clusters such as /rl/), the only possible nasal cluster is /mn/, as in *mnung* 'sitting motionless by themselves' (Henderson 1976:530). Tsou, spoken in Taiwan, is another case of a language with a very permissive phonotactics. Here all nasals can be the first member of a cluster, but /m/ can combine with a greater number of consonants: /mp, mf, mts, ms, mz, mn, mʔ, mh/ are all allowed, whereas /n/ can only precede /m, t, s/ and /ŋ/ only /v, h/ (Wright & Ladefoged 1994).

3.5.3 Liquids

Assessing the absolute salience of liquids is quite difficult, especially because, if the realization of /l/ is quite consistent cross-linguistically, /r/ can be realized as a coronal trill, a velar trill, a tap, a flap, a velar fricative, a uvular fricative, etc. However, it is arguable that if in a language rhotics behave like sonorants, they must have been phonetically sonorant at some point in their history, i.e., non-fricative continuant sounds. Since salience is based on phonetics, if each language realizes /r/ in a phonetically different way, it is likely that the liquid salience scale is language-specific. The definition of rhotics has always been problematic in phonology, since there do not seem to be features that are common to all the members of the class and sounds that function as rhotics in some languages, do not in others. As Lindau (1985) points out, "the relations between members of the class of rhotics are more of a family resemblance". Phonologically, though, there are good reasons to group rhotics together as a class. First of all, they are very often the only allowed second member of an onset cluster or the first member of a coda cluster, independently from their actual place and manner of articulation. They also show an affinity with vowels, especially low ones: they tend to lengthen the preceding vowels, color them (rhoticization) and can act as a syllabic nucleus in a number of languages, e.g., Czech, Serbian, etc. Moreover, different rhotics alternate with each other: in Fula, /r/ is realized as the approximant [ɹ] before a consonant and as the trill [r] elsewhere (Ladefoged & Maddieson 1996:216). Acoustically speaking, it is quite likely that apical and uvular trills are more salient than taps and flaps, if only because of the greater length of the former. Interestingly, the lateral flap is a member of both the class of rhotics and the class of laterals (Ladefoged & Maddieson 1996:243). Rhotics are traditionally grouped with laterals under the broader definition of liquids, which basically indicates a class of segments that behave, phonotactically, as more sonorous than nasals and less sonorous than vowels. In some languages the lateral and the rhotic are allophones of the same phoneme, e.g., in Korean [r] is the medial allophone and [l] the final one (and in some varieties [n] occurs word-initially). Unlike rhotics, though, laterals normally imply an occlusion and are therefore more restricted from occurring in clusters and in the coda. Although they display less variation in production, there can still be significant inter- and intra-personal differences (Dart 1991). Laterals are particularly prone to coarticulation, e.g., they tend to palatalize much more often than rhotics, and word-finally they are likely to lose occlusion and to be realized as approximants or vowel-like segments. Given the great variability in their production, an analysis based exclusively on acoustic data is unlikely to be conclusive. Instead, I will look at the behavior of /r/ and /l/ cross-linguistically, both diachronically and synchronically, and try to formulate some predictions concerning which one is the segment more likely to be preserved in adverse conditions. First of all, in the passage from Latin to Romance languages, OR clusters underwent important changes: while /r/ was normally preserved in clusters, /l/ changed to /j/ in Italian and to /r/ in Portuguese:

(53)

Latin	*Italian*	*Portuguese*	*Gloss*
brakja	brattʃa	brasa	'arms'
flore(m)	fjore	flor	'flower'
plagia(m)	spjaddʒa[21]	praja	‘beach’

However, in the Portuguese-based creole of São Tomé, O/r/ transformed into O/l/ (Ferraz 1987), as in *blaza, floli* for *brasa, flor.* Assimilation processes should suggest a tendency for less salient segments to assimilate to more salient ones, but if the process is articulation-based, rather than perceptual, the outcome can be different. As a matter of fact, it is /r/ that

---

[21] Initial /s/ in *spiaggia* is said to have an intensifying value (Nocentini & Parenti 2010).

assimilates to /l/ in Toba Batak (Hayes 1986), but the opposite holds true for Hungarian and Italian. In Urban Jordanian Arabic, coronal stops assimilate to following labial stops, but not to dorsal ones, non-sibilant fricatives assimilate to sibilants, e.g., /ħadiiθ ʃaajiʕ/ → [ħadiiʃ ʃaajiʕ] 'a common talk', and laterals assimilate to rhotics, e.g., /lel raajig/ → [ler raajig] 'a calm night' (Zuraiq & Zhang 2006). In Veneto Italian, /l/ is weakened to [e̯] and sometimes deleted, but that does not apply to /r/. However, in non-rhotic dialects of English, /r/ in the coda disappeared, whereas /l/ remained, although velarized. Lenition and assimilation phenomena do not seem to bring sufficient evidence for the greater salience of /r/ as opposed to /l/. However, the liquid salience scale /r/ > /l/ will remain as a hypothesis to test, given the existence of word-initial /#rl/ and the absence of /#lr/.

3.6 Testing salience

Following the definition of salience given in (49), it is possible to make the following generalizations: the more salient a segment is, the more likely it is to appear as the first member of a word-initial plateau cluster. This would explain why /s/, /m/ and /r/, respectively, are the most frequent initial segment in obstruent, nasal and liquid clusters. If salience is a property that allows segments to be identified without the support of an adjacent vowel, word-finally the situation should be the opposite: [s], [m] and [r] should be easier to discriminate than other segments of the same natural class as the second member of a word-final cluster.

(54)

O = obstruent, N = nasal, L = liquid

Word-initially: [s]O > [f]O > [k]O > [p]O > [t]O; [m]N > [n]N; [r]L > [l]L

Word-finally: O[s] > O[f] > O[k] > O[p] > O[t]; N[m] > N[n]; L[r] > L[l]

With regards to [ʃ], I assume that its position in the scale is between [s] and [f], based on the average COG values found by Gordon et al. (2000). I therefore assign the following arbitrary numeric values to each segments of the scale: [s] = 6, [ʃ] = 5, [f] = 4, [k] = 3, [p] = 2, [t] = 1; [m] = 1, [n] = 0; [r] = 1, [l] = 0. Note that these values are merely conventional and do not imply that the difference in salience between [s] and [ʃ] is the same as the one between [ʃ] and [f]. They simply indicate hierarchical relationships, e.g., that [r] is more salient than [l]. To verify the salience hypothesis, a short perception experiment was devised. The participants were enrolled among Dutch and Italian students, from Utrecht and Padua universities, respectively, making sure that they were monolingual and with no significant linguistic experience with languages allowing extremely complex clusters (e.g., Hebrew, Georgian, Khasi, Slavic languages, etc.). Some of the participants received a small amount of money while others were volunteers. All participants declared not to have any visual or auditory impairment. The total number of participants was 64, of which 34 were Dutch and 30 Italian, and included both genders. The salience scale predicts that the higher the segment in the scale, the easier its correct identification.

3.6.1 Methods and materials

The stimuli consisted in 64 nonce disyllabic words, each of them containing a consonant cluster, either word-initially or word-finally. Two lists of words were prepared, so the total number of nonce words was actually 128, but each participant listened to either one of the two lists (see Appendix A). Both lists contained all the target plateau clusters. Stress always fell on the syllable containing the complex cluster, e.g., [ˈstapul] vs. [taˈmusk], in order to avoid sources of distraction from the consonant sequence. The stimuli were read by a phonetically trained Serbian[23] native speaker, who was recorded using Praat (Boersma & Weenink 2013) at a sample rate of 44100 Hz. Each word was analyzed with Praat to make sure that each segment was pronounced correctly (and in the case of stops, fully released) and subsequently white noise was added. Nonce words not containing any cluster were used as fillers. The clusters created were only plateau clusters, i.e., obstruent clusters, liquid clusters and nasal clusters, and were obtained through combining the following sounds: [ʃ, s, f, k, p, t, r, l, m, n]. [ʃs] and [sʃ] combinations were avoided for obvious articulatory and perceptual difficulties. The combinations were the following: [ʃf, ʃk, ʃp, ʃt, sf, sk, sp, st, fʃ, fs, fk, fp, ft, kʃ, ks, kf, kp, kt, pʃ, ps, pf, pk, pt, tʃ, ts, tf, tk, tp, rl, lr, mn, nm] both word-initially and word-finally. [ʃ] was included in the experiment because, albeit not being a distinctive phoneme in Dutch, Dutch speakers are familiar with it and were expected not to have problems in annotating it consistently as <sh> or <sj>. On the contrary, neither [x] nor [ŋ] were included, since Italian speakers are not acquainted

---

[23] The choice of the language of the speaker was based on two factors: (a) the speaker had to be neither Italian nor Dutch, and, (b) he had to be able to pronounce complex consonant clusters with ease, and Serbian, like other Slavic languages, has a very permissive phonotactics.

with them and it would have been hard to devise a straightforward way to note them graphically. As a matter of fact, when loanwords containing these sounds are adopted in Italian, [x] is normally realized as a stop, e.g., *Bach* [bak], and [ŋ] as [ŋɡ], e.g., *jogging* [dʒoɡːiŋɡ]. As for /r/, while its typical realization in Italian is [r], in Dutch, depending on the dialect and the speaker, it is pronounced as an apical trill [r], an alveolar tap [ɾ], a voiced uvular fricative [ʁ], a uvular trill [ʀ] or an alveolar approximant [ɹ]. However, since [r] is the most prototypical rhotic, this pronunciation variant has been used for the experiment.

Participants had to sit in a soundproofed room in front of a PC screen wearing headphones while instructions on the screen told them to listen to a series of a words and type on the keyboard what they thought they had heard. They could listen to each word only once. The task was self-paced and normally did not take longer than 15 – 20 minutes. The sound level was set at 94 dB for all participants. Regarding the transcription, participants were instructed to follow the orthographic conventions of their native language. In the particular case of [ʃ], Dutch speakers were told that they could use either <sj> or <sh> and Italians that they could use either <sc> or <sh>, as they preferred. Analyzing the participants' transcriptions, the absence of the target segment and/or the substitution of the target segment with another with different place or manner of articulation counted as incorrect, whereas inversion, addition of a segment, deletion or misperception of the preceding or following consonant were not considered relevant mistakes. For instance, if the target segment was [k] in the initial cluster [kp], the transcriptions <k>, <kp> and <pk> counted all as correct, whereas <p> and <tp> were judged as incorrect.

3.6.2 Results

In general, the predictions made by the salience scale for obstruents were not met. Whilst [s] was identified correctly most of the time, [f] and [p] were very often misheard or not heard at all, whereas [k] and [t] ranked much better than expected. With regards to liquids and nasals, [m] was identified correctly more often than [n] and [r] more regularly than [l], both word-initially and word-finally, confirming the sonorant salience scales. Other than salience, then, the following variables were considered: position (word-initial vs. word-final), context (following or preceding consonant), legality, language (Dutch vs. Italian) and NAD. The correlation between the correctness rate (CR, henceforth) and each of the variables was checked by running a bivariate Pearson correlation.

3.6.3 Results for Obstruents

For initial obstruent clusters, the correlation between the salience of the first consonant and the CR was not statistically relevant. On the contrary, correlation between CR and language, context, legality and NAD (using the values proposed in Baroni 2012a) turned out to be significant (see Table 4). Word-final clusters present quite a different picture. In final position, salience turns out to be relevant, as well as context and legality, whereas the correlation of the CR with language and NAD is not significant. However, the significance of salience here only indicates that different salience values predict a different CR, not that the higher the value, the higher the CR (see Table 5). Here and henceforth, significant values are indicated by a star if the *p* value is smaller than .05 and by two stars if the *p* value is smaller than .001.

Table 4: Pearson Correlation for word-initial obstruent clusters

Word-initial clusters		Salience	Language (Dutch vs. Italian)	Context	Legality	NAD
CR	Pearson Correlation	.027	.080**	.056*	.144**	.175**
	Sig. (2-code)	.259	.001	.017	.000	.000

Table 5: Pearson Correlation for word-final obstruent clusters

Word-final clusters		Salience	Language (Dutch vs. Italian)	Context	Legality	NAD
CR	Pearson Correlation	.085**	.042	.136**	.217**	.021
	Sig. (2-code)	.000	.075	.000	.000	.381

If we consider all the clusters without distinguishing between initial and final, CR correlates significantly with language, position, context and NAD, but not with salience and legality.

Table 6: Pearson Correlation for obstruent clusters

Obstruent clusters (both initial and final)		Salience	Language (Dutch vs. Italian)	Position	Context	Legality	NAD
CR	Pearson Correlation	.024	.062**	.099**	.092**	.015	.086**
	Sig. (2-code)	.155	.000	.000	.000	.362	.000

3.6.3.1 Correlation between CR and salience

The most frequent mistake in the identification of obstruents was the misperception of [ʃ], which was systematically annotated as <s> (= [s]) by most Dutch participants and sometimes by Italian participants as well. Here a language specific bias was obviously active, since [ʃ] is not distinctive in Dutch. Because of this, it appears from the results that [ʃ] is very hard to perceive, although in fact, [ʃ] was always heard, either as [s] or [ʃ], except once, word-finally, in which input [tʃ] was annotated <t>. Keeping this in mind, table 7 shows the correctness rates for each obstruent:

Table 7: CR for obstruents word-initially and word-finally (% of speakers who correctly identified the target segment in a plateau cluster)

	s	ʃ	f	k	p	t
Word-initial	98.4%	60.5%	78%	96.5%	69%	83%
Word-final	95.3%	82.4%	69.4%	97.25%	82%	99.7%

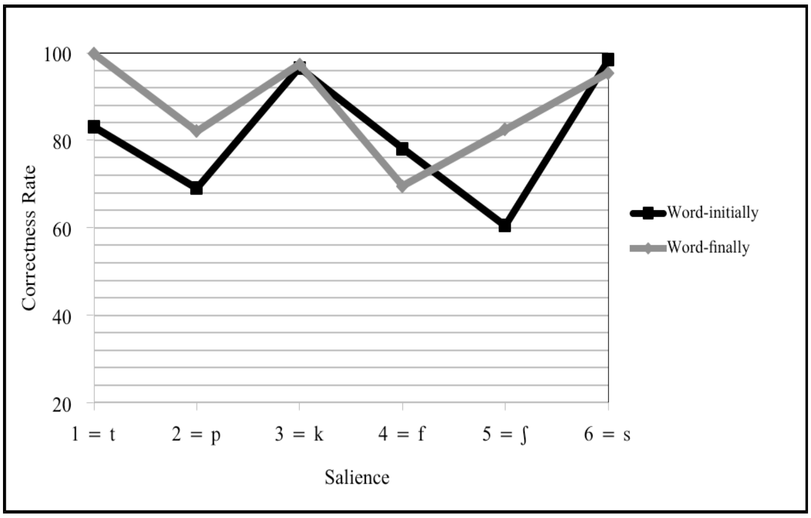


Fig. 1: Correlation between CR and salience for obstruents

From these results (table 7, figure 1), it is apparent that the obstruent salience scale proposed in (50a) does not make the correct predictions. Even if we do not consider the difference between the two sibilants, [k] is identified correctly more often than [f], both word-initially and word-finally, and in the latter position even more often than [s]. [t] is almost always recognized correctly word-finally, and word-initially ranks better than [f] and [p]. Not surprisingly, there is no significant correlation between salience and CR ($p > .05$). What the results seem to suggest is instead a certain weakness of labial segments in clusters, since [f] and [p] have the lowest CR both word-initially and word-finally. [s] and [ʃ] appear to be easier to identify word-initially, unlike stops, which are more accurately perceived word-finally, especially [p] and [t].

3.6.3.2 Other variables

Position. As a matter of fact, word-final clusters were transcribed correctly by 87.9% of participants, whereas initial clusters were perceived without errors only by the 83%. Moreover, the standard deviation of the CR for initial clusters was 20.66, against 14.22 for final clusters. This difference suggests that there was a greater deal of variation in the perception (correct vs. incorrect) of initial clusters than in that of final ones. I have three possible explanations for these results. Firstly, it might have been more difficult to perceive the beginning rather than the end of the nonce words used as stimuli because of the white noise. Secondly, a recency effect might have played a role, making the last sound of the sequence easier to remember. A third explanation might take into account top-down effects: since word-initial clusters are more common, participants might have had expectations about them and could have misperceived them in order to adapt them to their own language's phonotactics. On the contrary, since word-final clusters are generally marked, expectations on them were fewer.

Legality. The correlation between CR and legality is significant (*p* < .001). Initial clusters that are legal in at least one of the two languages are very rarely misperceived. Here I give the clusters and their CR:

(55)

sf, ks (100%) > st, sp, ps (98.4%) > tʃ, sk (95.3%) > pt (92.2%) > ts (67.2%)

Comparatively, initial clusters that are illegal in both languages are mistaken by a much greater number of participants.

(56)

tf, kf, fk (100%) > kʃ (98.4%) > ft (97%) > kp (95.3%) > tp, kt (87.5%) > fʃ (79.7%) > ʃf (73.4%) > pʃ (72%) > tk (67.2%) > ʃk, fs (65.6%) > ʃp (62.5%) > pf (59.4%) > fp (45.3%) > ʃt (43.75%) > pk (26.5%)

Notably, the clusters that turn out to be harder to identify correctly imply a lack of contrast, or put differently, a violation of OCP, e.g., OCP[continuant], as in [fs], OCP[labial], as in [pf, fp].

Word-final clusters that are legal in Dutch seem to be identified more easily by both Dutch and Italian participants.

(57)

kt, ft, st, sp, ps (100%) > pt, sk, ks (98.4%) > ts (87.5%)

(58) Final clusters illegal in both languages

ʃt, fk (100%) > tk (98.4%) > pk (95.3%) > ʃk, fs (93.75%) > kp, ʃp, fʃ (92.2%) > pf (90.6%) > pʃ (82.8%) > sf, tʃ (76.6%) > kʃ (75%) > ʃf (73.4%) > fp (70.3%) > kf (65.6%) > tp (56.25%) > tf (51.5%)

Among illegal clusters, those ending with [t] or [k] are identified correctly more often than those which end with a labial, [p] or [f]. Note that the cluster that seems most affected by its position is [tf], which is identified correctly 100% of the time word-initially and only 51.5% word-finally.

Language. Dutch speakers were expected to perform better than Italians in identifying final clusters correctly, since Italian does not allow more than one consonant word-finally. The correlation between the legality of final clusters in participants' native language and correctness is significant ($p < .001$), with Italian participants actually performing worse than Dutch. Italians misperceived 30.2% of final clusters, vs. 21.7% of Dutch native speakers. Unexpectedly, the correlation between CR and language is significant for initial clusters as well ($p < .001$), with Italians misperceiving 27.6% of initial clusters vs. 25.4% of Dutch.

Context. The correctness in the perception of the first segment in initial clusters and of the last segment in final clusters seemed to depend on, among other factors, the adjacent consonant, i.e., the context in which it occurred. In particular, labial [f] and [p] were very likely to be misperceived or even not heard at all when preceding or following another labial. The correlation between CR and context is significant word-initially ($p < .05$) and even more significant word-finally ($p < .001$) (see Baroni, in press, for details).

NAD. For each cluster, the Net Auditory Distance has been calculated, considering the difference in MOA (Manner of Articulation) and POA (Place of Articulation). The calculation has been made following Baroni (2012a), thus assigning the following MOA and POA values:

(59)
/s/: MOA 4, POA 0.
/ʃ/: MOA 4, POA 1.
/f/: MOA 4, POA 3.
/k/: MOA 4, POA 2.
/p/: MOA 4, POA 3.
/t/: MOA 4, POA 0.

After the calculation was made, the following NAD values were obtained for the clusters:

(60)
NAD /sf, fs, ps, sp, tf, ft, pt, tp/ = 3.
NAD /sk, ks, ʃf, fʃ, pʃ, ʃp, tk, kt/ = 2.
NAD /ʃk, kʃ, ʃt, tʃ, fk, kf, pk, kp/ = 1.
NAD /ts, st, fp, pf/ = 0.

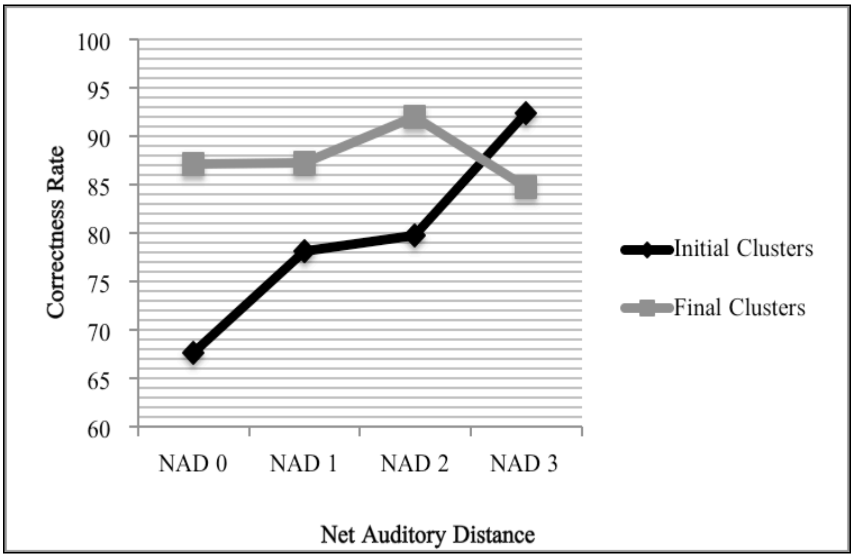


Fig.2: Correlation between CR and NAD for initial vs. final clusters

The correlation between NAD and the average CR of the clusters with the same NAD value appears to be significant for initial clusters (*p* < .001) but not for final clusters (*p* > .05). As fig. 2 shows, the higher the NAD, the higher the CR for initial clusters. The same does not hold true for final clusters, since the clusters with the highest NAD (=3) were identified correctly less often than clusters with a smaller NAD (=2). This might suggest that word-finally, the consonants of a cluster must differ enough to maintain a certain degree of modulation but an excess of contrast might be hard to maintain and perceive word-finally.

3.6.4 Results for nasals

CR correlates significantly with salience (with [m] > [n]) and position, but not with language (see table 8). [m], being more salient than [n], is identified correctly by a greater number of participants. Word-initially, [m] is annotated <m> by 100% of participants and word-finally by 72%, whereas [n] is transcribed as <n> by 94% word-initially and 53% word-finally. The difference in salience between the two appears to be more evident in final position, where very often both [mn] and [nm] are heard as [m]. As a matter of fact, position (word-initial vs. word-final) is statistically more significant ($p < .001$) than salience ($p < .05$).

Table 8: Pearson Correlation for nasal clusters

		Salience	Position (initial vs. final)	Language (Dutch vs. Italian)
CR	Pearson Correlation	.155*	.427**	.087
	Sig. (2-code)	.013	.000	.167

It is important to point out that the comparison between [mn] and [nm] might not have been fair, since the former is attested in both languages and the latter is not. However, [mn] is a very peripheral cluster in Dutch and Italian, occurring exclusively in words containing the Greek stem *mnemo-* 'memory-related'.

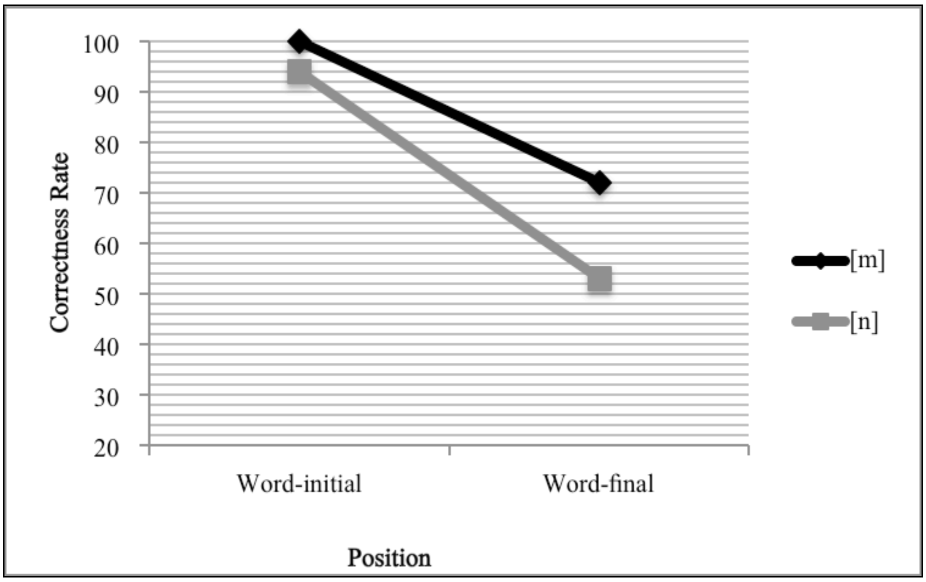


Fig.3: Correlation between CR and position for nasals

There seems to be no significant difference between Italian and Dutch participants in nasal cluster identification. The percentages are very similar: [m] was identified correctly word-initially by 100% of Dutch and Italians and word-finally by 73% of Dutch and 70% of Italians. [n] was transcribed correctly by 97% of Dutch and 90% of Italians word-initially, and by 55.9% of Dutch and 50% of Italians word-finally.

3.6.5. Results for liquids

Table 9: Pearson Correlation for liquid clusters

		Salience	Position (initial vs. final)	Language (Dutch vs. Italian)
CR	Pearson Correlation	.415**	.198**	.145*
	Sig. (2-code)	.000	.001	.020

There were only four nonce words in the experiment containing liquid clusters, one beginning and one ending with [rl] and one beginning and one ending with [lr]. 98.4% of participants identified [r] correctly in word-initial [rl], but only 34.4% heard [l] in word-initial [lr]. Most of the time, [lr] was annotated <dr>. Word-finally, participants performed slightly worse with [r] and much better with [l]: [r] was identified correctly by 87.5% and [l] by 79.7%. Neither of the two clusters is legal in either of the two languages. Germanic proper names that end with [rl] in German and English, such as *Karl, Earl,* contain a schwa in Dutch, e.g., *Karel.* The correlation between salience and correctness is significant ($p < .001$). Also the correlation with position (word-initial vs. word-final) appears to be significant (*p* = .001), with [l] easier to perceive word-finally than word-initially (see fig. 4).

Dutch and Italian participants performed quite differently in the identification of liquid clusters, especially with regards to the lateral. [r] was identified correctly by both groups quite easily word-initially (100% of Dutch, 96.7% of Italians), but word-finally Dutch ranked much better (97% vs. 76.7%). Italians performed very poorly with word-initial [l] (only 13.3% heard it correctly, vs. 53% of Dutch), but were better than the Dutch at identifying it word-finally (86.7% of Italians vs. 73.5% of Dutch). Importantly, word-finally, Italians found it easier to hear [l] than [r] (CR: 86.7% for /l/ vs. 76.7% for /r/), suggesting that the salience of liquids might not be universal but might depend on positional and language-specific factors. The correlation between CR and language is significant (*p* < .05).

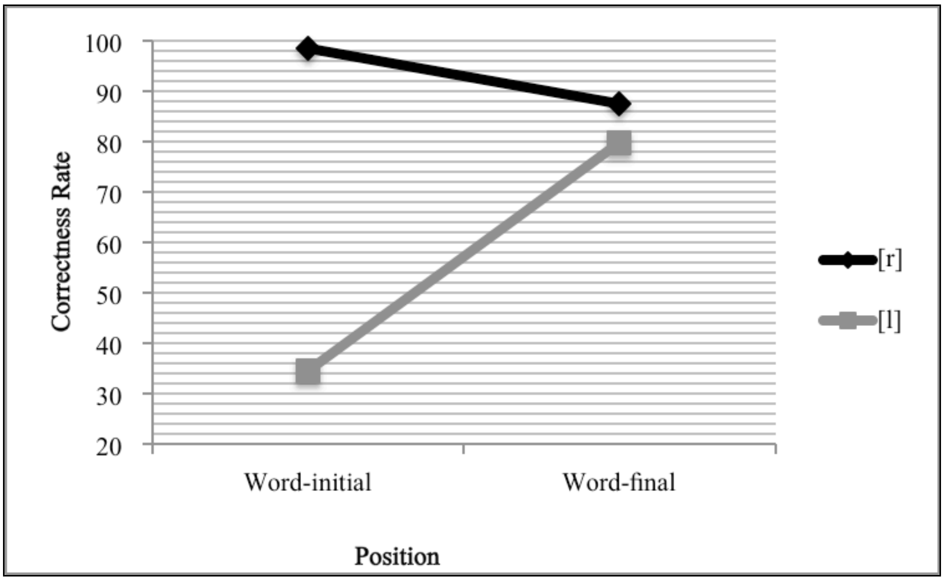


Fig. 4: Correlation between CR and position for liquids

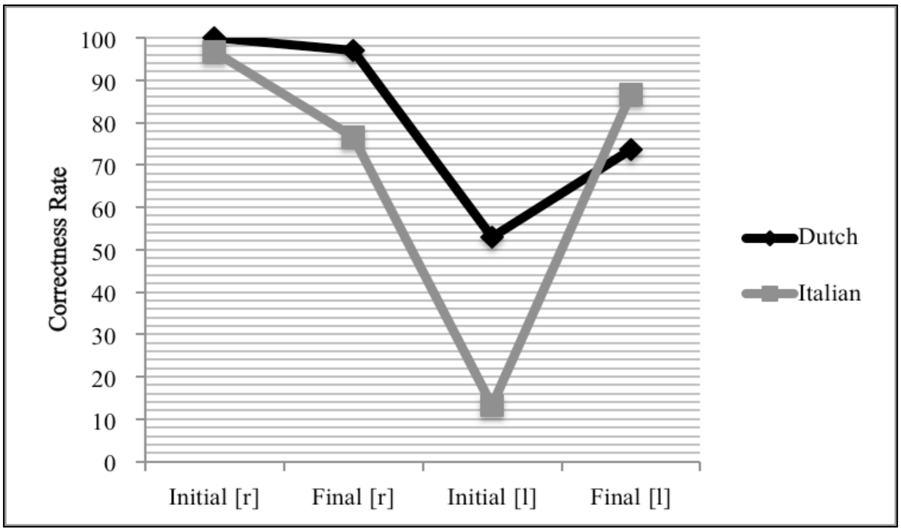


Fig. 5: Correlation between CR and language for liquids

3.7 Discussion

Generally speaking, the salience scale was confirmed for liquid and nasal clusters, but not for obstruent clusters. There are two possible explanations for the results: either the salience scale for obstruents is wrong and must be reformulated, or factors other than salience played a greater role in consonant identification. I argue that both possibilities are worth considering. Leaving aside the difference between [s] and [ʃ], it seems established that sibilants are the most salient obstruents, since, even if [ʃ] was systematically mistaken for [s], it was almost never confused with a non-sibilant fricative or with silence. Given these results, sibilants can maintain their position in the salience hierarchy. When it comes to non-sibilant fricatives and stops, the results suggest that [f] is very likely to be confused with silence or non-linguistic noise. As a matter of fact, [f], together with [p], turned out to be the perceptually weakest segment and probably the most affected by white noise. Since stops imply an abrupt pause in the signal, their manner of articulation was more likely to be noticed, whereas [f] seemed particularly unfit to combine with other obstruents and likely to be confused with silence. This does not appear that surprising if we look at the history of Indo-European languages: Latin initial /f/ turned into /h/ and then to ∅ in Spanish, cf. Latin *ferro,* Spanish *hierro* [jerːo] 'iron'. Proto-Germanic **af* 'off, away' conserved the final fricative in Swedish *av* [ɑːv], English *of* [əv] and Dutch *af* [ɑf] but turned it into a stop in German *ab* [ap], and lost it completely in Danish, *af* [æ]. Therefore, it is not unusual for /f/ to undergo debuccalization and deletion spontaneously. Among stops, [p] turned out to be the most difficult to perceive, whereas [k] was identified almost as easily as sibilants word-initially and more than sibilants word-finally. Unexpectedly, [t], acoustically the least salient obstruent, was heard correctly most of the time. I argue that this should not lead us to believe that [t] is more acoustically salient than [p] or [f], but that by virtue of a top-down effect, participants judged [t] more likely than [f] or [p] to occur in a consonant cluster. The fact that [t] was very often hallucinated is significant: many participants transcribed consonant clusters adding a [t] which was not there before, after or between the two obstruents. The same did not happen with any other obstruents. Moreover, it is well known that, cross-linguistically, sequences of homorganic obstruents are disfavored, but coronal-coronal sequences are generally tolerated more often than dorsal-dorsal or labial-labial sequences. The stimuli contained instances of labial-labial sequences, such as [fp], [pf], that were often normalized by speakers. Of the two instances of coronal-coronal sequences, [st] and [ts], the former is legal in both languages and the latter *almost* legal. [x] or [χ] were not included in the stimuli, so there were no instances of violation of OCP[dorsal]. This might have favored [k], since it was never found adjacent to an obstruent with the same place of articulation. In light of these considerations, the obstruent salience scale might be reformulated in the following way:

(61) sibilants > non-sibilants (dorsal > labial > coronal)
s, ʃ > k > f, p > t.

The results of the experiment are problematic if compared to Male̋cot’s (1958), Wang & Bilger’s (1973), Miller & Nicely’s (1995) and Winters’ (2000) findings, but they are consistent with the “perceptual vulnerability” of labials proposed by Hume (1998) and the salience of [k] is confirmed by Hume et al.’s (1999) experiment as well.

4

ACOUSTIC REDUCTION

After having expounded experimental results on the relative salience of certain sounds with respect to other sounds of the same class, i.e., sibilants with respect to obstruents, /m/ with respect to /n/, /r/ with respect to /l/, in this section I wish to present data from different corpora of spontaneous speech in several languages, with a particular focus on the CLIPS corpus (Savy & Cutugno 2009). The aim of this chapter is to find a correlation between the acoustic salience of certain consonants and their presence/absence in casual speech. As a matter of fact, it is to be expected that segments that are both acoustically salient and relatively easy from an articulatory point of view, e.g., /m/, will be conserved most of the time in running speech. Poorly salient and highly predictable segments, such as /t/ and /n/, are likely to be often omitted, since they are easily reconstructed by the listener. As previously stated, I rarely expect /s/ to be deleted, despite its being relatively articulatorily costly, because I assume that it impacts to a great extent the representation of the words it forms a part of.

4.1 Casual speech: an introduction

Even though linguistic studies tend to focus on an idealized form of language (Saussurian “*langue*”, Chomskyan “competence”), in fact casual speech and all its related phenomena are not exceptions, but the norm (cf. Warner 2011). The truth is that the majority of what we hear during the day is reduced, blurry pronunciations of words in context, and not idealized, carefully uttered words in isolation. It is from this stream of coarticulated sounds that children learn their own language and that we disambiguate and understand speech every day. Nevertheless, casual speech is still underinvestigated. Studies on casual English and German are scarce (Zwicky 1972a, Brown 1977, Dalby 1984, Kaisse 1985, Kohler 1990, Keating 1998, Shockey 2003), and only recently has there been a growing interest in acoustic reduction, especially in psycholinguistic approaches (Ernestus 2000, Hanique & Ernestus 2012, Mitterer 2000, 2008, Mitterer et al. 2008, Brouwer et al. 2012, Pate & Goldwater 2011). To my knowledge, studies on Italian casual speech are even rarer (Farnetani 1995, Landi & Savy 1996, Savy & Cutugno 1997, 1998) and those that approach acoustic variation from a phonological point of view can be counted on the fingers of one hand (Gnerre 1976, Mioni 1976, Farnetani & Shockey 1992, Farnetani & Busà 1993). Compared to some years ago, though, the study of conversational speech is much more accessible, thanks to the availability of freely accessible corpora: CLIPS for Italian (Savy & Cutugno 2009), Ernestus (Ernestus 2000) for Dutch, the Nijmegen Corpus of Spoken French (Torreira et al. 2010), DanPass for Danish (Grønnum & Tøndering 2007), the Kiel Corpus of Spontaneous Speech for German (Kohler et al. 1995/1997), SCRIBE for British English (1989/90), Santa Barbara Corpus of Spoken American English (Du Bois et al. 2000/2005), BEA for Hungarian (2013), and many others. Regardless of the language, casual speech phenomena, being grounded in the same perceptual and articulatory mechanisms, are universal: coarticulation, reduction or deletion of vowels, fricativization of stops, debuccalization of obstruents, palatalization, nasalization, labialization, voicing, etc. However, it is important to point out that languages differ in the way that speakers coarticulate sounds and that children become sensitive to the assimilatory patterns of their own language very early. An experiment conducted by Skoruppa et al. (2013) found that English toddlers are able to compensate for English place assimilatory processes, such as *te*[m] *pounds*, as well as French toddlers are able to compensate for voicing assimilation, as in *montre le bu*[z] *là-bas* 'show the bus down there'. However, French children did not compensate for place assimilation, since it is not a characteristic of French native running speech (*montre la lune* [lym] *par ici* 'show the moon over here' is not attested). Nevertheless, it is possible that this kind of assimilation is phonological and therefore forms somehow part of speakers' grammatical knowledge, but that in conversation all kinds of assimilation can potentially take place. Shockey (2003:3) makes an important distinction between merely phonetic reduction, driven by physical factors, and language-specific reduction, which is controlled by cognitive mechanisms. One of the ideas defended in this thesis is that acoustic variation and reduction are not unrestrained, but that the drive to reduce articulatory effort has to be balanced with the need to preserve lexical distinctivity, in the spirit of both Natural Phonology (Stampe 1979) and OT (markedness vs. faithfulness). Lexical distinctivity is preserved if the invariant remains untouched and I wish to show that casual speech phenomena, in spontaneous conversation, are pervasive, but do not apply blindly, they apply only to units that do not form part of the invariant of a word.

4.2 Previous studies

4.2.1 English

In one of the first studies on casual spoken English, Zwicky (1972a) stresses the phonetic naturalness of casual speech phenomena, as opposed to morphophonological alternations of the type [ɪ] ~ [æ] in *sing – sang*. Moreover, he points out that reduction processes do not apply to the entire lexicon but are somehow sensitive to the lexical/grammatical category that a word belongs to. He cites the example of the generalized elision of the vowels in verb forms such as *is, am, are*, which lose their syllabicity, and notes that words which are phonologically very similar, such as *in* and *and*, do not undergo the same process. Vowels can indeed be deleted in *in* and *and* but these function words always maintain their syllabicity, e.g., *a radio* [n̩] *a television set* vs. *he*[z] *my son*. Zwicky also notes that vowel elision, as in *opera* [ɒprə], *buttoning* [bʌʔn̩ɪŋ], is only possible when the vowel is followed by a sonorant, but does not apply before an obstruent, even though the result would be potentially pronounceable, e.g., *[æʔkə] is not a possible rendition of *Attica* (Zwicky 1972a:609). Dalby (1986) examines in depth the factors that favor the reduction of unstressed vowels in casual American English in two studies. In the first one, several TV broadcasts were analyzed and the following tendencies in unstressed vowel deletion were identified:

- There was great interpersonal variation.
- Unstressed vowels that were the nuclei of a final syllable of a word were deleted less often than those in medial position (the correlation was significant, $p < .001$).
- The place and the manner of both the preceding and the following

consonants were relevant to vowel reduction. Vowels were more likely to be deleted if preceded by a liquid (16%) or a labial (obstruent or nasal, 11%) and less likely if preceded by a coronal obstruent (6%) or a velar (4%). As for the following consonants, vowels were more likely to be deleted if followed by a coronal (obstruent or sonorant, 9%) and less likely if followed by a velar (8%) or a labial (6%). As for manner, vowel deletion occurred more often after continuants (fricatives and sonorants) than after stops, and more often before stops than before continuants.

- The highest rate of deletion (22%) was found between a liquid and a coronal stop, such as in *ability*, followed by the liquid__velar context (16%), liquid__labial and labial__labial (9%).

In his second study, Dalby elicited fast speech artificially from three subjects, and compared their pronunciations in fast and slow tempo. He found that 111 phonotactically illegal clusters occurred in the corpus, of which 48 were in initial position. Some of the clusters violated OCP restrictions, e.g., [tl] in [tlɛgfɹ̩z] for *telegraphers,* [pf] in [pfɛʃn̩l] for *professional,* [ðl] in [ðleɪm] for *the lame.* Other clusters were clearly violations of the Sonority Hierarchy and/or displayed a sequence of more than two consonants: [hsk] in [hskæsɪk] for *her scholastic,* [tspɹ] in [tspɹɪnski] for *Zabrinsky,* [mk] in [mkæŋkl̩] for *mechanical,* [lð] in [lðə] for *will the,* [ŋg] in [ŋgoʃjeɪɾəɹz] for *negotiators,* [lɹ] in [lɹɛnzoz] for *Lorenzo's,* and [db] in [dbeɪ] for *debate*. Some other clusters were simply unattested in careful speech, although they did not violate either OCP or the Sonority Hierarchy, e.g., [ml] in [mlɛɹiə] for *malaria,* [mɹ] in [mɹɒko] for *Marocco,* [kn] in [knju] for *can you,* [ðm] in [ðmɪsiːbi] for *the Mississippi,* [vn] in [vnilː] for *vanilla*. Dalby's findings demonstrate that phonotactically illegal clusters are so only at the phonological level, because phonetically they are possible. It is therefore arguable that articulatory factors play a minor role in shaping phonotactic preferences, and that certain sequences are illegal because they are not easy to perceive. Shockey (2003), in her book about spoken English, proposes a Vulnerability Hierarchy: [t], [ð], [ə] are defined as incredibly vulnerable, [n], [d], [l], [z] moderately vulnerable and [f], [m], [ʃ], [tʃ], [dʒ] practically invulnerable. It is noteworthy that the vulnerable consonants are all coronal whereas the almost invulnerable ones are either labial or stridents. Even though [z] is a strident as well, its high frequency in English (e.g., as the realization of both the plural morpheme and the genitive) increases its chances of being reduced. Moreover, whilst [t, d, ð] tend to disappear, nasals and coronal fricatives are simply prone to assimilation, but they are normally preserved, albeit modified, and tend to conserve their salient characteristics (e.g., nasality, stridency). Shockey further notes that liquids and nasals are likely to become syllabic in casual speech, sometimes even word-initially, as in American English [l̩æskə] for *Alaska* (Shockey 2003:22). Non-coronal nasals can become syllabic as well, even though it happens less frequently, e.g., Southern British [ɛgzm̩] for *eggs and bacon*, [juɣŋ̩] for *you can*. Obstruents can function as a syllabic nucleus too, and unsurprisingly, [s] and [ʃ] do so more often than other obstruents, e.g., American English [mæks̩məm] for *maximum*, Shepherd's Bush English [ʃbweɪsʔ͜t] for *should waste*. Also non-sibilant fricatives, such as [f], are sometimes syllabic, e.g., East London [f̩gɑʔ] for *forgot*. Other common phenomena of spoken English are the fricativization of stops, the tapping of coronal stops and nasals, contextual voicing and devoicing, the tendency to simplify syllabic structure towards the CV type, the loss of final [t, d] and of initial [h], the loss of occlusion in syllable-final pre-consonantal nasal stops (e.g., VNC → ṼC), etc. Sometimes, the reduction of specific words is so extreme that the articulatory motivation seems hard to reconstruct; examples are [jɔ̃] for *you know* and [jɔ̃w̃ʌ̃mĩ] for *you know what I mean*. Shockey calls these extremely reduced forms "icons" (Shockey 2003:46) and further notes that these pronunciations often apply to place names, e.g., [fænʃɔ] for *Featherstonehaugh.*

4.2.2 Dutch

Dutch is one of the most studied languages when it comes to acoustic reduction and casual speech, mostly thanks to the work of Ernestus (2000). This author created a corpus of spoken Dutch and analyzed acoustic reduction and voice assimilation. Notably, even though she stresses the importance of ease of articulation, she points out that acoustic salience can play a role in the preservation of a segment (Ernestus 2000:110). As in many Germanic languages, [t] is probably the most likely segment to be absent from spontaneous speech. In her corpus, Ernestus found that [t] is particularly prone to disappear after [s] and before labial stops (nasal or oral). As a matter of fact, [s] and the noise burst of [t] have very similar acoustic properties and therefore the coronal stop is not very salient after the fricative and speakers can dispense with it without hindering communication. Instead, the acoustic absence of [t] before labial stops is probably due to the fact that the labial gesture masks articulations realized within the vocal tract (Browman & Goldstein 1990:360, Ernestus 2000:114).

The rhotic liquid is another segment that is frequently absent from spoken Dutch, especially syllable-finally after schwa. In that context, though, its realization is more vocalic than consonantal and is therefore easily confused with schwa. [r] is also often missing from the high frequency word *precies* 'precisely'. The presence of the nasal [n] is very often merely signaled by the nasalization of the preceding vowel. It is pointed out that very frequent items were realized most of the time as idiosyncratic forms, for instance [tyk] for *natuurlijk* /nɑtyːrlək/ 'of course'. In following studies (Mitterer & Ernestus 2006, Mitterer et al. 2008) experimental evidence has been brought that listeners automatically compensate for casual speech reductions, in particular, they believe they hear [t]'s that are actually not pronounced. I argue that this finding must be somehow related to the tendency of many Dutch participants to the experiment (see discussion of the results, 3.7) to reconstruct an epenthetic [t] that was not present in the acoustic input, e.g., [kp] noted as <ktp>.

4.2.3 French and Spanish

Besides studying spoken Dutch, Ernestus collaborated in the creation of two spoken language corpora at Nijmegen University, one for French (Torreira et al. 2010) and one for Spanish (Torreira & Ernestus 2010). The two corpora were obtained using similar techniques and that made the comparison between casual speech phenomena in the two languages particularly feasible. From one study comparing the pronunciation of intervocalic voiceless stops in the two languages (Torreira & Ernestus 2011a), it emerged that in Spanish intervocalic stops display incomplete closure more often and undergo voicing most of the time. The same phenomena took place in French, but to a much lesser extent. On the other hand, French vowels undergo devoicing more often than Spanish ones, although they are generally longer. Another study on French (Torreira & Ernestus 2011b) focused on the realization of *c'etait* 'it was' in French, which was reduced in more than half of the tokens of the sequence. Specifically, /setɛ/ was pronounced [stɛ], even though this pronunciation variant is normally ignored in the available descriptions of the language. Similarly, Spanish /s/ is traditionally described as always voiceless, even intervocalically, but an acoustic study based on casual speech revealed that over a third of intervocalic /s/ in the Nijmegen corpus was realized as voiced [z], especially in frequent words and in redundant suffixes (Torreira & Ernestus 2012).

4.2.4 German

German acoustic reduction is well documented by Kohler (1990, 1999). For instance, he notices that the articles *dem* 'the-DAT.MASC.SG' and *einen* 'a-ACC.MASC.SG' can undergo different degrees of reduction (Kohler 1990:74):

(62)
(a) [deːm → dem → dəm → dm → bm → m]
(b) [ʔaɪnən → anən → an/ənən → ən/nən → n]

He further notices that /r/ vocalizes to [ɐ] every time that it is not followed by a vowel and that the sequence /ə/N is often reduced to a syllabic nasal. As already cited, Kohler coins the term "articulatory prosodies" (Kohler 1999) to identify the phonetic essence of words in highly connected speech, which grossly correspond to what I call the invariant. As in other Germanic languages, in spoken German [t] is often missing, coronal stops assimilate to labial and dorsal consonants and some words can be realized without vowels, e.g., *zum* /tsʊm/ → [tsm]. Phenomena that appear to be more specific to German are the following: (1), final /l/ may be deleted, even before a vowel, especially in frequent words such as *mal* 'times', *soll(en)* 'should', *will* 'to want-1sg'; (2), a stop in a nasal environment is often realized as a glottalized nasal, e.g., /kœntən/ → [kœnn̰n]. I argue that the conclusions drawn by Kohler on spoken German can be applied to casual speech in general: "[a]rticulatory components become dissociated from segmental entities of speech sound size, manifesting themselves at highly variable points in time (…). [For example, n]asalization may become a feature of a syllable or a whole syllable chain and not be tied to a delimitable nasal consonant. The same applies to labi(odental)ization" (Kohler 1999:92). Another interesting study conducted by Koesters Gensini (2000) tested the prediction made by Meinhold (1973) and Kohler (1995), according to whom coarticulation and reduction are blocked by morphology (in case of potential ambiguity). The findings contradict the two authors' expectations. As a matter of fact, in the spontaneous speech of five German informants (all from the Bonn – Köln area), morphological syncretism was significantly increased. Given the personal verbal suffixes /-ə, -st, -t, -ən, -t, -ən/, /-ə/ and /-t/ were often reduced to zero, /-st/ to [-s] and /-ən/ was realized as [-n̩] (Koesters Gensini 2000:67). As for nominal morphology, case suffixes were not expressed most of the time. The definite and indefinite articles were realized as, respectively, [də] and [n], instead of standard *der, die, dem, den* (definite) and *eine, einer, einem, einen* (indefinite). A number of nouns lacked the case suffix as well. Given these data, Koesters Gensini concludes that the primacy of morphology over phonetics is a controversial subject and suggests that the decodification of the acoustic signal must necessarily work on a level greater than the segmental one (Koesters Gensini 2000:76).

4.2.5 Danish

Conversational Danish is particularly relevant for the topic of acoustic reduction, since Danish exhibits a great deal of consonantal lenition and schwa-deletion. Unfortunately, not many works on Danish phonology and phonetics are available, let alone on casual Danish. Basbøll (2005:259) underlines the fact that Danish is peculiar in that even in very careful, almost artificial pronunciation, obstruents are pronounced in their lenited form. Notable processes are: (1) the allophonic realization of non-initial /d/ as syllabic [ð̩] and of non-initial /ɡ/ as [j] or [ʊ], depending on the vocalic environment; (2) the fusion of syllable-final /r/ with the preceding vowel; (3) the assimilation of schwa to a following approximant, cf. (28a-b). Constructing psychologically real UFs for Danish words is not always a simple task and most of them are likely to be based on diachrony and/or orthography. As a matter of fact, "careful pronunciations need not be, and are very often not, the most frequent or normal pronunciations of the word forms involved. On the contrary, they will be somewhat artificial and pedantic, particularly perhaps in languages that have heavy stresses and thus normally a great deal of reduction (e.g. English, Danish, Russian). In such cases, it may be that speakers may even construct full-vowel plans which are virtually never realized as such (…). Possibly such abstractness may be due to conventional orthography" (Linell 1979:56, n12). Thanks to the efforts of Nina Grønnum (2006), a corpus of spontaneous Danish, DanPass, is today available online. According to the results of a study on the reduction of stops (Pharao 2009), based on conversations extracted from DanPass, Danish stops show a tendency towards reduction in intervocalic context, but are rarely completely deleted, with /d/ being the most likely to dissolve. As for reduction, velar stops appear more likely to be reduced to fricatives than coronal and labial, but according to Pharao (2009:120) the weakness of velars depend on the rate of confusability with other segments. As a matter of fact, if a labial stop becomes a fricative, it may be confused with [f], and similarly, [t] may be confused with [s], whereas, since there are no velar fricatives in Danish, [k] can spirantize without potential ambiguity.

4.3 Italian

This section constitutes the core of this chapter and will give an overview of Italian connected speech. Italian, being a syllable-timed language, does not exhibit phonological vocalic reduction as opposed to stress-timed languages like German, English, Dutch and Russian[24] (Bertinetto 1989). However, phonetic reduction of vowels is omnipresent, as well as consonant lenition. In order to collect data on spontaneous spoken Italian, I selected four dialogues from the CLIPS corpus (Savy & Cutugno 2009), a project carried out at the University of Naples and guided by Federico

---

[24] As for deletion, Bertinetto & Loporcaro (2005:140) point out that the only vowel that undergoes phonological deletion in Modern Italian is word-final /e/, as in *maggiore* vs. *maggior* "bigger, greater", whereas other instances of vowel deletion are lexicalized or limited to some word classes (e.g., 3rd person plural verbs, such as *vivon* instead of *vivono* 'they live' in pre-consonantal context.)

Albano Leoni. The corpus offers several advantages: it is composed of several sub-corpora, each one representing a different speech modality: radio-TV, dialogues, read speech, telephone, orthophonic. In order to do justice to the very complex linguistic situation of Italy, where every region has its own dialect and standard Italian is spoken with notable prosodic, rhythmic and segmental differences, speakers from many areas were recorded: North (Genoa, Turin, Bergamo, Milan, Venice, Parma), Center (Florence, Perugia, Rome), South (Bari, Catanzaro, Naples), Sicily (Palermo), Sardinia (Cagliari). For the purposes of this thesis, I chose to analyze four dialogues, each one involving a pair of speakers from a different area (Turin, Venice, Rome, Palermo). Each dialogue consisted of a Map Task (Bard et al. 2001), i.e., speaker A and speaker B look at two maps, which are similar but not identical. Each one, in turns, has to explain the other how to get from one point to another. Importantly, the two speakers already knew each other, so they were communicating in a spontaneous, friendly manner. Given the nature of the task and the fact that the maps looked similar, all the four dialogues contain a group of recurring words, e.g., *sinistra* 'left', *parte* 'part, side', *punto* 'point, dot', *farfalla* 'butterfly', *televisore* 'television', etc. This group of frequent lexical items offers particularly interesting insights on Italian casual speech phenomena. As for the acoustic analysis, monosyllabic function words, unintelligible words and words occurring only a few times were ignored, as well as words not containing any of the target consonants. In general, I preferred to focus on content words consisting of at least two syllables, because monosyllables and function words are sometimes reduced to such an extent that it becomes impossible to analyze them acoustically (for a list of all the words analyzed, see Appendix B).

4.3.1 Aim of the study

The analysis was carried out with a double intent: (1) to provide a description of the recurring casual speech phenomena of Italian, trying to distinguish between universal and language-specific tendencies; (2) to verify the existence of a correlation between the results of the experiment on consonant salience (chap. 3) and the degree of reduction undergone by consonants in running speech.

4.3.2 Procedure

After having transcribed the four dialogues, a conspicuous number of target words were selected and all (or most) of the realizations of those words were analyzed using Praat (Boersma & Weenink 2013). The transcription of vowels is not particularly narrow, since they are less relevant for our purposes, whereas particular attention was given to the following consonants: /s, ʃ, f, k, p, t, k, r, l, m, n/. All transcriptions were made by ear by the author, who checked the audio files several times and compared them to the spectrograms. In the case that there was disagreement between the author's perception and the spectrogram, the latter was judged more trustworthy. For instance, if the author was unsure whether a schwa was present or not, the presence of vocalic formants on the spectrogram aided the decision. Section 4.3.3.2 presents the pronunciation variants of liquids, section 4.3.3.3 deals with nasals and section 4.3.3.4 with obstruents. For each group of consonants, percentages are given about their realizations and the position of the word in which they occur. All the data obtained were checked for statistical significance but the results are presented only when at least one of the factors taken into consideration proved to be significant (with a *p* value smaller than .001 or .05). While I am aware of the importance of primary and secondary stress in acoustic reduction, I did not consider stress as a variable, given that in connected speech entire words could be realized as phonetically unstressed. However, further research will surely have to take lexical stress into account.

4.3.3 Casual Italian phenomena

4.3.3.1 Vowels

As noted by Savy & Cutugno (1997), Italian is not exempt from vowel reduction. In an unstressed position, all vowels are likely to be centralized, and importantly, vowel centralization does not seem to be a characteristic exclusive to some speaking style or rate but can be considered as a structural phenomenon. All vowels in unstressed positions are normally realized in the central area of the vocalic space. Examples of vowel reduction extracted from the dialogues are given in (63).

(63)

(a) Reduction to schwa

/a/ → [ə]: /far.ˈfal.la/ → [fɚfalːə], /al.ˈlo.ra/ → [aloɾə], /ˈri.ga/ → [ɾigə].

/e/ → [ə]: /ˌpra.ti.kaˈmen.te/ → [preɾijamẽntə], /ˈsem.pre/ → [sẽmprə], /ˌte.le.vi.ˈzo.re/ → [tələvizore].

/o/ → [ə]: /ˈpun.to/ → [pũtə], /o.ˈbli.kwo/ → [oblikwə], /ˈal.to/ → [altə].

/i/ → [ə]: /si.ˈnis.tra/ → [səsðr]

(b) Centralization

/a/ → [e]: /ˌpra.ti.kaˈmen.te/ → [preɾijamẽntə]

/a/ → [ɨ]: /far.ˈfal.la/ → [fɨfalːa].

Before proceeding, it is important to point out again that a lexically stressed vowel is not necessarily phonetically stressed in connected speech. As a matter of fact, in the corpus, many lexically stressed vowels were reduced because the whole word they belonged to was unstressed. For instance, the word *allora* 'by the way', which was used most of the time as a conversational filler, was frequently unstressed and underwent a great deal of reduction that, quite surprisingly, affected the lexically stressed /o/ to a much greater extent than the two lexically unstressed /a/'s.

(64) Tokens of *allora* with centralization of the stressed vowel

[alːøra, əlørə, alːwəɾa, alːəɾə, alːəɾe, alːœr, alːœra]

As shown in (64), the stressed vowel of *allora* may be realized in running speech as schwa, as central [ø, œ] or as the diphthong [wə], even in tokens where the other two vowels were fully pronounced. Besides centralization and reduction to schwa, deletion was another pervasive phenomenon that could potentially affect all vowels, regardless of timbre and stress, although obviously unstressed vowels were more likely to be deleted.

(65) Vowel deletion

/a/ → [∅]: /far.ˈfal.la/ → [frfalːə, vrfalːa], /al.ˈlo.ra/ → [ərɽ̥ː, rː], /ˌpra.ti.ka.ˈmen.te/ → [prtikamẽnte].

/E[25]/ → [∅]: /ˈsɛm.pre/ → [sʲmprʲ, smpɸ, sɛmpr], /ˈsem.pli.tʃe/ → [sepʲtʃ].

/i/ → [∅]:/ˌpra.ti.ka.ˈmen.te/ → [prtkɛnte, prmdə], /si.ˈnis.tra/ → [s˜sr], /im.ˈpɔr.ta/ → [˜pɔrtə].

/O/ → [∅]:/al.ˈlo.ra/ → [aɽ̥ːa, ərɽ̥ː, əlrə, alːrː, ɛrə, alrːə, alr, ɛlːrɔ], /tʃo.ˈɛ/ → [cɛ, tʃɛ, tʃʷɛ]

/u/ → [∅]: /ˈpun.to/ → [pmto, p˜tə]

4.3.3.2 Liquids

Italian possesses two non-nasal sonorants, /l/ and /r/, traditionally referred to as liquids. Their phonotactic distribution is quite similar, i.e., they can occupy the coda position (together with nasals), they are the preferred second member of an onset cluster and they can occur word-finally (albeit only in a small number of function words, such as the definite article *il* 'the-MASC.SG' and the preposition *per* 'for'). According to the hypothesis of chapter 3, /r/ is more salient than /l/, and the results of the perception experiment confirmed that it was easier for the participants to identify correctly the rhotic rather than the lateral in the clusters [#rl, rl#, #lr, lr#]. However, as mentioned earlier, the great acoustic variability of liquids, and of the rhotic in particular, made it difficult to claim that the salience relationship between /r/ and /l/ is universal. In Italian, /r/ is typically

---

[25] Italian has four distinctive mid vowels, /ɛ, e, ɔ, o/ but since my analysis did not focus on vowels, I did not distinguish between /ɛ/ and /e/, on the one hand, and /ɔ/ and /o/, on the other.

realized as an alveolar apical trill, which is considered the prototypical phonetic manifestation of rhotics crosslinguistically. In the four dialogues that I analyzed, all the eight speakers produced standard realizations of /r/, even though the two Venetian speakers sometimes pronounced it slightly labialized. The pronunciation of /l/ is much less controversial and the eight speakers had similar realizations. As a matter of fact, Šimáčkova (2009) points out that lambdacism (the deviant pronunciation of the lateral liquid) is much less stigmatized than sigmatism and rhotacism (non-standard realizations of /s/ and /r/, respectively). In order to verify whether greater acoustic salience translated into greater chances of preservation in spontaneous speech, I looked at a number of words containing /r/ and /l/ in the corpus, occurring in different phonological contexts: onset (word-initial), onset (intervocalic), onset (post-consonantal), coda (word-internal), coda (word-final) and as the second member of an OL cluster. Overall, 1102 tokens of /r/ were compared to 256 tokens of /l/. From the raw numbers it is already evident that /r/ is much more frequent in the corpus than /l/. As a matter of fact, /r/ has less restrictions of occurrence, e.g., it forms part of clusters more easily than /l/. I distinguished between the following possible realizations of the two liquids: fully pronounced (FULL; close to [r] and [l], syllabic or not), reduced (RED; e.g., [ɾ, ɹ, ɻ, ɐ] for /r/, [ɾ, j] for /l/), deleted (DEL), assimilated to adjacent segments (ASSIM; e.g., palatalized). The results are given below.

Table 10: Liquids in the onset (word-initially)

/r/	Full	Red	Del	Assim
	71/90	10/90	2/90	7/90
%	**79%**	**11%**	**2%**	**8%**
Realizations	[r]	[ɻ, ɹ, ɾ]	∅	[rʷ]
/l/	Full	Red	Del	Assim
	61/82	4/82	5/82	12/82
%	**74%**	**5%**	**6%**	**15%**
Realizations	[l]	[w]	∅	[ʎ, lʲ, ɫ]

The two liquids in word-initial position occur with similar frequency in the dialogues: overall, 90 tokens of /r/ and 82 tokens of /l/. However, as shown by table 10, /r/ is fully realized more often than /l/, but is also reduced more often. Conversely, /l/ is reduced less often but is more likely to be deleted or assimilated. More specifically, /l/ is particularly prone to palatalization when occurring before a front vowel (as in *linea* 'line', almost always realized with initial [ʎ]) and to velarization when preceding /u/ (as in *lupo* 'wolf', almost always realized with initial [ɫ]). It is important to point out that what I call here "word-initial" context does not always correspond to "utterance-initial", thus, word-initial consonants in spontaneous speech may display the same behavior as intervocalic consonants, e.g., undergo lenition, deletion, etc. Nonetheless, it appears that both liquids are fully pronounced relatively often in this position.

Table 11: Liquids in the onset (intervocalically)

/r/	Full	Red	Del	Assim	Other
	162/205	30/205	11/205	1/205	1/205
%	**79%**	**15%**	**5%**	**0.5%**	**0.5%**
Realizations	[r]	[ɾ, V˞, ʔ, ɻ, ʋ, ɹ]	∅	[rʷ]	[l]
/l/	Full	Red	Del	Assim	Other
	62/82	13/82	6/82	*	1/82
%	**76%**	**16%**	**7%**	**-**	**1%**
Realizations	[l]	[ɾ, r, ɥ]	∅	*[ɥ]	[ɭ]

First of all, the most striking fact is that /r/ occurs intervocalically more than twice as much as /l/, probably by virtue of its higher sonority. Besides this, the two liquids behave similarly when between two vowels. The rhotic is fully pronounced slightly more often than the lateral and is reduced and deleted slightly less often. Reduced realizations of /r/ include [V˞], standing for any rhoticized vowel, and [ʔ], indicating that /r/ might contain an occlusion element, which is not that surprising considering that, articulatorily, a trill is a sequence of very rapid closing and opening gestures. There is one instance of /r/ realized as [l], which cannot really be described as a lenition since the lateral is more consonantal than the rhotic. However, if one follows Scheer (2004) who describes [r] as a trilled [l], this substitution could be interpreted as an articulatory undershoot, where the speaker did not manage to produce the vibration. Among the reduced realizations of /l/, the labiopalatal glide [ɥ] appears in the word *televisore* 'television', and is both reduced (liquid to glide) and assimilated (to the following syllables /vi/, containing both labiality and palatality). The two liquids, being among the most sonorous segments in the phonological inventory, almost never occur after a coda consonant. The only case I could find in the dialogues is word-internal /rl/, of which I analyzed four tokens, realized as follows: [l̰ː, V˞l, rl, rl]. The third and the fourth tokens are unreduced, while the first consists of a glottalized geminate lateral. In the second token, the rhotic colors the preceding vowel but is not fully pronounced. In all four cases, onset /l/ is preserved.

Table 12: Liquids in the coda (word-internally)

/r/	FULL	RED	DEL	ASSIM	OTHER
	238/336	38/336	37/336	22/336	1/336
%	**71%**	**11%**	**11%**	**6.5%**	**0.5%**
Realizations	[r]	[V˞, ʀ, ʔ, ɾ, ɻ, ɾ̃]	∅	[ʂ, ʐ, s, s̰]	[l]
/l/	FULL	RED	DEL	ASSIM	OTHER
	40/73	5/73	20/73	7/73	1/73
%	**55%**	**7%**	**27%**	**9.5%**	**1.5%**
Realizations	[l]	[ʔ, ˡ]	∅	[ɬ, m, n]	[s]

In the word-internal coda, the behavior of the two liquids differs greatly. Again, as in the intervocalic position, /r/ occurs considerably more often than /l/ (336 tokens vs. 73). Moreover, the lateral is deleted more than twice as often as the rhotic (27% vs. 11%). The rhotic is fully pronounced more frequently and is less likely to assimilate. In fact, /r/ is assimilated to the following consonants only in [rs] sequences, as in *verso* 'towards', *forse* 'perhaps', where the rhotic is almost always pronounced as a retroflex sibilant (voiced or voiceless). Alternatively, the whole sequence is produced as a long [s]. Both liquids can be reduced to [ʔ] (although not that often). I interpret the cases where /l/ is velarized as assimilation processes, and not as a neutralization effect of the coda position, given that velarization always occurs after a back vowel. Similarly, realizations of /l/ as [m, n] occur in nasal contexts. It is hard to decide whether the data in table 12 prove the greater salience of /r/ with respect to /l/ or they simply confirm that rhotics are more sonorous than laterals and therefore more apt to occupy the coda.

Table 13: Liquids in the coda (word-finally)

/r/	Full	Red	Del	Assim
	6/13	2/13	1/13	1/13
%	**46%**	**15%**	**8%**	**8%**
Realizations	[r]	[ɾ]	∅	[r̟ᶴ]
/l/	Full	Red	Del	Assim
	5/13	0/13	5/13	3/13
%	**38%**	**0%**	**38%**	**24%**
Realizations	[l]	-	∅	[ʒ, e]

As for word-final liquids, I had to identify tokens of the grammatical words *per* 'for' and *del* 'of the-MASC.SG' that were intelligible enough in order to be acoustically analyzed. The sample is probably too small, but /r/ still appears to be somewhat more resistant than /l/ to deletion and assimilation.

Table 14: Liquids in OL clusters

/r/	Full	Red	Del	Assim
	379/458	17/458	58/458	4/458
%	**83%**	**4%**	**12%**	**1%**
Realizations	[r]	[ɾ, ʔ, V˞]	∅	[ʁ, ʐ]
/l/	Full	Red	Del	Assim
	5/6	1/6	0/6	0/6
%	**84%**	**16%**	**0%**	**0%**
Realizations	[l]	[ʲ]	-	-

If we look at OL clusters, the comparison is particularly unfair. In the four dialogues, I was able to identify 458 tokens of words containing O/r/ clusters and only six containing O/l/ clusters. Of these six, four were instances of /kl/ in *bicicletta* 'bicycle'. These results are not surprising, given that Latin clusters of the O/l/ type evolved into O/j/ and Italian words displaying O/l/ clusters are normally loanwords or belong to the literary/scientific lexicon. Having said that, the resistance of /r/ in a cluster is nevertheless striking: it is fully pronounced as an apical trill 83% of the time, and given the high number of tokens, the percentage is noteworthy.

Table 15: Overall comparison of /r/ and /l/

/r/	FULL	RED	DEL	ASSIM	OTHER
	859/1102	97/1102	109/1102	35/1102	2/1102
%	**78%**	**9%**	**10%**	**3%**	**<1%**
/l/	FULL	RED	DEL	ASSIM	OTHER
	173/256	23/256	36/256	22/256	2/256
%	**68%**	**9%**	**14%**	**9%**	**<1%**

In general, /r/ is represented in the selected dialogues much more than /l/ (almost four times as much). This capacity of /r/ to occur in a wider range of contexts than /l/ may be due both to its greater sonority and its higher acoustic salience. As expected, /r/ is preserved (i.e., unreduced/not deleted) more often than /l/, while both undergo reduction to a similar extent (9% of the tokens for both liquids). The most striking difference is evident in the tendency to assimilate, which is much greater for /l/ than for /r/ (three times higher). However, according to the one-way ANOVA test, the difference in realization between the two liquids is not statistically significant in any position.

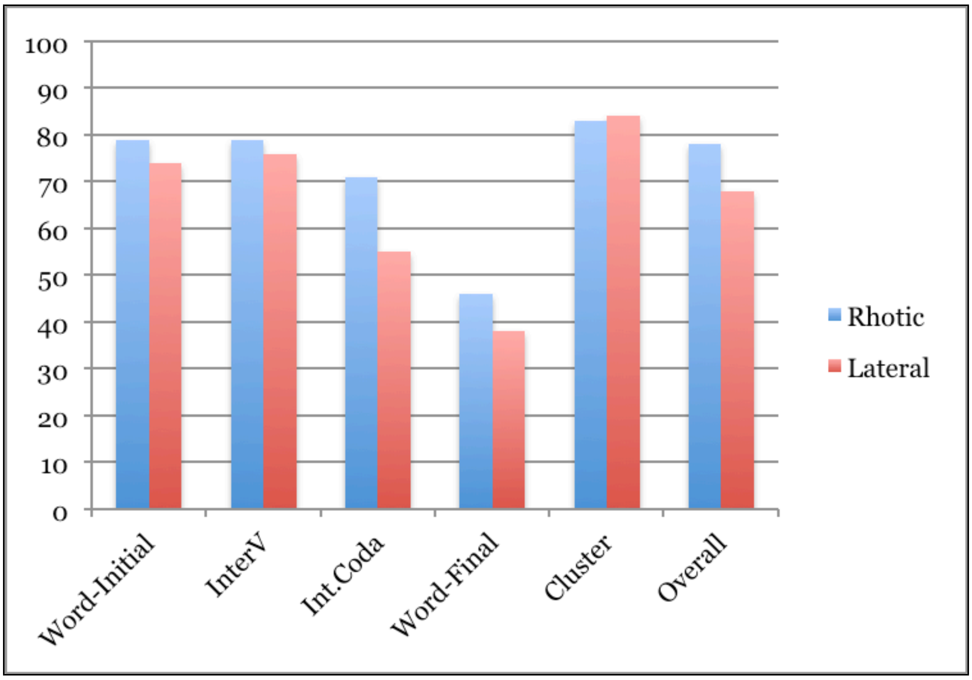


Fig. 6 Full pronunciation of liquids

The graph (fig. 6) shows that /r/ and /l/ are fully pronounced to a similar extent word-initially, intervocalically and in clusters, whereas in the internal and final coda the rhotic appears to be more resistant. Here both the difference between the rhotic and the lateral and the position in the word are significant ($p < .05$ and $p < .01$ respectively), with the latter being more significant (table 16).

Table 16: Two-way ANOVA (Full pronunciation of liquids)

Analysis of variance	**Df**	**SumSq**	**MeanSq**	**F-value**	**Pr(>F)**
Liquid	1	140.08	140.08	8.0124	0.036648*
Position	5	2238.75	447.75	25.6101	0.001428**
Residuals	5	87.42	17.48		

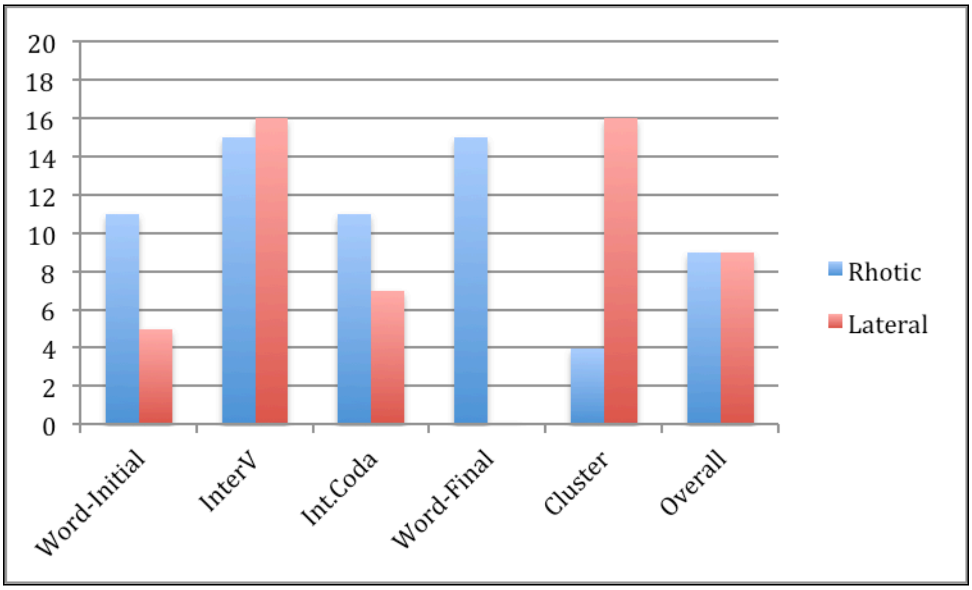


Fig. 7 Reduction rate for liquids

As for reduction (fig. 7), word-finally the lateral is never reduced, whereas in clusters it undergoes reduction to a much greater extent than the rhotic. /r/ proves to be weaker word-initially, but the overall comparison between the two liquids shows an almost identical reduction rate. The lateral appears to be more prone to deletion (fig. 8) in almost all contexts except in clusters, where only /r/ is deleted. Note however, that the clusters containing /l/ are substantially fewer than those containing /r/ in the corpus (and in general in the Italian lexicon). Of the two liquids, the lateral is also the more likely to assimilate to adjacent segments, especially word-initially and word-finally (fig. 9).

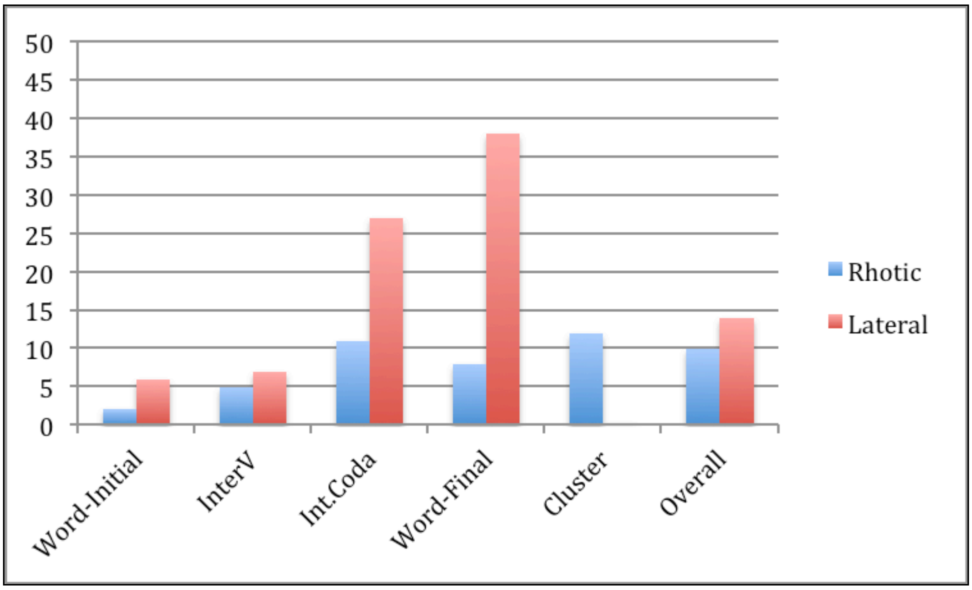


Fig. 8 Deletion rate for liquids

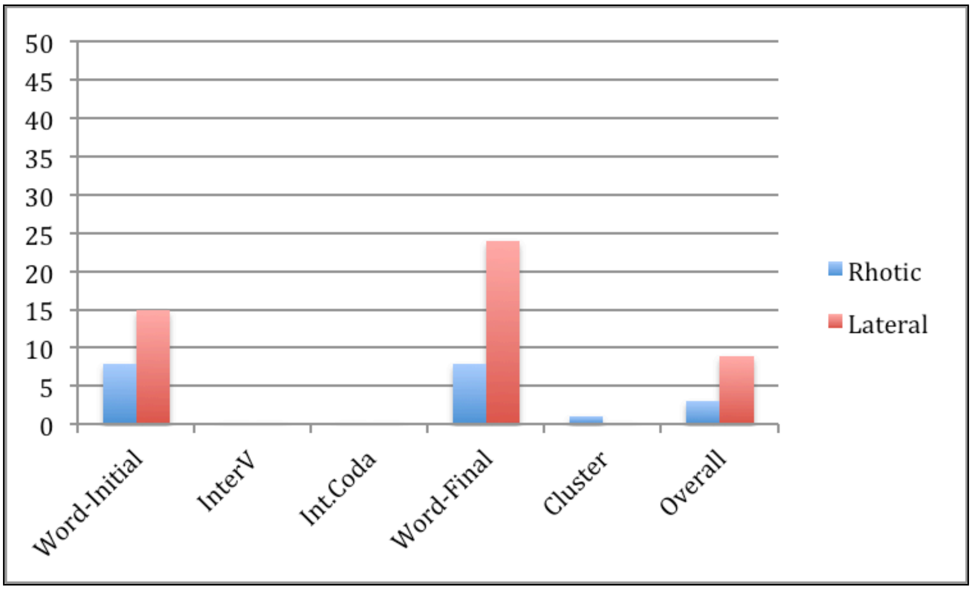


Fig. 9 Assimilation rate for liquids

Importantly, while the full realization of liquids appears to depend heavily on the type of liquid and the position where it occurs, differences in the reduction, deletion and assimilation rates are not statistically significant (*p* > .05). Overall, 1358 tokens of liquids were analyzed, of which 1102 rhotics and 256 laterals. /r/ therefore appears to be significantly more frequent than /l/. As a matter of fact, if the null hypothesis is that the two liquids should occur with the same frequency, one would expect, out of 1358 tokens, 678 /r/'s and 678 /l/'s, but this is not the case, and a chi-squared test proves that the difference between Observed (O) and Expected (E) (Pierrehumbert 2003, Lentz 2011:66) is significant.

Table 17: Liquids O/E ratio

Liquid	**Frequency**	**Observed**	**Expected**	**O/E**
/r/	high	1102	678	1.62
/l/	low	256	678	0.37
Chi squared equals 527.037 with 1 degree of freedom. *p* < .001				

Table 17 shows the O/E ratios of the two liquids. Frequency is high if the O/E ratio is greater than 1, low if it is lower than 1 and normal if it is about 1.

4.3.3.3 Nasals

Italian has three phonologically distinctive nasal consonants, /m/, /n/ and /ɲ/. [ŋ] and [ɱ] only occur before velar and labiodental consonants, respectively. The distribution of /ɲ/ is very limited: it mainly occurs intervocalically, where it is always realized as a geminate, and appears word-initially in a handful of (not so frequent) words. Unlike other nasals, /ɲ/ cannot occupy the coda (unless it is the first half of a geminate), and never occurs word-finally. The two remaining nasals, /m/ and /n/, are distinctive only pre-vocalically (either as singletons or geminates) but neutralize in pre-consonantal context, where nasals take the place features of the following consonants. The results of the experiment indicated that [m] was easier to perceive correctly than [n] in /#mn, m#n, #nm, nm#/ clusters. Therefore, I expect that [n] is more likely to be reduced or deleted in casual speech. Acoustic data extracted from the four CLIPS dialogues confirm these expectations. For nasals, as for liquids, I considered the following possible realizations: FULL, RED (nasal flap, deletion with nasalization of the preceding vowel, placeless nasal), DEL, ASSIM (e.g., palatalized).

Table 18: Nasals in the onset (word-initially)

/m/	FULL	RED	DEL	ASSIM
	27/29	1/29	0/29	1/29
%	**93%**	**3.5%**	**0%**	**3.5%**
Realizations	[m]	[w̃]	-	[m^j]
/n/	FULL	RED	DEL	ASSIM
	19/27	4/27	1/27	3/27
%	**70%**	**15%**	**4%**	**11%**
Realizations	[n]	[ɾ̃, Ṽ]	∅	[ɲ, n^j]

Word-initially both nasals are able to escape reduction, deletion and assimilation, as is to be expected in a strong position. Nonetheless, /n/ proves to be weaker than /m/, being fully pronounced only 70% of the time (vs. 93%). The coronal nasal is also assimilated more easily than the labial one, a fact which could be explained either by resorting to markedness (labials are more marked than coronals) or to articulatory factors (the labial gesture tends to mask other gestures, therefore labials rarely assimilate).

Intervocalically (see table 19), again, /m/ is fully pronounced more often than /n/. However, /m/ is deleted in 10% of the tokens and /n/ only in 4%. Crucially, most of the tokens of intervocalic /m/ consisted of the initial segment of the adverbial suffix *–mente* (grossly corresponding to English –*ly*), so the higher rate of deletion for /m/ could also be due to its predictability in that position.

Table 19: Nasals in the onset (intervocalically)

/m/	FULL	RED	DEL	ASSIM	OTHER
	103/127	9/127	13/127	1/127	1/127
%	**81%**	**7%**	**10%**	**1%**	**1%**
Realizations	[m]	[β̃, m, Ṽ, ʔ, ɥ]	∅	[ɲ]	[b]
/n/	FULL	RED	DEL	ASSIM	OTHER
	98/156	20/156	6/156	32/156	-
%	**63%**	**13%**	**4%**	**20%**	-
Realizations	[n]	[ɾ̃, ʔ, ʔ̃, N, Ṽ, j]	∅	[ɲ, nʲ, m]	-

Table 20: Nasals in the onset (word-internally, post-consonantally)

/m/	Full	Red	Del	Assim
	22/23	0	0	1/23
%	**96%**	**0%**	**0%**	**4%**
Realizations	[m]	-	-	[ɲ]
/n/	Full	Red	Del	Assim
	38/38	-	-	-
%	**100%**	-	-	-
Realizations	[n]	-	-	-

The word-internal, post-consonantal onset position is known to be one of the strongest (i.e., less likely to undergo deletion/reduction) and unsurprisingly, both nasals are fully pronounced in most of the tokens, with no significant differences (see table 20).

In the internal coda, /n/ occurs significantly more often than /m/ (see table 21). This is not surprising, given that the phoneme /n/, in Italian, includes also the allophones [ŋ, ɱ], which occur before velar and labiodental consonants. Therefore, the higher rate of assimilation of /n/ is to be expected. Moreover, as pointed out by Greenlee & Ohala (1980), [ŋ] is the nasal consonants that most resembles to a nasal vowel, and in fact most of the realizations of /n/ as nasality of the preceding vowel occur before a velar consonant. /m/ is restricted to appear only before another labial consonants, which clearly limits its combinatory possibilities (only 28 tokens of coda /m/).

Table 21: Nasals in the coda (word-internally)

/m/	FULL	RED	DEL	ASSIM
	23/28	4/28	1/28	0/28
%	**82%**	**14%**	**4%**	**0%**
Realizations	[m]	[Ṽ, N]	∅	-
/n/	FULL	RED	DEL	ASSIM
	124/293	136/293	14/293	19/293
%	**42%**	**47%**	**5%**	**6%**
Realizations	[n]	[Ṽ, N, V̰]	∅	[m, ŋ]

Table 22: Overall comparison between /m/ and /n/

/m/	FULL	RED	DEL	ASSIM	OTHER
	175/207	14/207	14/207	3/207	1/207
%	**85%**	**7%**	**7%**	**1%**	**<1%**
/n/	FULL	RED	DEL	ASSIM	OTHER
	279/514	160/514	21/514	54/514	-
%	**54%**	**31%**	**4%**	**11%**	-

In total, 514 tokens of /n/ and 207 tokens of /m/ were analyzed, suggesting that /n/ is more than twice as frequent as /m/ (see table 22). /m/ was fully pronounced more often than /n/, in almost all positions (fig. 10). /n/, in turn, underwent deletion and assimilation to a considerably greater extent (fig. 11, fig. 12). However, it was /m/ that was deleted more often, most notably in intervocalic position (fig. 13). It must pointed out, though, that in any of the position considered the difference between /m/ and /n/ proved to be significant according to the one-way ANOVA test.

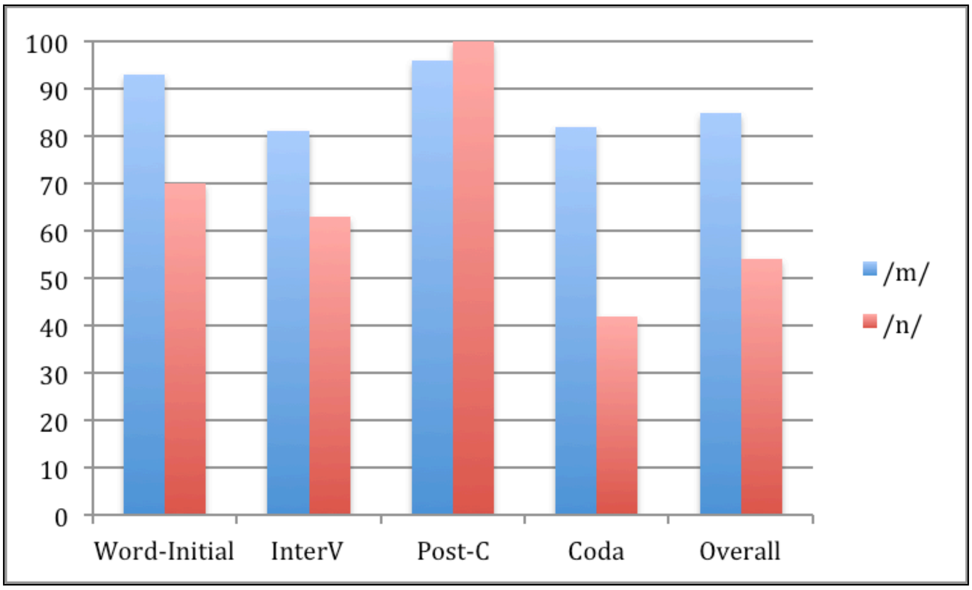


Fig. 10 Full realization of nasals

/m/ is significantly fully realized more often than /n/ (*p* < .05), whereas their position in the word does not seem to have a relevant effect on their pronunciation (*p* > 0.5, see table 23).

Table 23: Two-way ANOVA (Full realization of nasals)

Analysis of variance	**Df**	**SumSq**	**MeanSq**	**F-value**	**Pr(>F)**
Nasal	1	1166.4	1166.4	8.5046	0.04341*
Position	4	1533.4	383.35	2.7951	0.17170
Residuals	4	548.6	137.15		

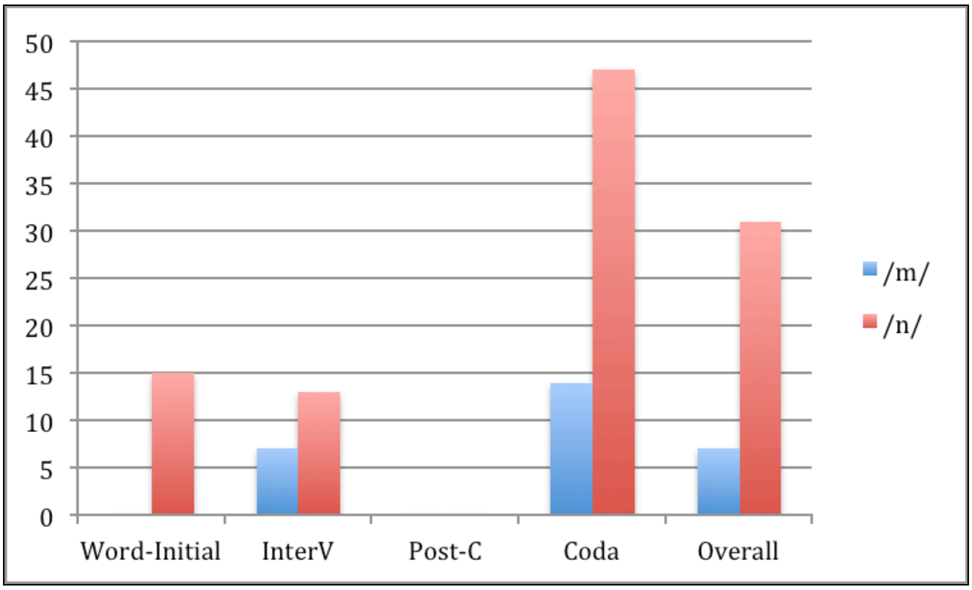


Fig. 11 Reduction rate for nasals

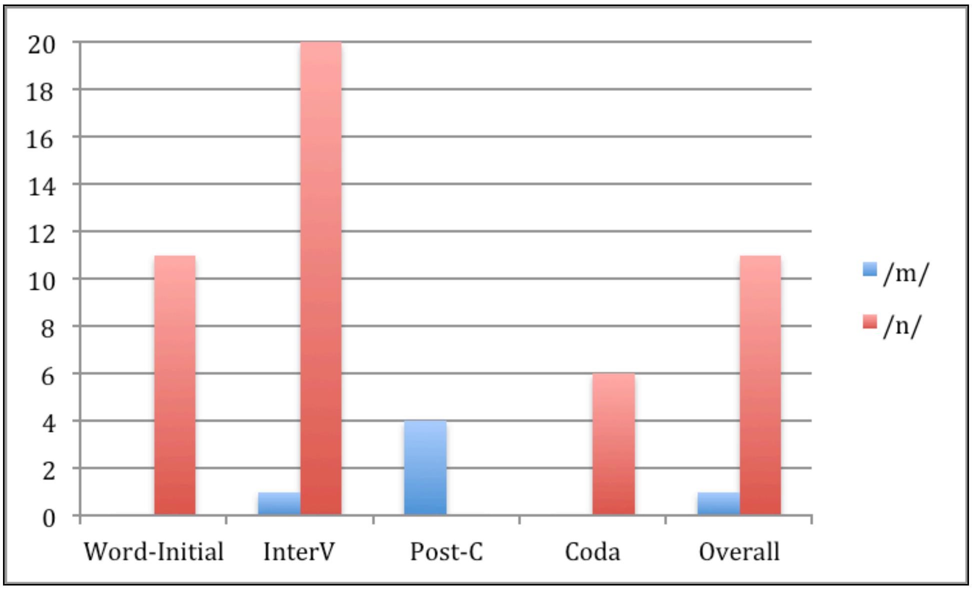


Fig. 12 Assimilation rate for nasals

The difference in reduction, deletion and assimilation rate between the two nasals does not appear to be significant and neither does their position in the word. As for liquids, though, the difference in frequency is significant, according to the chi-square test.

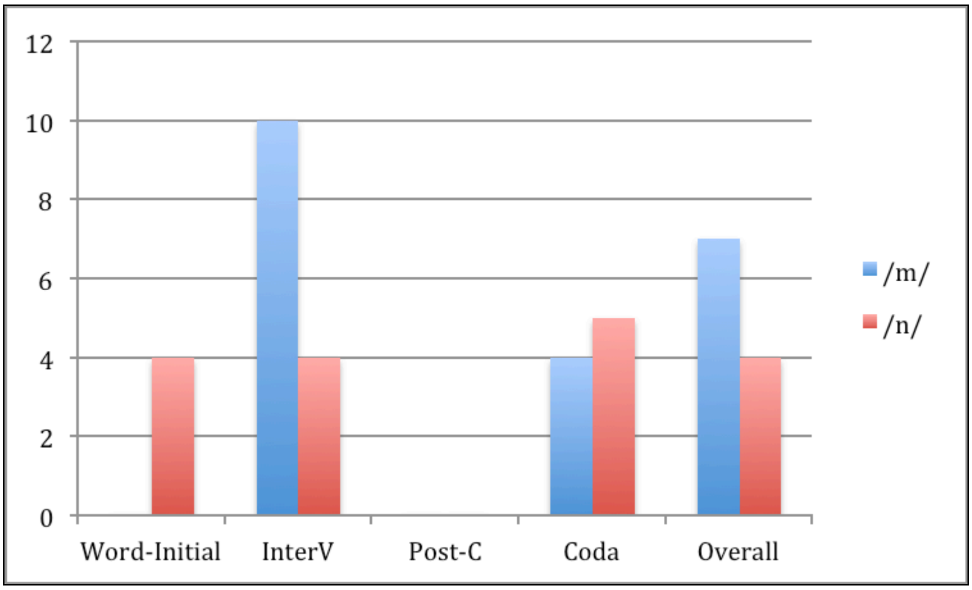


Fig. 13 Deletion rate for nasals

Table 24: Nasals O/E ratio

Nasal	**Frequency**	**Observed**	**Expected**	**O/E**
/m/	low	207	256	0.80
/n/	normal	514	465	1.10
Chi squared equals 14.542 with 1 degree of freedom. $p < .001$.				

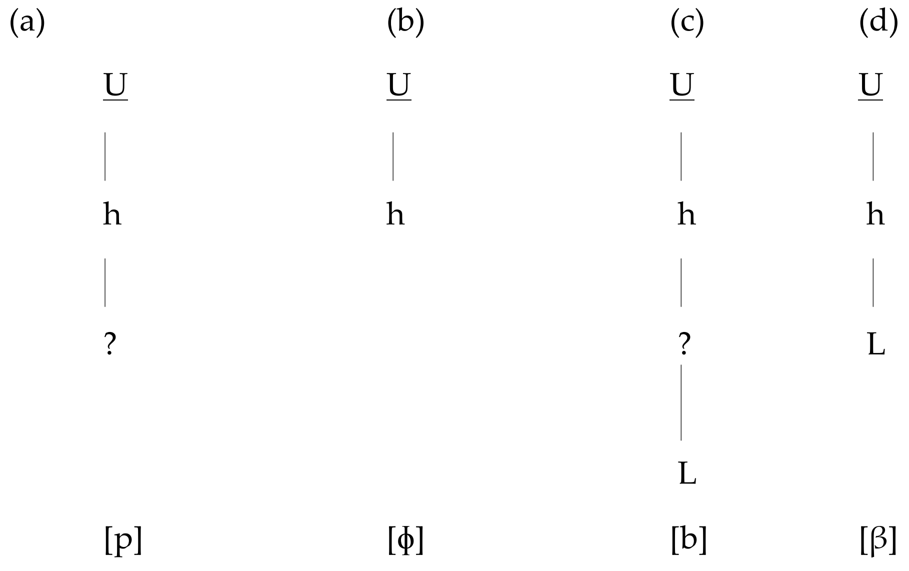
The O/E ratio of both /m/ and /n/ is close to 1, but the frequency of /m/ qualifies as low because fewer tokens than expected were found. The expected values were calculated as follows: overall, the tokens of nasals were 721, of which 321 in the coda. Since the phoneme /n/ occurs before /t, k, f, s/ and /m/ only before /p/, the values were adjusted for the token frequency of the obstruents.

4.3.3.4 Obstruents

In this section I will analyze the realizations of the obstruents [s, ʃ, f, k, p, t] in the different contexts in which they occur in the dialogues. I chose not to consider their voiced counterparts for two main reasons: (a), the voiced sibilants in Italian are not phonologically distinctive, i.e., [z] is an allophone of /s/ and /ʒ/ does not occur in the native lexicon, and (b), since in the perception experiment (chap. 3) only voiceless obstruents were employed, a direct comparison between perceptibility and preservation in casual speech could not be made for voiced obstruents. As for liquids and nasals, I considered four different possible realizations: FULL, RED, DEL, ASSIM. However, it was not always easy to understand whether a given consonant had undergone reduction, assimilation or both. For ease of exposition, I sometimes had to classify certain processes in a seemingly arbitrary way. Namely, if an obstruent was realized as its voiced counterpart adjacent to a sonorant, it was classified as assimilation. Conversely, any other type of lenition (e.g., fricativization, voicing and fricativization, glottalization, etc.) was treated as reduction. For instance, if underlying /p/ (represented in 66a) is realized as [b] between two vowels, it would count as a case of assimilation (as in 66c), whereas if /p/ is pronounced [ɸ], it would be considered as a type of reduction (as in 66b). Basically, a process qualifies as reduction if a manner element is lost and as assimilation if an element (manner or melody) is acquired. When **?** is lost and **L** is acquired (as in 66d), the process is still considered a reduction process, since I assume that in obstruents the opposition in continuancy is more perceptually salient than the opposition in voicing.

In word-initial position (see table 25), all obstruents are preserved (fully realized) most of the time, with /k/ being the most likely to be deleted (14% of the tokens) and unsurprisingly the fricatives being the strongest group. It has to be pointed out, though, that /ʃ/ and /t/ occur quite rarely word-initially, at least in the selected dialogues. Whereas this low frequency rate is to be expected for the post-alveolar sibilant, it is quite striking for /t/, which is normally considered to be the most frequent/least marked consonant. /s/ undergoes fortition in two tokens, being realized as an affricate, whereas /k/ is pronounced as [t] in three tokens, which could be due to the preference for more front consonants to occur word-initially.

(66)



Between two vowels (see table 26), obstruents in general are quite infrequent. As a matter of fact, in intervocalic position Italian prefers obstruents to be geminate (and [ʃ] is actually phonologically a geminate and realized as such in most tokens). /s/ and /p/ are rare in this position, with /s/ almost always realized as [z], which is the standard in Turin and Venice but not in Rome and Palermo. Intervocalic /f/ appears in only four tokens, whereas /k/ and /t/ have the highest rates of occurrence. However, if the former is easily reduced or assimilated between two vowels, /t/ appears to be more resistant. Considering that, crosslinguistically, labials prefer to occur initially and dorsals finally, this could be an indication that coronals are fit for the intervocalic position, a fact corroborated by their use as epenthetic segments to break potential hiatuses in a number of languages.

Table 25: Obstruents in the onset (word-initially)

/s/	FULL	RED	DEL	ASSIM	OTHER
	188/192	0/192	0/192	4/192	2/192
%	**97%**	**0%**	**0%**	**2%**	**1%**
Realizations	[s]	-	-	[z]	[ts, ^{d}z]
/ʃ/	FULL	RED	DEL	ASSIM	OTHER
	37/38	0/38	0/38	1/38	-
%	**97%**	**0%**	**0%**	**3%**	-
Realizations	[ʃ]	-	-	[ɕ]	-
/f/	FULL	RED	DEL	ASSIM	OTHER
	95/97	0/97	0/97	2/97	-
%	**98%**	**0%**	**0%**	**2%**	-
Realizations	[f]	-	-	[v]	-
/k/	FULL	RED	DEL	ASSIM	OTHER
	162/211	11/211	29/211	6/211	3/211
%	**78%**	**5%**	**14%**	**3%**	**<1%**

Realizations	[k]	[j, ɰ, x, h, k̚, ɣ]	∅	[g, kʲ, v]	[t]
/p/	FULL	RED	DEL	ASSIM	OTHER
	189/208	14/208	1/208	4/208	-
%	**91%**	**7%**	**1%**	**2%**	-
Realizations	[p]	[v, β, w, ɸ, f]	∅	[b]	-
/t/	FULL	RED	DEL	ASSIM	OTHER
	36/38	0/38	0/38	2/38	-
%	**95%**	**0%**	**0%**	**5%**	-
Realizations	[t]	-	-	[d]	-

Table 26: Obstruents in the onset (intervocalically)

/s/	FULL	RED	DEL	ASSIM	OTHER
	7/25	0/25	0/25	16/25	2/25
%	**28%**	**0%**	**0%**	**64%**	**8%**
Realizations	[s]	-	-	[z]	[θ, ð]
/ʃ/	FULL	RED	DEL	ASSIM	OTHER
	23/23	0/23	0/23	0/23	-
%	**100%**	**0%**	**0%**	**0%**	-
Realizations	[ʃ]	-	-	-	-
/f/	FULL	RED	DEL	ASSIM	OTHER
	4/4	0/4	0/4	0/4	-
%	**100%**	**0%**	**0%**	**0%**	-
Realizations	[f]	-	-	-	-
/k/	FULL	RED	DEL	ASSIM	OTHER

	14/57	19/57	10/57	14/57	-
%	**24.5%**	**33%**	**18%**	**24.5%**	-
Realizations	[k]	[j, x]	∅	[ɟ, g̊, g]	-
/p/	Full	Red	Del	Assim	Other
	15/23	7/23	0/23	1/23	-
%	**65%**	**30%**	**0%**	**5%**	-
Realizations	[p]	[ɸ, β, w, v]	-	[b]	-
/t/	Full	Red	Del	Assim	Other
	56/100	23/100	14/100	7/100	-
%	**56%**	**23%**	**14%**	**7%**	-
Realizations	[t]	[ð, t̚, h, j, tˢ, ɾ, r]	∅	[d]	-

Table 27: Obstruents in the onset (word-internally, post-consonantally)

/s/	Full	Red	Del	Assim	Other
	95/104	0/104	0/104	2/104	7/104
%	**91%**	**0%**	**0%**	**2%**	**7%**
Realizations	[s]	-	-	[ʂ]	[ts]*
/f/	Full	Red	Del	Assim	Other
	46/46	0/46	0/46	0/46	-
%	**100%**	**0%**	**0%**	**0%**	-
Realizations	[f]	-	-	-	-
/k/	Full	Red	Del	Assim	Other
	110/122	2/122	2/122	6/122	2/122
%	**90%**	**1.6%**	**1.6%**	**5%**	**1.6%**

Realizations	[k]	[ʔ]	∅	[c]	[q, ɸ]
/p/	FULL	RED	DEL	ASSIM	OTHER
	73/76	3/76	0/76	0/76	-
%	**96%**	**4%**	**0%**	**0%**	-
Realizations	[p]	[pɸ]	-	-	-
/t/	FULL	RED	DEL	ASSIM	OTHER
	580/741	21/741	88/741	40/741	-
%	**78%**	**3%**	**12%**	**5%**	-
Realizations	[t]	[ts, ɾ, θ, ʔ, ð]	∅	[d]	-

In the word-internal, post-consonantal position (see table 27), coronals appear to be particularly frequent, with 104 tokens of /s/ and 741 tokens of /t/, followed by the velar stop, with 122 tokens. As expected, in this strong syllable-initial position all obstruents tend to be fully pronounced, with /t/ being the weakest, as it is deleted in 12% of the tokens. The post-alveolar sibilant does not occur in any token.

In Italian obstruents cannot appear in the coda unless they are the first half of a geminate. The only exception is /s/, which can precede another consonant both word-initially (where it is traditionally analyzed as extrasyllabic or as part of a coda, cf. Iverson 1990, Kaye 1990) and word-internally (and in a few words even word-finally, e.g., *autobus* 'bus'). Tables 28 and 29 present the results for /s/ word-initially and word-internally.

Table 28: Word-initial /s/+C

/s/	FULL	RED	DEL	ASSIM
	125/125	0/125	0/125	0/1245
%	**100%**	**0%**	**0%**	**0%**
Realizations	[s]	-	-	-

Table 29: /s/ in the coda (word-internally)

/s/	FULL	RED	DEL	ASSIM	OTHER
	296/304	1/304	3/304	1/304	3/304
%	**97%**	**0.5%**	**1%**	**0.5%**	**1%**
Realizations	[s]	[h]	∅	[z]	[θ, ð]

Both initially and medially, /s/ is preserved in /s/+C clusters, which are quite numerous in the dialogues: 125 word-initial and 304 medial /s/+C clusters. Given the salience of the sibilant and its crosslinguistic unmarked status, these results are not surprising and confirm the hypothesis that salient consonants tend to be preserved despite articulatory complexity and lack of vocalic support.

Table 30: Obstruents in OL clusters (word-initially)

/f/	Full	Red	Del	Assim
	11/11	0/11	0/11	0/11
%	**100%**	**0%**	**0%**	**0%**
Realizations	[f]	-	-	-
/k/	Full	Red	Del	Assim
	5/9	2/9	0/9	2/9
%	**56%**	**22%**	**0%**	**22%**
Realizations	[k]	[ɰ]	-	[g̚]
/p/	Full	Red	Del	Assim
	86/98	6/98	1/98	5/98
%	**88%**	**6%**	**1%**	**5%**
Realizations	[p]	[v, β, ɸ]	∅	[b, ʙ, t]
/t/	Full	Red	Del	Assim
	48/49	1/49	0/49	0/49
%	**98%**	**2%**	**0%**	**0%**
Realizations	[t]	[t^{s}]	-	-

It is unusual for sibilants to combine with liquids in clusters, and as a matter of fact, SL clusters are not found in the dialogues. The obstruent most likely to combine with a liquid in a cluster is the labial /p/, followed by /t/ (see table 30). Word-internally (see table 31), I was only able to find OL clusters beginning with a stop. Here, /tr/ is the most frequent, followed by /pr/. Before a sonorant /p/ is more likely to undergo reduction, whereas /t/ is deleted in 24% of tokens. In the few tokens with /kL/, /k/ is never reduced or deleted.

Table 31: Obstruents in OL clusters (word-internally)

/k/	Full	Red	Del	Assim
	6/6	0/6	0/6	0/6
%	**100%**	**0%**	**0%**	**0%**
Realizations	[k]	-	-	-
/p/	Full	Red	Del	Assim
	61/84	15/84	1/84	7/84
%	**73%**	**18%**	**1%**	**8%**
Realizations	[p]	[ʔ, v, ɸ, β, w, f, pɸ]	∅	[b, ʁ]
/t/	Full	Red	Del	Assim
	119/188	21/188	45/188	3/188
%	**63%**	**11%**	**24%**	**2%**
Realizations	[t]	[ð, θ, ʔ, t^s, t^θ]	∅	[s, d]

Table 32 presents a general overview of the phonetic realization of obstruents in the four dialogues extracted from CLIPS. By looking at the percentages, it appears that the most resistant obstruent (the least likely to undergo reduction, deletion and assimilation) is /f/, followed by /ʃ/ and /s/. It would seem, therefore, that fricatives are more resistant than stops, which is not unexpected, given their longer duration and, in the case of stridents, their higher salience with respect to stops. Among the stops, the most resistant is /p/, followed by /t/ and /k/. Importantly, obstruents appear to behave overall similarly in Italian, since the one-way ANOVA test did not prove their difference to be significant in any position.

Table 32: Overall comparison between obstruents

/s/	FULL	RED	DEL	ASSIM	OTHER
	711/750	1/750	3/750	23/750	14/750
%	**95%**	**<0.5%**	**<0.5%**	**3%**	**2%**
Realizations	[s]	[h]	∅	[z]	[ð, θ, ts]
/ʃ/	FULL	RED	DEL	ASSIM	OTHER
	60/61	0/61	0/61	1/61	-
%	**98%**	**0%**	**0%**	**2%**	-
Realizations	[ʃ]	-	-	[ɕ]	-
/f/	FULL	RED	DEL	ASSIM	OTHER
	156/158	0/158	0/158	2/158	-
%	**99%**	**0%**	**0%**	**1%**	-
Realizations	[f]	-	-	[v]	-
/k/	FULL	RED	DEL	ASSIM	OTHER
	291/399	34/399	41/399	28/399	5/399
%	**73%**	**8.5%**	**10%**	**7%**	**1.5%**
Realizations	[k]	[j, ɰ, x, h, k̚, ɣ, ʔ]	∅	[g, kʲ, v, ɟ, c]	[t, ɸ, q]
/p/	FULL	RED	DEL	ASSIM	OTHER
	424/488	44/488	3/488	17/488	-
%	**87%**	**9%**	**1%**	**3%**	-
Realizations	[p]	[v, w, f, β, ɸ, pɸ, ʔ]	∅	[b, t, ʙ]	-

/t/	FULL	RED	DEL	ASSIM	OTHER
	720/928	45/928	102/928	49/928	-
%	**78%**	**5%**	**12%**	**5%**	-
Realizations	[t]	[h, j, r, ts, tˢ, ð, t̚, ɾ, θ, ʔ]	∅	[d]	-

(67) Full realization rate:
f > ʃ > s > p > t > k

The most likely obstruent to be lenited or somehow reduced in pronunciation is /p/, followed very closely by /k/ and then /t/, whereas fricatives are almost never reduced.

(68) Reduction rate:
p > k > t > s > ʃ, f

As for deletion, the obstruent that undergoes deletion the most often is /t/, followed by /k/, while /p/ and the fricatives are almost never deleted.

(69) Deletion rate:
t > k > p > s, f, ʃ

The obstruent which most often assimilates to adjacent segments is /k/, followed by /t/, which is to be expected, given that these two consonants are normally described as the “weakest”, i.e., the more likely to be the target of assimilatory processes (most notably, palatalization). However, in the case of /t/, its assimilation always consisted of voicing. The same can be said for /s/, which is voiced intervocalically (or at least in sonorant context) in 3% of its occurrences. /p/ undergoes assimilation in the same percentage of instances, followed by /ʃ/ and /f/.

(70) Assimilation rate:
k > t > s, p > ʃ > f

Nonetheless, the data presented in table 32 have to be filtered by the frequency of occurrence of each obstruent in the dialogues. For instance, the post-alveolar palatal only occurs 61 times, as opposed to the 928 words containing (or expected to contain) /t/. As shown in figure 14, the coronal consonants /t/ and /s/ are the most frequent, followed by the labial and the velar stop, while /ʃ/ and /f/ are relatively rare.

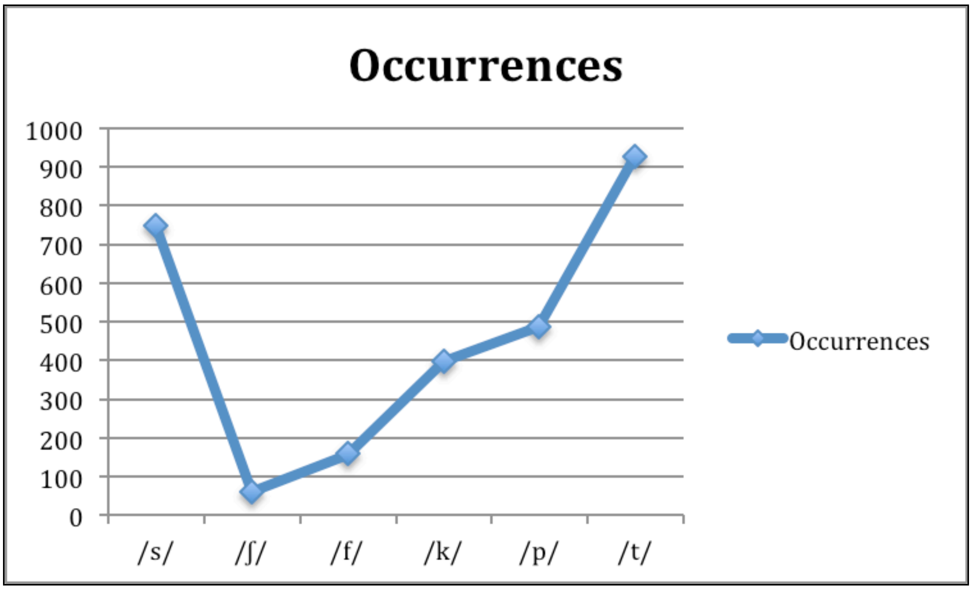


Fig. 14 Absolute frequency of obstruents

In total, 2784 obstruents were analyzed, but comparing their O/E ratios, it turns out that /s/ and /t/ are significantly more frequent than /p/ and /k/, which in turn are significantly more frequent than /f/ and /ʃ/.

The expected values in table 33 are based on the null hypothesis that all obstruents have the same chance to occur (464 = 2784/6). The frequency rate of /k/ and /p/ almost equals the expected values, whereas /s/ and /t/ prove to be highly frequent and /f/ and /ʃ/ occur much less than expected. These results fit very nicely with the markedness hypothesis, since unmarked coronal obstruents occur considerably more often than dorsal and labial ones, and fricatives are notably rarer than stops (with the exception of /s/).

Table 33: Obstruents O/E ratios

Obstruent	**Frequency**	**Observed**	**Expected**	**O/E**
/s/	high	750	464	1.61
/ʃ/	low	61	464	0.13
/f/	low	158	464	0.34
/k/	normal	399	464	0.85
/p/	normal	488	464	1.05
/t/	high	928	464	2
Chi squared equals 1202.453 with 5 degrees of freedom. $p < .001$.				

Besides absolute frequency, obstruents tend to appear more frequently in certain positions in the word. For instance, /s/ shows a clear preference for occurring in the coda, including the “extra-syllabic” coda in word-initial

#SO clusters and word-internal coda, whereas it disfavors rising sonority clusters and the intervocalic position (fig. 15). As already stated, /ʃ/ cannot combine freely in Italian, and as figure 16 shows, it only occurs word-initially and intervocalically, with a clear preference for the former position. /f/ is mainly found in the onset, especially word-initially and before vowels. It can occur, albeit rarely, as the first member of an OL cluster, but is basically absent from the other positions. Contrary to the crosslinguistic tendency for /k/ not to be initial, in the sample under analysis /k/ is mostly found word-initially (fig. 17). It sometimes appears word-internally after a consonant and intervocalically, whilst it hardly ever forms clusters with sonorants. Also /p/ is word-initial most of the time, is frequently the first element of an initial cluster, but is harder to find intervocalically. Finally, the distribution of /t/ is notable with respect to the other obstruents, in that it appears more often word-internally than word-initially (fig. 18). It is most frequently found post-consonantally (after a sonorant or /s/) and intervocalically.

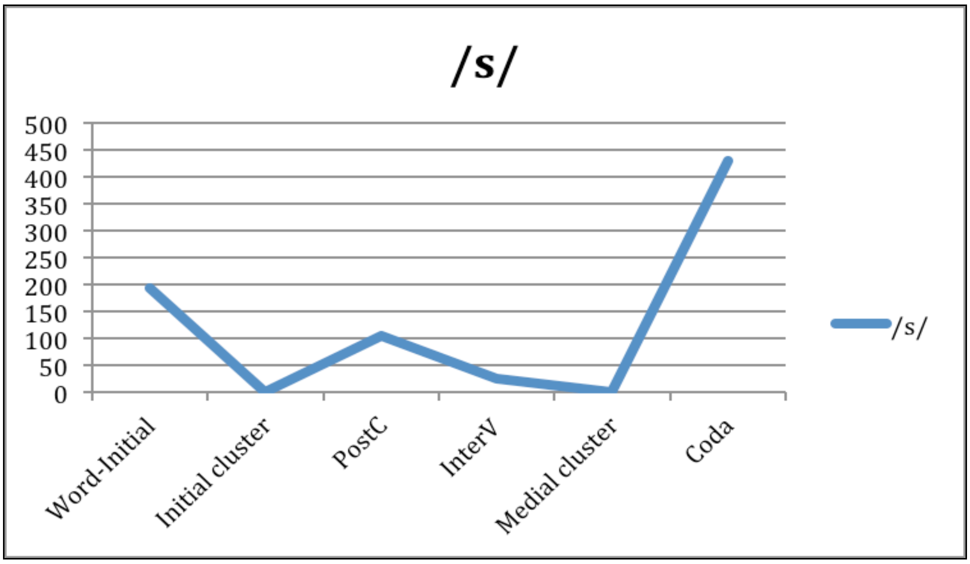


Fig. 15 Number of occurrences of /s/ in each position

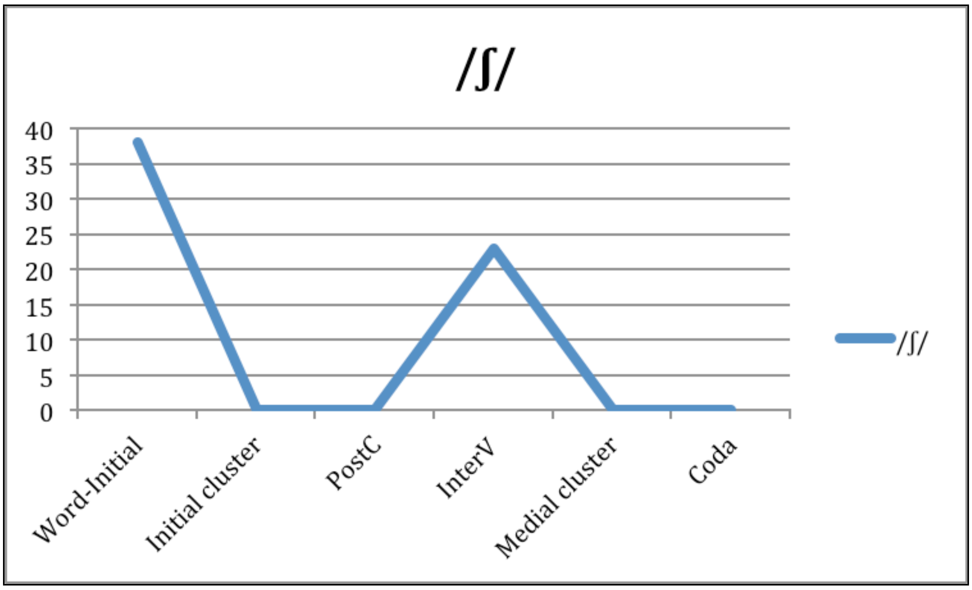


Fig. 16 Number of occurrences of /ʃ/ in each position

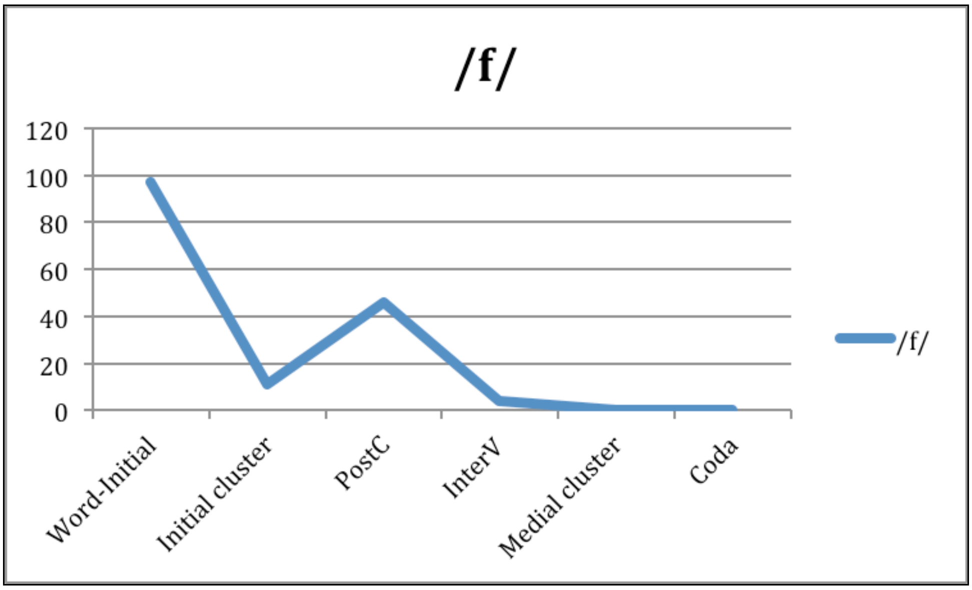


Fig. 17 Number of occurrences of /f/ in each position

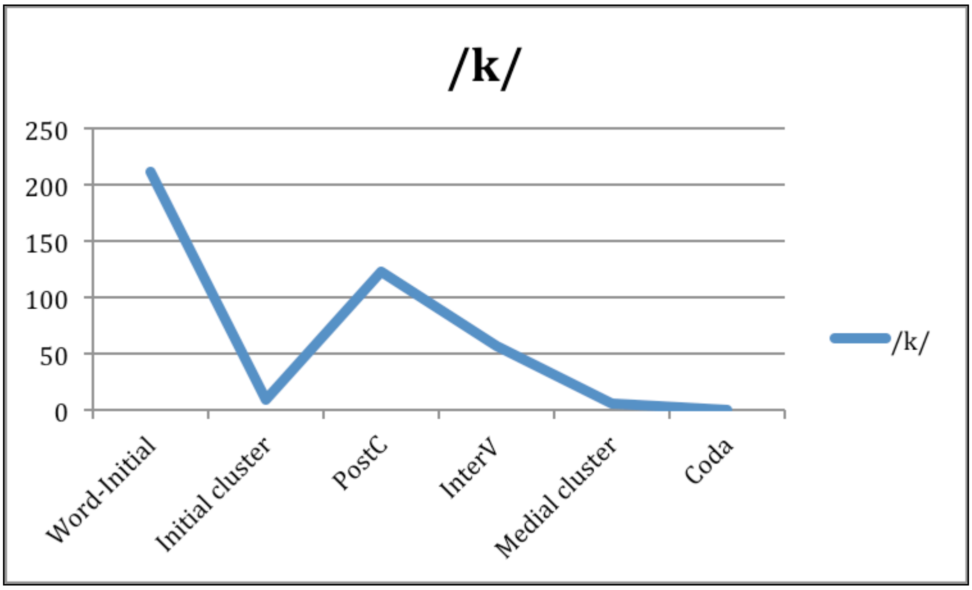


Fig. 18 Number of occurrences of /k/ in each position

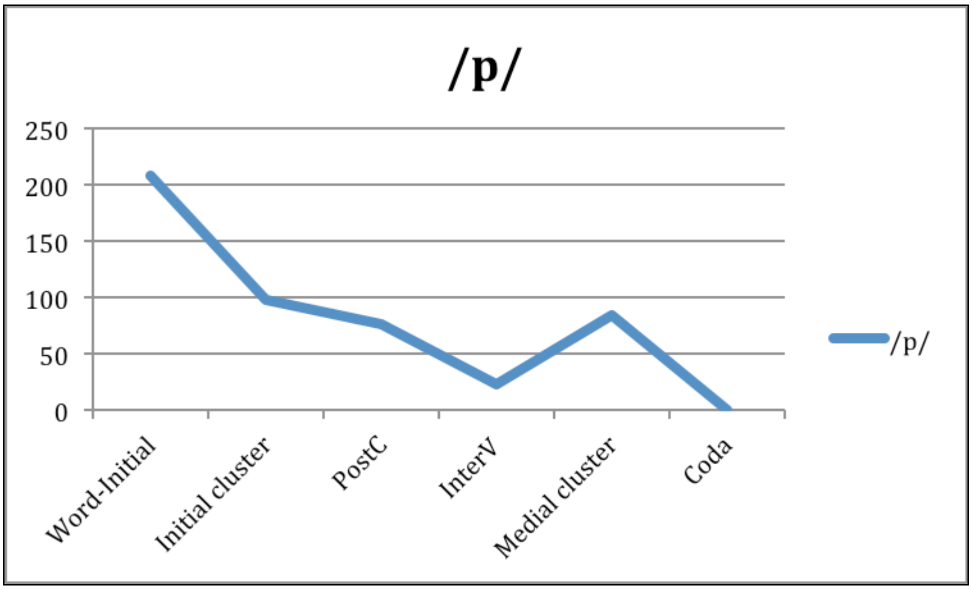


Fig. 19 Number of occurrences of /p/ in each position

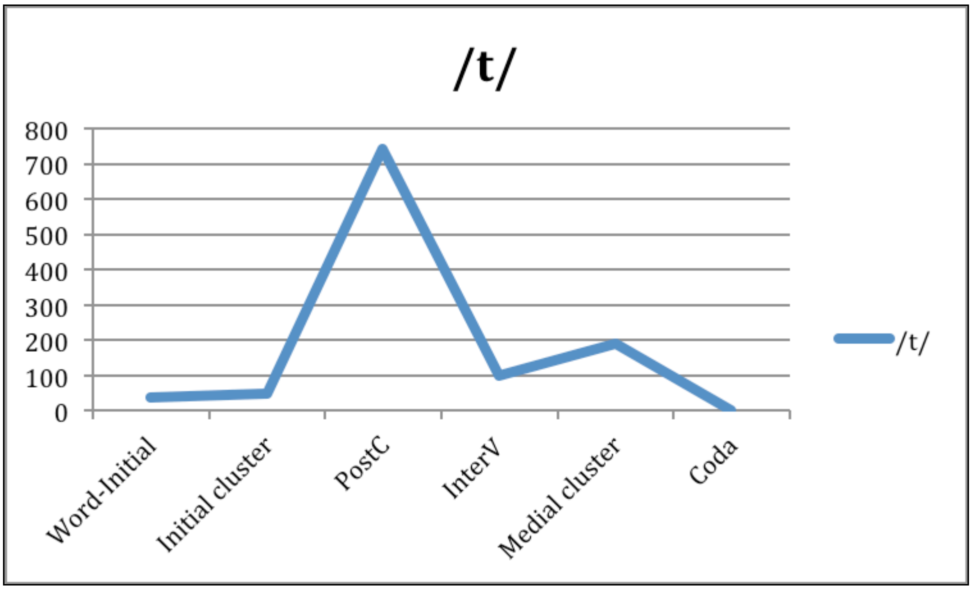


Fig. 20 Number of occurrences of /t/ in each position

The six obstruents under examination also display different behaviors depending on the position in which they occur. Figure 21 shows that /s/ is mostly pronounced [s] except intervocalically, where it assimilates to the voicing of the surrounding vowels. Figures 22 and 23 show the behavior of the other two fricatives: the post-alveolar sibilant and the labiodental fricative are basically always fully pronounced.

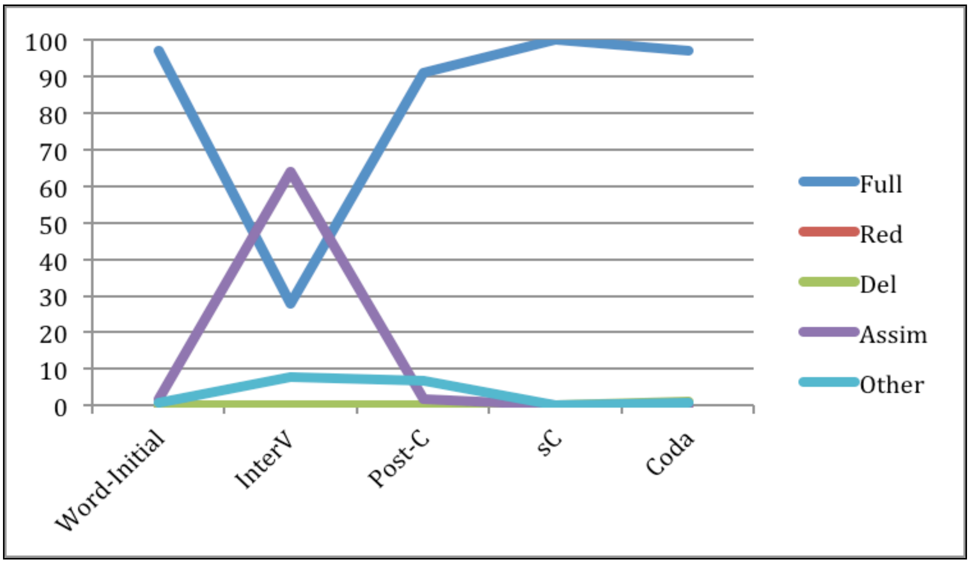


Fig. 21 Realizations of /s/ in each position

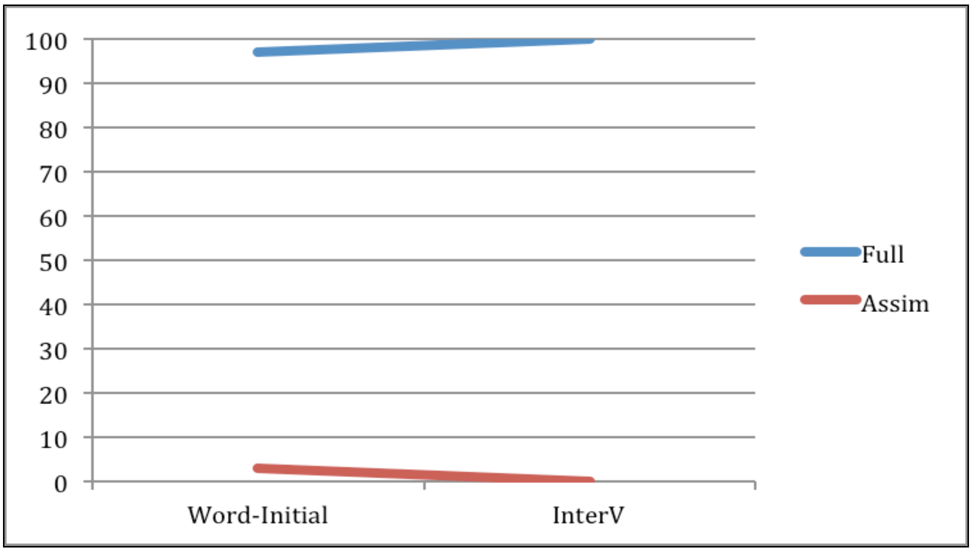


Fig. 22 Realizations of /ʃ/ in each position

The realization of the three stops varies to a much greater extent. /k/ (fig. 39) is mostly pronounced [k] word-initially, post-consonantally and as the first member of a medial cluster, but its reduction rate rises significantly intervocalically. The intervocalic position is also the one in which /k/ is more frequently deleted, whereas assimilation occurs both intervocalically and before a sonorant in initial clusters.

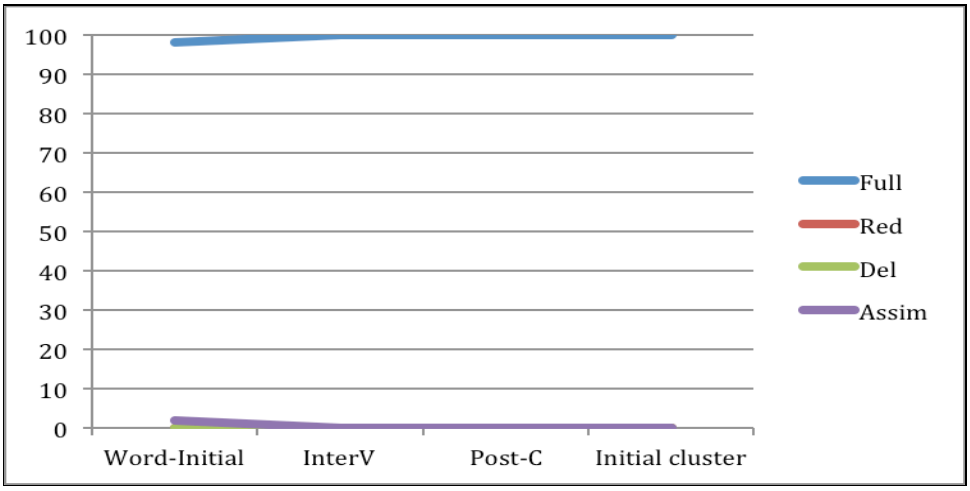


Fig. 23 Realizations of /f/ in each position

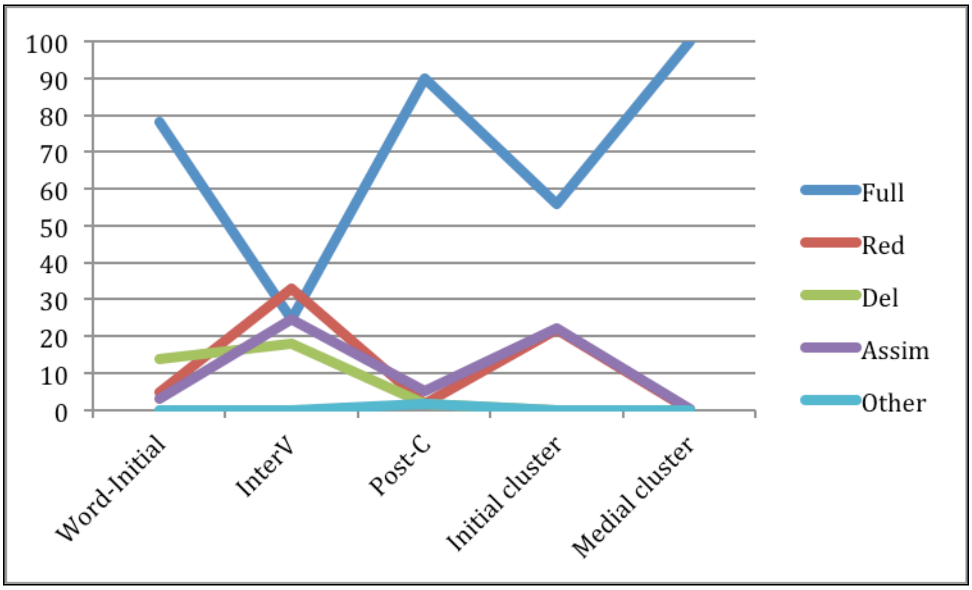


Fig. 24 Realizations of /k/ in each position

As for /p/ (fig. 25), [p] is a common pronunciation not only word-initially and post-consonantally but also intervocalically. However, between two vowels /p/ undergoes reduction quite significantly. Cases of assimilation and deletion appear to be infrequent. Finally, figure 26 shows that /t/ is mostly pronounced [t]. Intervocalically, it is more often reduced than deleted or assimilated, whereas deletion is more frequent in medial OL clusters.

Comparing the realization of all the obstruents and the position in which they occur, the only factor which turns out to be statistically significant is the position in the word as a predictor for both full pronunciation and reduction. The difference between obstruents did not prove to be significant, suggesting that they all behave similarly in casual speech, and neither did frequency. Tables 30 and 31 show the results of the two-way ANOVA tests checking for the significance of the correlation between obstruent type and position (word-initial, intervocalic, post-consonantal) in the cases of full realization (table 33) and in the cases of reduction (table 34). Both tables show that position is significant ($p < .05$) but obstruent type is not ($p > .05$), suggesting that all obstruents tend to be fully pronounced in the same context (mostly word-initially and post-consonantally) and reduced in the same context (intervocalically), with no substantial difference within the obstruent class.

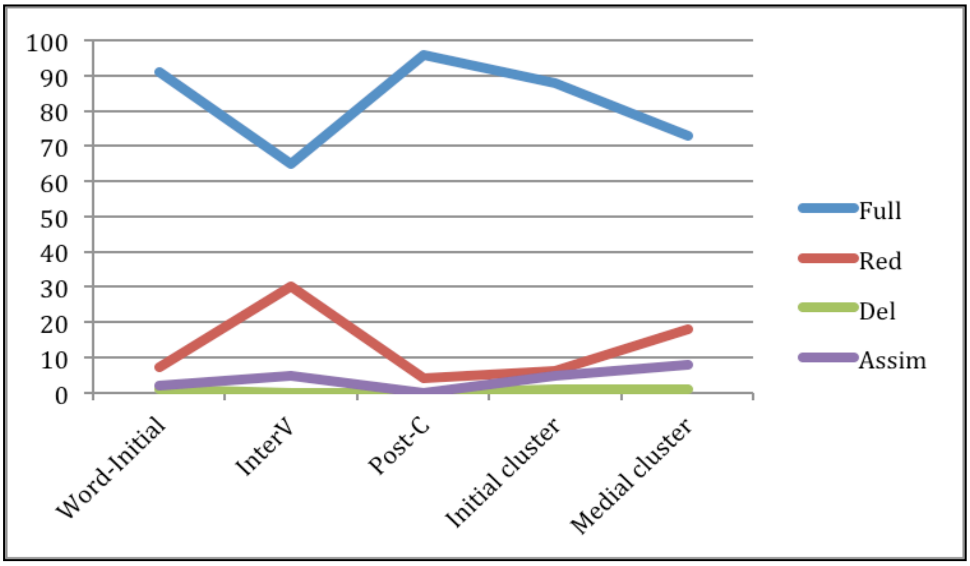


Fig. 25 Realizations of /p/ in each position

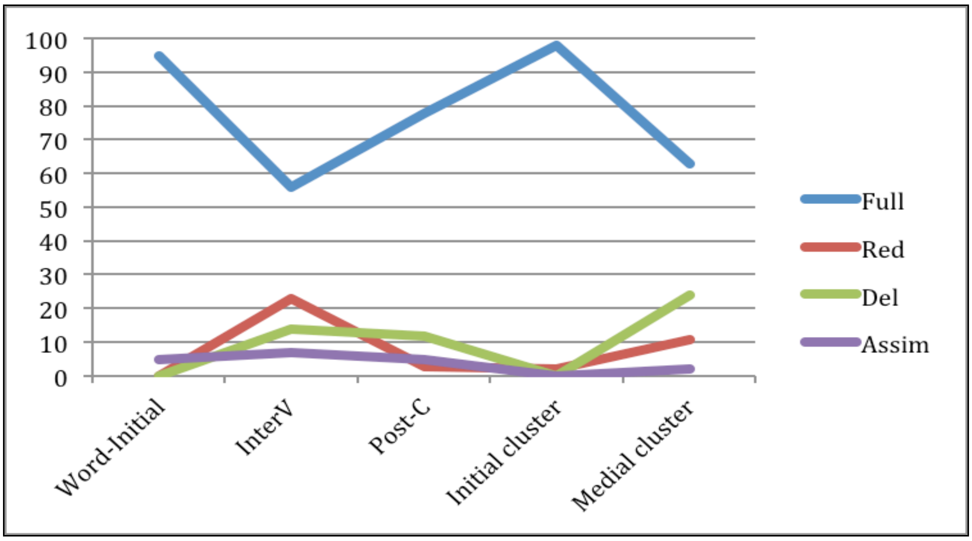


Fig. 26 Realizations of /t/ in each position

Table 33: Two-way ANOVA (Full realization of obstruents – without /ʃ/)

Analysis of variance	**Df**	**SumSq**	**MeanSq**	**F-value**	**Pr(>F)**
Obstruents	4	2143.3	535.83	1.9625	0.19347
Position	2	4491.2	2245.62	8.2247	0.01146 *
Residuals	8	2184.3	273.03		

Table 34: Two-way ANOVA (Reduction of obstruents – without /ʃ/)

Analysis of variance	**Df**	**SumSq**	**MeanSq**	**F-value**	**Pr(>F)**
Obstruents	4	550.82	137.70	2.0178	0.18474
Position	2	765.22	382.61	5.6065	0.03006 *
Residuals	8	545.95	68.24		

Regardless of the results presented in tables 33 and 34, it can still be argued that the difference between, at least, /s/ on the one hand and stops on the other hand is notable and probably a larger amount of data would yield more statistically significant results.

4.4 Summary

The casual speech phenomena identified in the four CLIPS dialogues under analysis prove that, (1), not all consonants are equally frequent in spoken Italian, and (2), not all consonants undergo reduction/deletion to a similar extent. As for sonorants, /r/ and /n/ are substantially more frequent than, respectively, /l/ and /m/. One could go so far as to say that, at least in spoken Italian, /r/ is the unmarked liquid and /n/ is the unmarked nasal, since both /l/ and /m/ are excluded from certain environments where /r/

and /n/ are allowed. For instance, OL clusters are typically composed of an obstruent followed by a rhotic, whereas clusters with a lateral are comparatively scarce. Similarly, since nasals in Italian always agree in place of articulation with the following consonant, and coronals are substantially more frequent than labials, the sequence /nT_{cor}/ is much more common than /mP/. The sequence [ŋK] appears to be more frequent than /mP/ as well, but the velar nasal is not a distinctive phoneme in Italian, therefore I deemed it to belong to the phoneme /n/. Crucially, there is an important difference between the two liquids and the two nasals: /r/, albeit being more frequent than /l/, is fully pronounced more often than /l/, and the two undergo reduction, deletion and assimilation to a similar degree. Conversely, the most frequent nasal, /n/, is reduced and assimilated more readily than /m/, which in turn is fully pronounced more often (but curiously, also deleted more regularly). Therefore, frequency[26] does not predict the reduction/deletion rate of sonorants, whereas salience might. As a matter of fact, /r/, which is more frequent and more salient than /l/, is also more resistant, and /m/, which is less frequent but more salient than /n/, is also reduced less often. As for sonorants, I would conclude that the results of the experiment presented in chapter 3 accord quite nicely with the realizations of liquids and nasals in casual Italian: the more salient a segment, the more likely it will be correctly identified in a cluster and the more likely it will be preserved (fully pronounced, unreduced, unassimilated) in spontaneous speech.

---

[26] Not even sonority is explanatory. As a matter of fact, assuming that /r/ is more sonorous than /l/ and that /n/ is more sonorous than /m/, /r/ and /n/ should undergo deletion and reduction to a similar extent, either more or less frequently than /l/ and /m/, respectively.

Moving the focus to the obstruent series, the typology of their realizations and their frequency rates indicate, similarly to sonorants, that frequency cannot be the only predictor for their pronunciation in running speech. Overall, stops are more frequent than fricatives, with the exception of /s/, which is more frequent than both /k/ and /p/. Fricatives are fully realized more often than stops, however, while tokens of /f/ and /ʃ/ are quite rare, /s/ is the second most frequent obstruent. As for stops, /t/ is more frequent than both /k/ and /p/, and the labial stop is slightly more common than the velar one. Of the three stops, /p/ is, at the same time, the most often fully pronounced and the most often reduced. The highly frequent /t/ undergoes deletion more systematically than the two others, whereas /k/ appears to be the most prone to assimilation. How to interpret these data in comparison with the results of the experiment, which suggested that /k/ and /t/ were effectively recognized more easily than /p/? First of all, it is not contradictory that /k/ is at the same time the most salient and the most easily assimilated of the three stops. The experimental findings of Hume et al. (1999) indicate that /k/, in isolation, has more robust acoustic cues than /p/ and /t/, but that its overall perceptibility and recognizability varies to a great extent depending on the following vowel. More specifically, participants in the experiment found that the dorsal transitions were quite distinct from the labial and coronal ones before /a/ and /u/, while before /i/ it was the labial that proved to be the most perceptually distinct (Hume et al. 1999:2072). The labial stop was hard to identify correctly in plateau clusters and in the dialogues it is very often both fully pronounced and reduced. Nevertheless, figure 40 clearly shows that all instances of reduced /p/ basically occur in the intervocalic context.

Elsewhere, /p/ is realized as [p]. Considering that crosslinguistically /p/ prefers to occupy the word-initial position (and also in the dialogues, see fig. 34), I suggest that labial transitions are particularly salient phrase-initially, whilst they become less salient intervocalically and lose much of their audibility in the coda and in combination with other obstruents, probably because labial stops are the ones which exhibit the shortest duration. The coronal stop is deemed to be the least salient by Hume et al. (1999), but my experimental results suggest that it is easier than /p/ and /f/ to perceive in plateau clusters. It is also the most frequent obstruent in the dialogues and the most frequent consonant in general after /r/. It undergoes reduction intervocalically relatively less often than the other two stops, suggesting that coronals might be "fitter" for that position. Its deletion rate is the highest among obstruents, a fact that could be explained both by its frequency and its predictability. We should be reminded that since /t/ is so frequent/unmarked, listeners expect to hear it much more than they expect to hear any other consonant. Vowels are deleted and reduced in casual speech more readily than consonants because they are also more easily reconstructed. Listeners know that a word cannot consist of only consonants (in Italian), and the perceptual absence of vowels is readily compensated. The same can apply to /t/: given its high occurrence rate, listeners expect it to be there, and when it is deleted or obscured by another articulatory gesture, they can easily reconstruct it. Moreover, it is noteworthy that the majority of the deletions of /t/ take place after /s/, which, given its salience and its spectral similarity to /t/, tends to acoustically mask the stop. The absolute rarity of the post-alveolar sibilant limits the possibility of interpreting much about its perceptual and cognitive salience. It is hard to determine whether it is never reduced or deleted because it is salient, because it is infrequent (and therefore, very informative) or because it is intrinsically a geminate. /f/ too proves to be quite resistant, since it generally escapes deletion, reduction and assimilation. However, it basically occurs only in two positions: word-initially and post-consonantally, i.e., the two strongest positions in a word. This tendency corresponds quite well to the predictions made for /p/, namely, that labials maintain their salience in the (non-intervocalic) onset and lose it elsewhere. /s/ is, simultaneously, the second most frequent obstruent in all the four dialogues and the third most frequent consonant (after /r/ and /t/), the most perceptually salient and the third least reduced/deleted consonant (after the other two fricatives). Again, the realizations of /s/ prove that frequency by itself cannot predict how often a consonant will be reduced or deleted. Given its frequency, /s/ might be expected to be deleted to a similar extent to /t/, but it is not the case. Its articulatory complexity would also predict a much higher rate of deletion. Instead, /s/ is at the same time particularly frequent and exceptionally resistant, considering that it also occurs in acoustic environments that normally favor reduction and deletion. Further taking into account the results of the perception experiment, it no longer seems controversial to say that the coronal sibilant is mostly preserved in casual speech because its perceptual salience plays an ultimate role in speech recognition. Finally, it appears that all consonants, in all positions, were fully realized more often than they were deleted, reduced or assimilated[27], a fact which

---

[27] There are the following exceptions: /l/ in the coda is fully pronounced as often as it is deleted, /n/ in the coda and /k/ intervocalically are reduced more often than they are fully pronounced, /s/ undergoes intervocalic voicing more often than it does not.

accords well with the syllable-timed nature of Italian. Unlike stress-timed languages (e.g., English, Dutch), Italian tends to exhibit less variation in vowel quality and less consonantal lenition. As nicely formalized by Schwartz (2010:6), in stress-timed languages more robust CV transitions interact with vowel diphthongization and consonantal lenition, whereas in syllable-timed languages, pureness in vowel quality interacts with less robust CV transitions and lack of significant lenition. Schwartz assumes that “robust CV transitions (…) allow listeners to reconstruct consonants provid[ing] a perceptual license for consonant lenition (…). [T]he consonant may be reconstructed on the basis of vowel formants” (Schwartz 2010:5). Since in Italian CV transitions are less robust than in English or in Dutch, both vowels and consonants are reduced to a lesser extent.

5

SALIENCE AND INFORMATIVENESS IN PHONOLOGY

In the previous chapters, the relationship between perceptibility and resistance to reduction, deletion and assimilation in spontaneous speech has been investigated. As a result, it appears that highly perceptually salient consonants, such as /s/, are both easily perceived and rarely deleted or reduced. However, the behavior of certain consonants is more difficult to explain, since it appears to be context-dependent, e.g., /k/ is highly salient in clusters and before non-front vowels but readily undergoes reduction, deletion and assimilation intervocalically and adjacent to front vowels. Finally, other segments, such as /t/, albeit not particularly salient from a perceptual point of view, are not only identified fairly easily, but also readily reconstructed when deleted or reduced. Speaking in OT terms, it can be said that, at least in Italian, faithfulness constraints protecting /s/ are ranked high, whereas those protecting /t/ are ranked low. Another possibility is that there are no faithfulness constraints specified for /t/, but that a general MAX-C constraint is responsible for its preservation, whereas more specific constraints are somehow informed of the relevance of /s/. Alternatively, faithfulness constraints might always be general, and the MAX-INVARIANT constraint proposed previously might be specified, each time, on a word-specific basis. There are good reasons to posit the existence of MAX-INVARIANT, since in a great number of cases reduction in spontaneous speech is word-specific and the same segments can be deleted and reduced in certain words and yet be fully pronounced in others. It is very probable that this difference depends on the lexical frequency of the word, but since it is not desirable for a phonological theory to include frequency in the grammar, I proposed that representations be uneven. Basically, the construction of uneven representations would work on similar lines to those of neural networks, where each time the pronunciation of a certain unit is encountered in speech, its representation is strengthened (cf. Boersma et al. 2012:5). The more the representation of a segment is strengthened in the lexical entry where it belongs, the higher the position in the ranking of the constraints protecting it. If a segment (or a subsegmental unit, such as an element) is present in basically all the pronunciation variants of a word, then it is likely to be included in the *invariant*, i.e., the "core" essence of a word, and MAX-INVARIANT will prevent that unit (be it segmental or subsegmental) from undergoing deletion. Nevertheless, there are several problems with positing the existence of such a constraint. First of all, the reasoning might appear to be circular. OT already contains the seeds of circularity, since it is an output-oriented theory and analyses carried out in this framework are often unfalsifiable. Some may argue that proposing the existence of MAX-INVARIANT and the fact that its position in the hierarchy is fixed and undominated does not explain anything, or rather, states the obvious: the invariant is never deleted, so there must be a constraint that forbids the deletion of the invariant. However, one of the main intentions of this thesis is precisely to escape such circularity. Collecting a large number of pronunciation variants of several words and consonants, I tried to determine what units tend to be preserved and what units are more easily disposed of. As a result, it turned out that at least one of the reasons why certain units are preserved more frequently than others (and are therefore more likely to form part of the invariant) is their perceptual salience. Highly salient units, being more readily perceived and identified, have higher chances of being stored and acquiring a strong representational status. On the other hand, salience is also relative, so MAX-INVARIANT must necessarily be, at least to some extent, word-specific. As long as a word contains stridents and stops, it is easy to predict that stops will be more likely than stridents to be deleted or reduced, but when no such great difference in acoustic salience is present in a word, other factors surely come into play. One of those factors could be predictability. /t/ is readily dismissed as it is readily reconstructed in speech by listeners. Put differently, /t/ is an "all-purpose segment", following Steriade's definition (2001:64). Crucially, I argue that salience is secondary to predictability. The logic behind this claim is very simple: if the preservation of the integrity of certain segments in speech depended exclusively on perceptual salience, vowels would be preserved more frequently than consonants, but in fact they undergo deletion and reduction to a considerably greater extent. Vowels are prototypically syllabic nuclei, and the nucleus is, universally, the head of the syllable. Since vowels are expected to be present, they can be dispensed with in pronunciation. Consonants are less predictable than vowels for two reasons: (1) unlike the nucleus, the onset and the coda are not compulsory in a syllable, and (2), consonants display a greater variety of places and manners of articulation. Phrased alternatively, consonants are more informative than vowels.

Among consonants, the most frequent, which in Italian (and possibly in many other languages) is probably /t/, can also be easily dismissed, since it is uninformative[28].

(71) Degree of informativeness
Vowels < /t/ < Other consonants

Tentatively, the invariant might consist of a combination of or a compromise between the acoustically salient and the highly informative units composing a word or a segment.

5.1 Phonological salience

This section will deal with the following issue: what is the phonological counterpart of acoustic salience? There are at least two possible answers: headedness, as employed in ET, and the element **A**.

5.1.1 Headedness

Headedness of an element normally translates into acoustic prominence (Carr 2005). For instance, labials and velars are characterized by similar spectral characteristics, e.g., energy at low frequencies, however this acoustic gravity is prominent in labials and less so in velars. Therefore, labials are said to be **U**-headed and velars are said to be headless, yet to contain **U** (Backley 2011). Unlike "dark" labials and velars, coronals and palatals are bright, and this shared brightness is represented by the element **I**, which characterizes both (Botma 2004). The purest realization of

---

[28] Cf. Hume (2011:98-101) for the role of informativeness in phonology.

I is energy at high frequencies, and its vocalic realization is the front vowel [i]. Therefore, palatals are **I**-headed and coronals, like velars, are headless, while still containing **I**. Headedness also seems to play a role in casual speech, as headed elements are less likely to undergo deletion but more likely to spread. Thus, the front vowel [i], which consists of headed **I**, readily palatalizes adjacent consonants, especially those that are headless. In general, headless velars are the most prone to palatalization, and this is largely reflected in the phenomena analyzed in the previous chapter. Similarly, the coronal stop [t] is the most likely to be deleted, which is consistent with its headlessness. Labials and palatals, which contain respectively **U** and **I**, tend to appear in strong positions in the word, i.e., word-initially and post-consonantally. The fricative /ʃ/, which is **I**-headed, is never deleted, while **U**-headed /p/ and /f/, albeit less resistant, disfavor weak positions, such as the intervocalic one. Of the two nasals under analysis, /m/ and /n/, the first contains **U**, which is preserved most of the time, whereas /n/, whose structure combines **A** and **I**, frequently loses its unheaded element or replaces it with another one. As for the liquids, I will deal with them in the next paragraph. In sum, headedness translates into the prominence of certain acoustic characteristics in a segment and also into its resistance to reduction, deletion and assimilation. Headless consonants are more likely to host new elements and lose theirs, headed consonants are more likely to spread their elements while conserving them, but being headed does not necessarily translate into perceptual salience. For instance, /k/ and /p/ are both characterized by the element **U**, but, at least in isolation, /k/, which is headless, appears to be more salient than /p/.

5.1.2 Is **A** *the salience element?*

The element **A** is one of the topics most discussed by phonologists working with ET (see Schane 2005, Pöchtrager 2006, Kaye & Pöchtrager 2009 and Pöchtrager 2012). In actual fact, **A** occurs in the internal structure of all the loudest segments: it characterizes [a] and other low vowels (which are described as the most sonorous segments), all sonorants (either as the head, as in coronal sonorants, or as the operator, as in labial, palatal and dorsal ones), sibilants (as the head in [s] and as the operator in [ʃ]), uvulars and pharyngals. The direct interpretation of **A** is [a] if headed and [ə] if unheaded (Backley 2011). Crucially, every language has presumably at least one low vowel and the schwa is the default epenthetic vowel in a number of languages. All these data seem to point to the fact that there is a connection between lowness in vowels, on the one hand, and sonorancy, nasality and stridency in consonants, on the other hand. Vowels that are **A**-headed are more sonorous than those which are not, and consonants containing **A** are more perceptually salient than those which do not. In theoretical frameworks like Dependency Phonology (DP henceforth; Anderson & Ewen 1987) and Radical CV Phonology (RCVP henceforth; van der Hulst 1994, 1995), the roles that are played by **A** and **L/N** in ET are subsumed by an element **V**, grossly corresponding to vocalicness, whereas **C** (consonantality) stands for total or partial occlusion (i.e., stopness and frication noise) and voicelessness (i.e., **H**). Szigetvári (1999:62) argues that the very nature of the V slot in the skeleton is loudness, as opposed to the nature of the C slot, which is quietness. This explains why when a vowel governs a consonant, the former makes the latter more vocalic and when a vowel licenses a consonant, it corroborates its nature of being "quiet" (typically, it confirms its "stopness", cf. Ségéral & Scheer 2001). In an attempt to unify all theories dealing with monovalent element-like units, I propose the following relationships between syllabic positions and phonetic expression:

(72)

Other things being equal,

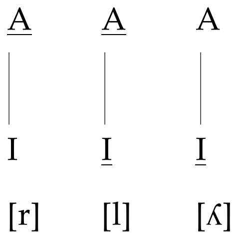
(a) a vocalic slot projects **A** on its melody tier.

(b) a consonantal slot projects **?** on its manner tier.

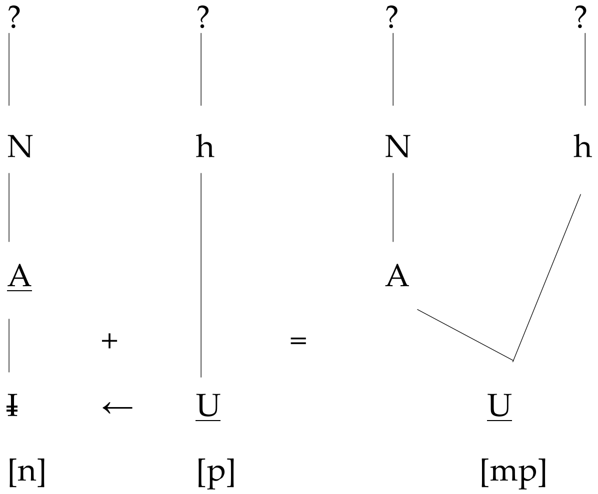
(c) **A** and **?** do not combine in the same phonological expression (Scheer 1999:218).

(72a) basically states that a vowel has to be loud/sonorous by default. (72b) states that a consonant should be the opposite of a vowel, namely, as little sonorous as possible, i.e., a stop. (72c) is the phonological formalization of obvious articulatory and perceptual factors: maximal aperture/sonority cannot co-exist with maximal closure/minimal sonority. (72) states that low vowels are less marked than non-low vowels, that sonorants (consonants containing **A** but not **?**) are more marked than stops (which contain **?** and do not contain **A**) and that the basic series of stops does not involve any **A**-stops, but only **U**-stops (i.e., /p, k/) and **I**-stops (i.e., /t, c/). Among sonorants, the greater salience of /r/ with respect to /l/ might be explained by the different roles played by **A** in the two segments. Although the two liquids share the same melody (i.e., **A-I**), there are good reasons to assume either that **I** plays a major role or that **A** plays a minor role in /l/, given its greater tendency to undergo palatalization. Finally, I argue that in nasals **A** does not indicate place but sonorancy, and as a matter of fact, place assimilation in nasals typically consists of the loss of the element **I**, which is replaced by **U̲** or **U**, as shown in (74).

(73) Melody of Italian liquids



(74) Nasal place assimilation



5.2 Unpacking Max-Invariant

The attempt of this paragraph is to unpack MAX-INVARIANT, namely, to determine whether MAX-INVARIANT is specified every time on a word-specific basis or if it is always possible to predict which units will form part of the invariant and therefore receive a sort of special protection.

5.2.1 Predictability of elements in C and V

One possibility is that MAX-INVARIANT is not affected by salience *per se* (otherwise vowels would be practically undeletable), but by salience *plus* informativeness, where by informativeness I mean the opposite of predictability. Generally speaking, melody is obviously loud and, of the three melodic elements, **A** is the loudest. Consonants, by default, should not be loud, therefore, loudness in consonants is informative and must be preserved.

(75)

Given the C and V slots, x *being the natural projection of C on its manner tier and* y *the natural projection of V on its melody tier, a phonological expression of the type: C–*y *is more informative than C–*x.

Expressed otherwise, C is not expected to contain loud elements, let alone the loudest element **A** – therefore, **A** in consonants tends to be preserved more often than in vowels, e.g., it is easier to delete /a/ than to delete /s/. Some may argue that sonorants undergo deletion and reduction quite often, but what sonorants normally lose in these processes are elements other than **A**. As discussed earlier, /r/ in many Germanic languages is reduced to a low vowel – it loses manner elements and **I** but maintains **A**. Nasals readily assimilate but do not lose sonorancy as easily. The fact that fricatives are more resistant than stops to deletion and reduction might also depend on informativeness. Consonants are expected to be as quiet as possible, i.e., to contain **?** on their manner tier. **h** is obviously less quiet than **?** and more informative in that position.

(76) Loudness/quietness scale of elements

A > I, U, N, R[29] **> h > ?**

Tentatively, the definition of the invariant can be reformulated as follows:

(77)

The invariant consists of the most informative and *the most perceptually salient elements composing a word.*

So far, I have only dealt with the unexpected presence of **A** in C. What about the opposite case, i.e., the presence of **?** in V? Data from acoustic reduction in Danish indicate that syllables that carry the stød (which can be arguably represented as an element **?** in the vocalic slot) undergo reduction to a substantially lesser extent than stødless syllables (Pharao 2009:130-131), suggesting that the presence of the occlusion element in vowels has quite a strong representational status. Moving back to the consonants under analysis in Italian, their deletion, reduction and assimilation rates accommodate nicely with their melodic structure: the only **A**-headed obstruent, /s/, is impressively resistant to deletion and reduction and occurs in every possible syllabic position (even in the nucleus when it is pronounced as syllabic). **U**-headed and **I**-headed consonants, /f, p, ʃ/ are almost never deleted, whereas /k, t/, which are headless, are the most likely to undergo deletion and assimilation. One

---

[29] It is certainly possible to make finer-grained distinction in loudness between **I, U, N** and **R** but it is not necessary for the current analysis.

way to decompose MAX-INVARIANT into a series of more specific constraints might be the following:

(78)

(a)

MAX-$\underline{A}$ IN C > MAX-$\underline{U}$, MAX-$\underline{I}$ IN C > MAX-A, I, U IN C

(b)

MAX-h IN C > MAX-? IN C

(78a) indicates that it is more important to preserve **A** in a consonantal position than any other elements, and that headed melodic elements are more important to preserve than unheaded ones. The ranking in (78b) expresses the fact that faithfulness to noise is ranked higher than faithfulness to stopness in consonants.

Crucially, informativeness may sometimes coincide with markedness but they are two separate concepts. The reasoning according to which faithfulness constraints tend to preserve informative material is closely reminiscent of De Lacy's *Preservation of the Marked* (2006). According to De Lacy, markedness constraints militate for the emergence of the unmarked, while faithfulness constraints preserve the marked. Nevertheless, I argue that markedness (as formalized by De Lacy) and the concept of informativeness proposed here make different predictions. For instance, the universal ranking *DORSAL > *LABIAL > *CORONAL >*GLOTTAL proposed by De Lacy implies that, among fricatives, /x/ is more marked than /f/, which is in turn more marked than /s/ and /h/. Informativeness would instead suggest that /s/ is more informative than both /f/ and /x/, because of its **A**-headedness. Similarly, while nasal consonants are undisputedly unmarked, nasal vowels are quite marked. However, since nasality is a **V**-element, and vowels are expected to host **V**-elements, nasal vowels are less informative than nasal consonants. Most importantly, faithfulness constraints based on informativeness stem from theory-internal conditions on the internal structure of segments, whereas faithfulness constraints preserving the marked are based on evidence coming from disparate sources, thus making informativeness more apt than markedness to be employed in a phonological theory.

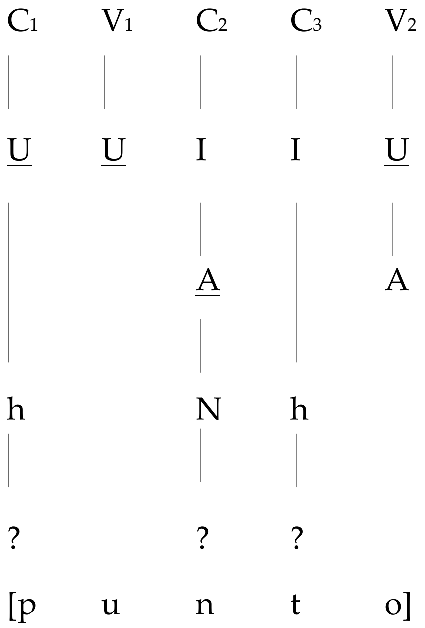
5.2.2 Predictability of structure vs. melody

In the previous paragraph I have proposed that faithfulness to melody in consonants is ranked higher than faithfulness to melody in vowels. I argue that presumably faithfulness to melody, in general, tends to be ranked higher than faithfulness to structure, where by structure I mean both the syllabic skeleton and the organization and distribution of melody on the skeleton. Let us consider the examples in (79-80).

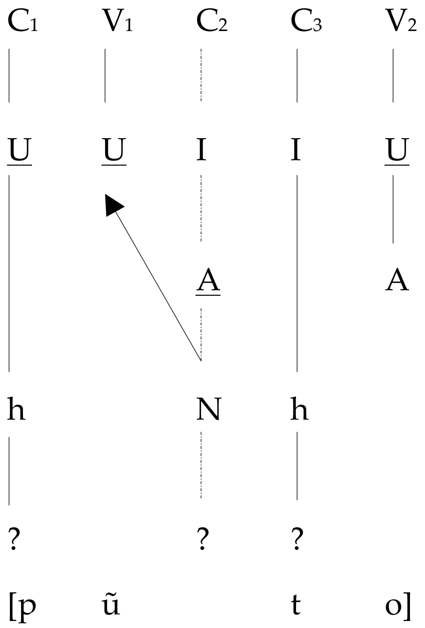
(79) Possible pronunciation variants of *punto* /pun.to/ ‘dot, point’:
[punto, pũto, pmto, ɸũNto…]

(80) Representations of possible distributions of the elements

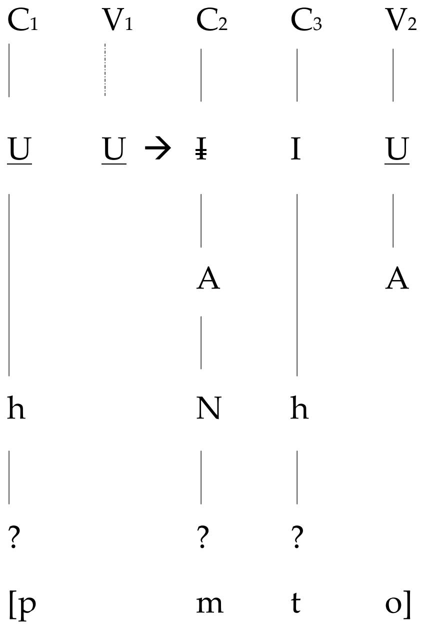
(a)



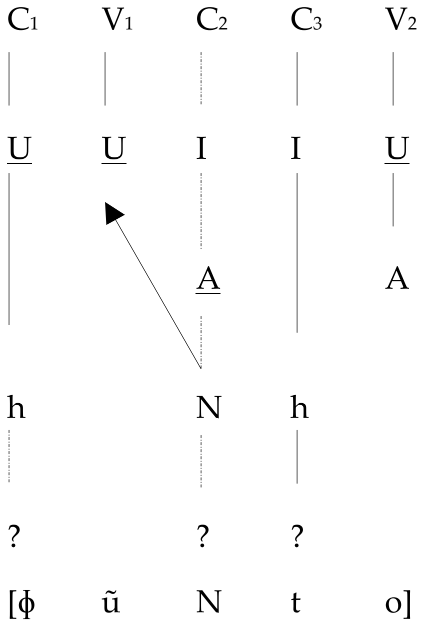
(b)



(c)



(d)



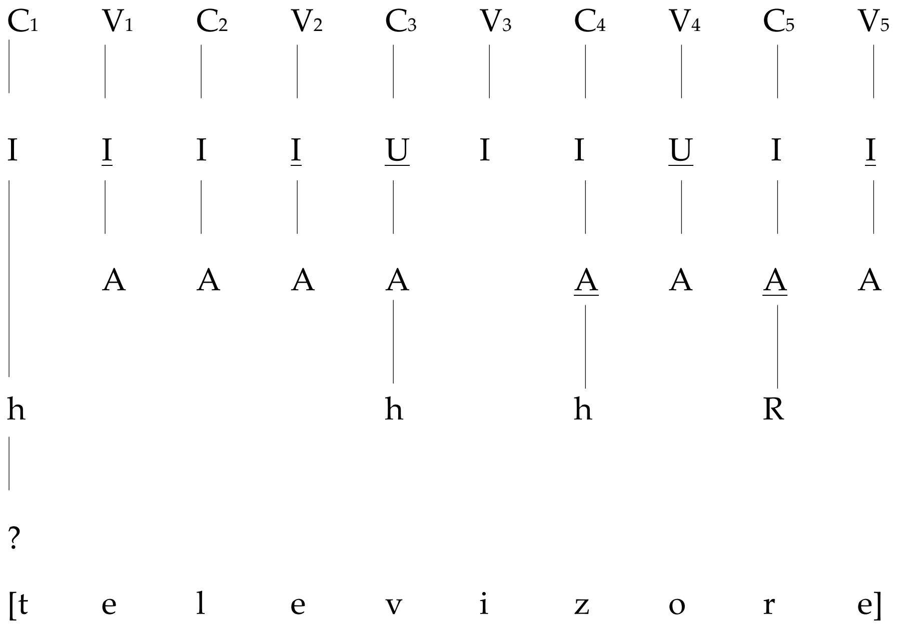
(80) shows that, among the potential realizations of /punto/, one of the main difference consists of whether the pre-consonantal nasal is

pronounced as an actual nasal stop or not. In two cases out of four, nasality is present but not the consonant itself. Put differently, the element **N** (or whatever element stands for nasality) is preserved in all the four tokens, whereas the phonological expression containing **A-I-N-?** in C_2 is not. (80a) represents the citation form, whereas in (80b) **A** and **I** are delinked from C_2 and **N** is linked to V_1. In (80c) it is **U** which is delinked from V_1 and associated to C_2, replacing **I**, and in (80d) **?** is delinked from C_1, while C_2 is delinked from **A** and **I** but not from **N**, which ends up being shared by C_2 and V_1. All of the phenomena described above are quite common in the spontaneous speech of all languages and basically consists in the re-distribution of the elements (mostly melodic ones) on the skeleton. Most importantly, what may appear at first sight as deletion, is in fact just a process of delinking followed by spreading. For instance, in (80c), it is more correct to say that the element **U** moves from a vocalic slot to a consonantal one, rather than saying that the vowel /u/ undergoes deletion. As a matter of fact, in the phonetic form [pmto], V_2 is missing, but its melodic content is preserved, albeit borne by another segment, i.e., [m]. Another example is given in (81-82).

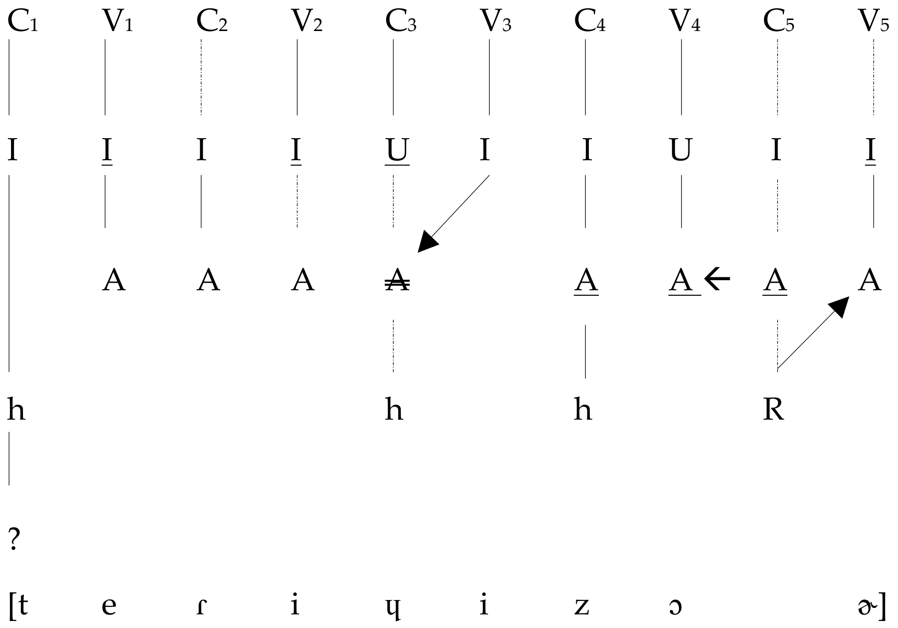
(81) Possible pronunciation variants of *televisore* /te.le.vi.zo.re/ ‘television’.
[televizore, teɾiɥizɔɚ, təlyvzɔwe]

(82) Representations of possible distributions of the elements

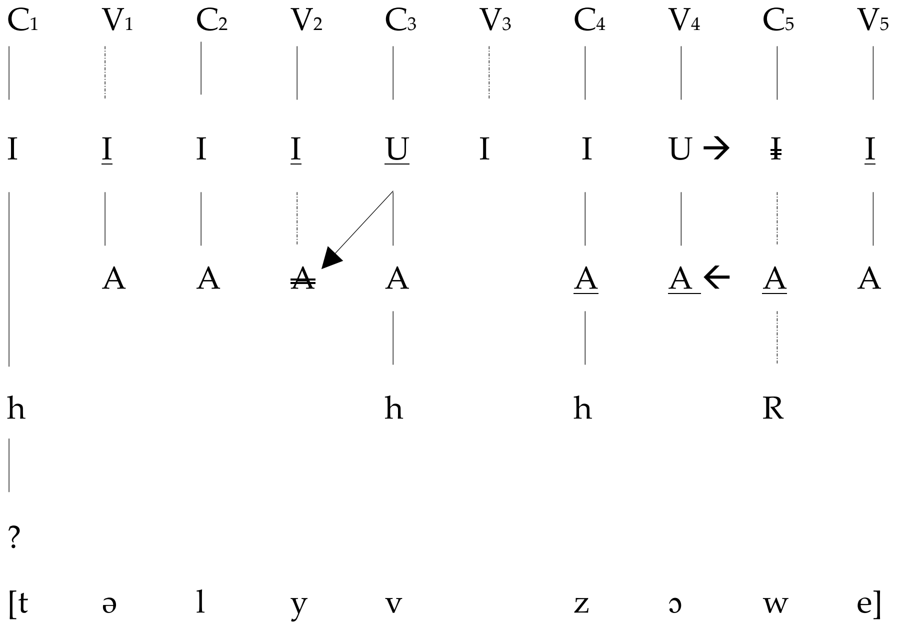
(a)



(b)



(c)



The phenomena represented in (82) are slightly more complex than the ones in (80). The citation form [televizore] is shown in (82a). In (82b), /l/ is lenited to [ɾ], which I interpret as the realization of unheaded **A** in C. I therefore represent it as the delinking of **I** from C2. /v/ is pronounced as [ɥ], implying both reduction and assimilation. Both **A** and **h** are delinked from C3 while **I** is shared between C3 and V3. The element **A̲** present in C5 spreads to the preceding **A** linked to V4, making it the head of V4 and turning **U** into the operator (therefore /o/ is lowered to [ɔ]). C5 is delinked from its melodic material and its element **R**, standing for rhoticity, spreads to V5, turning it into a rhotic vowel, which – however – is realized as schwa due to **I**-delinking. In (82c) C3 spreads its **U** element to the preceding vowel, displacing **A** and obtaining [y] as a result. V3 is delinked from **I** and does not receive phonetic interpretation, V4 is realized as [ɔ] for the same reasons as in (77b), and C_5 is delinked from all its elements but shares the element **U** with V_4, resulting in [w].

Just from looking at these two examples, it becomes apparent that both in *punto* and in *televisore* **A** is delinked from a C position. In (80b) and (80d), the structure of /n/ is reduced to **N**, while in (82b) and (82c) /v/ is reduced to a glide (losing both **A** and **h**) and **A** is delinked from the rhotic. These phenomena might seem problematic, since I assumed that **A** must be preserved when occurring in C. However, as stated in 5.1.2, there certainly is a connection between the element **A**, nasality and sonorancy – we should be reminded that by some authors (e.g., van der Hulst 1994) they are represented by the same element – therefore both **N** and **R** – identifying, respectively, the nasal and the rhotic – might simply be interpretations of **A**. As for /v/, I represented it as containing **A** in order to distinguish labial from labiodental fricatives, assuming that the former contain **$\underline{\text{U}}$-h** and the latter **$\underline{\text{U}}$-A-h**. Nevertheless, since Italian does not distinguish between these two categories[30] and labiodentals do not exhibit any phonological behavior proving that they contain **A**, they can be represented without the aperture element.

(83) Italian obstruents

/p, b/:	$\underline{\text{U}}$-h-?	/f, v/:	$\underline{\text{U}}$-h
/t,d/:	I-h-?		
/ts, dz/:	$\underline{\text{A}}$-I-h-?	/s/:	$\underline{\text{A}}$-I-h
/tʃ, dʒ/:	A-$\underline{\text{I}}$-h-?	/ʃ/:	A-$\underline{\text{I}}$-h
/k, g/:	U-h-?		

---

[30] Neither do most of the world’s languages. Ewe opposes /f, v/ and /ɸ, β/, but the former are pronounced with the upper lip noticeably raised.

(83) shows the internal structure of Italian obstruents, with labiodentals being melodically characterized exclusively by **U** and not by **A**. As for sibilants, it is still necessary to include **A**, at least in the structure of /s/, because /s/ patterns with sonorants in that they are the only consonants allowed to occur in the coda (without having to be the first half of a geminate).

Another striking characteristics of casual speech phenomena is that the deletion of a segment does not always correspond to the deletion of its internal structure. Most often, elements (mostly the melodic ones) simply occur in positions other from those in which they occur in the citation form. This fact suggests that melody and syllabic structure behave differently, which would not be surprising, since acquisition of prosody and acquisition of melody are quite separate domains (McMahon 2005). As a matter of fact, "[i]n the case of prosody (…), it appears that some specific-to-language innate component is supported by the small range of attested systems (…). For melody, however, the picture is very different. The range of possibilities, both in terms of allophonic variation and conditioning factors (…) is much broader" (McMahon 2005:269). Basically, since syllabic structure belongs to prosody, and prosody appears to be more likely than melody to form part of UG, being both evolutionarily older and acquired earlier by children, syllabic structure also proves to be less informative, more predictable, and therefore, less in need of being preserved intact in speech. On the contrary, it is evident that elements tend to be preserved to a considerably greater extent. If faithfulness constraints preserving more informative material are ranked higher than the ones preserving more predictable material, faithfulness to melody presumably dominates faithfulness to the distribution of melody on the syllabic skeleton, because the latter can be reconstructed by listeners resorting to their innate knowledge about prosody. In other words, listeners hearing [pmto] are able to extract the element **U** from [m] and assign it to the preceding vocalic slot, thus reconstructing /punto/, whereas a hypothetical form such as *[ʔəNto], where melody has been deleted from the first three segments, would be much harder to reconstruct.

5.3 An ET-OT analysis of casual Italian

In the first chapter of this thesis I discussed the various proposals that have been put forward in order to account for variation in OT. Those who have tried to model phonological variation within this framework, have all resorted, to a certain extent, to constraint reranking. To my knowledge, van Oostendorp (1997) is the only one who limits the arbitrariness of constraint reranking by positing that the re-ranking can only take place between faithfulness constraints and markedness constraints, whereas the ranking of a constraint with respect to another of the same family is fixed. In (84) I present an example, where F = faithfulness and M = markedness.

(84)
Careful speech ranking: F1 > F2 > F3 > M1 > M2 > M3
Allegro speech ranking: F1 > F2 > M1 > M2 > F3 > M3
Casual speech ranking: F1 > M1 > M2 > M3 > F2 > F3

It is evident from (84) that the relative ranking of faithfulness constraints, F1 > F2 > F3, and the relative ranking of markedness constraints, M1 > M2 > M3, remains untouched. However, the less careful the speech rate, the higher markedness constraints are ranked. Since I am dealing with the relationship between the SF and the AF/ArtF here, markedness constraints probably include (or coincide with) sensorimotor constraints, but given their similar role, I will not distinguish between the two constraint families.

I will now apply this model to the variation encountered in the dialogues under analysis in the pronunciation of *sinistra* 'left', *punto* 'point, dot' and *televisore* 'television'.

5.3.1 sinistra

In (25) a representation of the invariant of *sinistra* has already been proposed. Namely, it consists of the two stridents, the nasality element and the rhoticity element.

(85) Invariant of *sinistra*
/s N s R/

In (85) /s/ stands for the elements **$\underline{\text{A}}$-I-h** in a C slot, whereas for **N** and **R** the syllabic position is not specified. In 5.2.2 I have suggested that all the elements that somehow convey sonorancy are in fact interpretations of **A**. In order to formulate the faithfulness constraints that I need for my analysis of *sinistra*, I will group the elements **A**, **N**, **R** under the label **V**-

elements and the elements **?** and **h** under the label **C**-elements. MAX-INVARIANT can therefore be unpacked as follows:

(86)

MAX-V-IN-C	Do not delete sonorancy elements (**A**=**N**=**R**) associated to a consonantal slot.
MAX-I&V-IN-C	Do not delete **I** when associated to **V**-elements in a consonantal slot.

The constraints presented in (86) are undominated in the hierarchy. As for markedness constraints, I propose to consider the ones presented in (87).

(87)

*MELODY-IN-V	**A**, **I**, **U** are disallowed in a vocalic position.
SPREAD-N-TO-V	Vowels adjacent to nasal consonants must share **N** with them.
SPREAD-I̲-TO-C	Consonants adjacent to palatal vowels must share **I̲** with them.

SPREAD-N and SPREAD-I̲ are classical constraints required to explain assimilation, whereas *MELODY-IN-V militates against redundant/predictable material and is likely to be ranked low at careful speech rates and to be promoted only in very casual speech. We also need generic faithfulness constraints, such as the ones in (88).

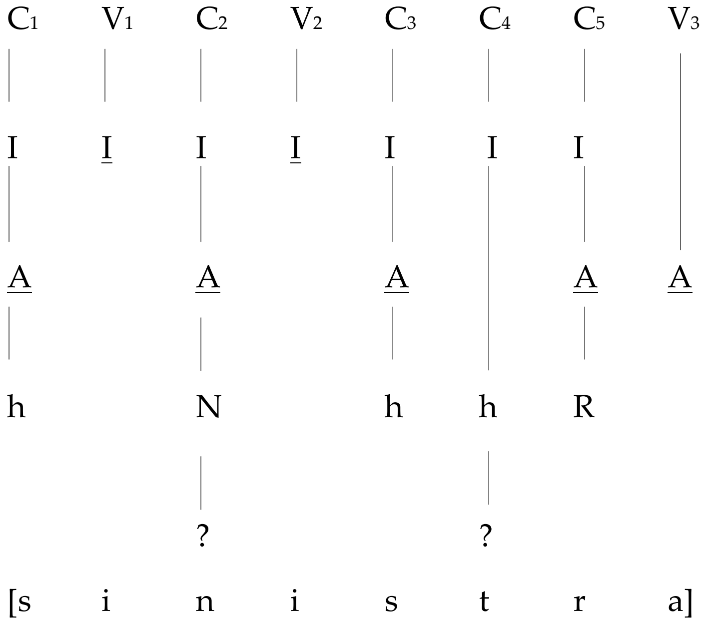
(88)

MAX-C	Do not delete consonantal positions and the elements associated to them.
MAX-I	Do not delete the element **I**.

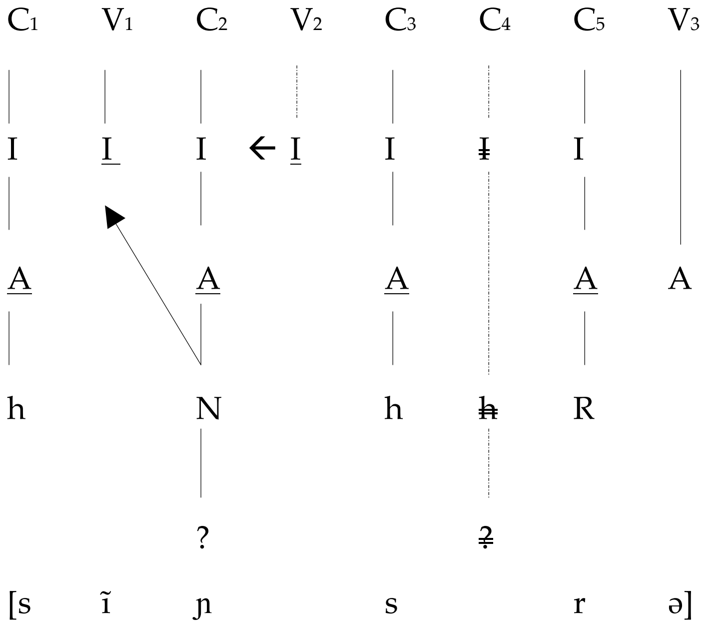
I selected three representative pronunciation variants of *sinistra*: careful speech [sinistra], moderate speech [sĩɲsrə] and casual speech [sNsr]. All of these pronunciations are actually attested in the CLIPS corpus. (89a-c) show their surface representations.

(89)

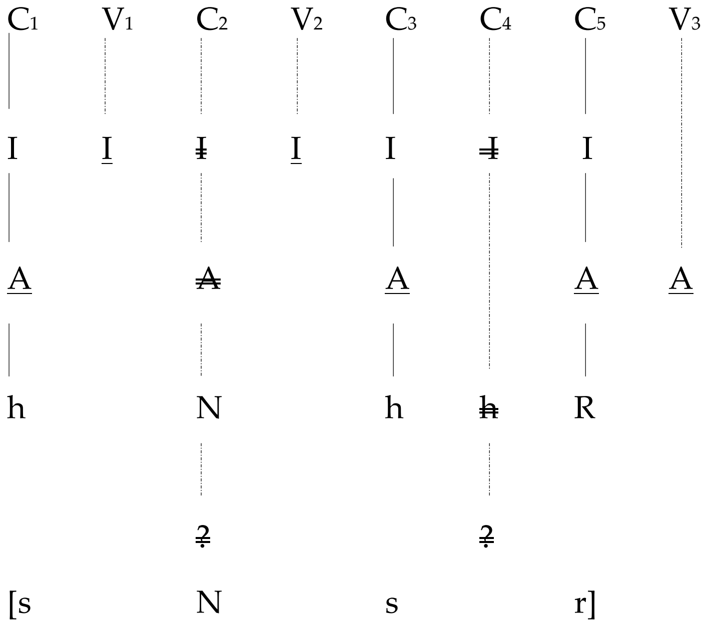
(a)



(b)



(c)



(89a) represents the citation form. In (89b) the headed **I** element is delinked from V2 and spread to C2, where it maintains its headedness, while **N** is shared by V_1 and C_2. C_4 is delinked from all its elements and in V_3 the element **A** loses its headedness. In (89c) all the vowels are delinked from their elements, as well as C_4, while C_2 loses its melody and its occlusion and only keeps **N**.

Tableau 11: *sinistra*, careful speech

/sinistra/ **s N s R**	Max-VinC, Max-I&VinC	Max-I	Max-C	Spread-NtoV	Spread-ItoC	*Mel-in-V
☞a) [sinistra]				**	***	***
b) [sĩɲsrə]			*!		*	***
c) [sNsr]		*!**	*			
d) [inita]	*!*****	***	***	**	**	***

In careful speech, all the relevant faithfulness constraints dominate markedness constraints. A hypothetical candidate such as (d) is ruled out under any ranking because Max-VinC, Max-I&VinC are undominated – they are a reformulation of Max-Invariant. [inita] would imply the deletion of the stridents and of the rhotic, i.e., loss of **V**-elements (**R** for the rhotic, **A** for the stridents) and of the combination of **V**-elements with **I** in a consonantal slot (i.e., the melody of stridents). Candidate (b) fails because of its violation of Max-C (/t/ is deleted), while the omission of **I** disqualifies candidate (c) as a possible winner. Candidate (a) is the winning candidate, being the most faithful one.

Tableau 12: *sinistra,* moderate speech

/sinistra/ **s N s R**	MAX-VINC, MAX-I&VINC	MAX-I	SPREAD-NTOV	SPREAD-ITOC	MAX-C	*MEL-IN-V
a) [sinistra]			*!*	***		***
☞b) [sĩɲsrə]				*	*	***
c) [sNsr]		*!**			*	
d) [inita]	*!*****	***	**	**	***	***

In Tableau 12 two markedness constraints, SPREAD-NTOV and SPREAD-ITOC, are promoted and dominate MAX-C. Therefore, candidate (a) [sinistra] is now ruled out because it violates SPREAD-NTOV (the vowels adjacent to the nasal consonant are not nasal as well), whereas candidate (c) is still out of the game because it deletes the element **I**. Candidate (b) wins the competition, allowing both **N**-spreading and **I**-spreading.

Tableau 13: *sinistra*, casual speech

/sinistra/ **s N s R**	Max-VinC, Max I&V inC	Spread NtoV	Spread I̲toC	*Mel inV	Max-I	Max-C inC	Max-C
a) [sinistra]		*!*	***	***			
b) [sĩɲsrə]			*!	***		**	*
☞c) [sNsr]					***	***	*
d) inita	*!*****	**	**	***	***	**	***

Under this ranking, markedness constraints dominate all faithfulness constraints except Max-Invariant (= Max-VinC, Max-I&VinC). Since Max-I is demoted, the winning candidate becomes (c), since it does not violate any of the relevant markedness constraints. It appears then that [sNsr] is less marked than [sinistra]. How can a phonetic sequence with no actual vowels be less marked than a sequence of well-formed syllables? The problem can be solved by assuming that it is not the alternation between consonants and vowels that is unmarked in speech, but rather the alternation between segments containing **C**-elements (such as the two stridents) and segments containing **V**-elements (the placeless nasal).

5.3.2 punto

The representation of the three pronunciation variants of *punto* under analysis have already been presented in (80a-c). I will not deal with the pronunciation variant shown in (80d) because it does not add much to the picture. Importantly, *punto* does not contain stridents and its invariant is composed of the element **U̲** which must be borne by the initial consonantal position and optionally by the following vowel and consonant, an element **N** and the segment /t/ (**I-h-ʔ**), followed by a vowel (whose melody ranges from **A-U̲** to **A**). The undominated constraint is still MAX-VINC, since the sonorancy represented by nasality cannot be dispensed with, while the preservation of **U̲** and /t/ can be stated using other constraints.

(90)

MAX-O, R_	Do not delete obstruents occurring after a sonorant.
ONSETC[PLACE]	Onset consonants must be specified for place (i.e., contain melody).

In the dialogues under analysis, there are 219 tokens of [nt] sequences, and [t] is deleted only in one token, suggesting that the preservation of the stop does not have to be encoded in the invariant, as it were specific to *punto*, but is presumably a more general process.

(91)

/Vnt/: 75/219 [Ṽt], 4/219 [Nt], 2/219 [Vt̰], 2/219 [t], 28/219 [n̩t], 2/219 [nd], 1/219 [∅], 1/219 [Nʔ], 1/219 [ʔṼ], 1/219 [d], 11/219 [ṼN], 10/219 [mt], 81/219 [nt].

There is in fact an absolute salience and a contextual salience and the role of the latter in the phonological grammar is nicely formalized by Steriade's P-map (2001). The P-map provides a series of universal constraint rankings grounded in the fact that certain contrasts are better audible in certain positions, e.g., the contrast in voice between two stops is better heard in intervocalic position. Similarly, it is well known that obstruents are clearly audible after a sonorant.

(92)

The invariant of a word consists of its most informative elements plus its most salient elements, where salience includes both absolute salience and contextual salience.

/t/ in *punto* is not salient *per se*, but forms part of all the pronunciation variants of the word because it is found in a strong position. Conversely, the second /s/ of *sinistra* is salient *per se*, because it is both in the coda (which is a weak position) and adjacent to segments that do not differ significantly from it (both the preceding vowel and following consonant contain **I**, which characterizes /s/ as well).

Other constraints that need to be taken into account for *punto* are presented in (93).

(93)

MAX-V	Do not delete vowels.
MAX	Do not delete segments (associations between elements and syllabic slots).

Spread-UtoC Consonants adjacent to vowels containing **U** must share **U** with them.

Tableau 14: *punto*, careful speech

/punto/	Max-VinC	Max-O,R_	OnsetC [place]	Max-V	Max	Spread-UtoC	Spread-NtoV
☞a) [punto]						*	*
b) [pũto]					*!	*	
c) [pmto]				*!	*		
d) [puto]	*!*				*		
e) [puno]		*!			*	*	*
f) [ʔunto]			*!			**	*

Under the careful speech ranking, faithfulness dominates markedness. Candidate (b) fatally violates Max, since nasality is not expressed by a consonant anymore but is carried out by a vowel. Candidate (c) is ruled out because of its violation of Max-V. Note, however, that neither in (b) nor in (c) are the elements **N** and **U** deleted. Candidates (d, e, f) are not attested. Candidate (d) omits nasalization, which is a violation of the undominated constraint (assuming that nasalization is one of the possible expressions of sonorancy in consonantal position). Candidates (e) and (f) violate constraints that, in a way, conspire to protect the invariant, since they are presumably universally high-ranked. In Steriade's terms, [puto], [puno] and [ʔunto] are more perceptually distant from [punto] than [pũto] and [pmto].

Tableau 15: *punto,* moderate speech

/punto/	Max-VinC	Max-O,R_	OnsetC [place]	Max-V	Spread-UtoC	Spread-NtoV	Max
a) [punto]					*	*!	
☞b) [pũto]					*		*
c) [pmto]				*!			*
d) [puto]	*!*						*
e) [puno]		*!			*	*	*
f) [ʔunto]			*!		**	*	

In Tableau 15 the two markedness constraints dominate Max, thus allowing candidate (b) to become the winner. Both candidate (a) and candidate (b) violate Spread-UtoC once, but candidate (a) also violates Spread-NtoV.

Tableau 16: *punto*, casual speech

/punto/	Max-VinC	Max-O,R_	OnsetC [place]	Spread-UtoC	Spread-NtoV	Max-V	Max
a) [punto]				*!	*		
b) [pũto]				*!			*
☞c) [pmto]						*	*
d) [puto]	*!*						*
e) [puno]		*!		*	*		*
f) [ʔunto]			*!	**	*		

In casual speech, Max-V is demoted below markedness constraints. Under this ranking, candidate (c) wins the competition, since it does not violate any of the top-ranked faithfulness constraints and neither of the two markedness constraints. In fact, in [pmto] **U** has been spread to the nasal consonant and there is no vowel left to nasalize.

5.3.3 televisore

The three pronunciation variants of *televisore* presented in (82a-c) are [televizore, teɾiɥizɔɚ, təlyvzɔwe]. The second token, [teɾiɥizɔɚ] shows typical lenition phenomena, such as the flapping of the lateral and the gliding of the fricative, as well as the fusion between the final vowel and the rhotic. I analyze the weakening of /l/ to [ɾ] as the loss of the **I** element, which is borne by the following vowel, which in turn loses its element **A**. The gliding of the fricative is instead explained by the loss of **h** and the acquisition of an **I** element, shared with the following vowel. Both phenomena occur in intervocalic position, which is the locus of lenition *par excellence*. The behavior of /l/ and /v/ can be seen as the effect of the following constraints:

(94)

*ComplexMelody V_V	A consonant cannot host more than one melodic element in intervocalic position.
*h,? V_V	A consonant cannot host **C**-elements in intervocalic position.

In [təlyvzɔwe], one vowel is reduced to schwa and another is deleted, which is normal for unstressed vowels. The second vowel loses **A** and receives **U** from the following consonant, while the rhotic is reduced to a glide, which undergoes **U**-coloring because of the preceding vowel. While the reduction of liquids to glides is quite a widespread phenomenon crosslinguistically, it is problematic for my theory since I assumed rhoticity to be one of the expressions of sonorancy in consonantal position and therefore, highly informative and undeletable. Nevertheless, if /s/ is indeed almost untouchable, sonorants are not. I therefore propose to distinguish between the mere occurrence of **V**-elements in a C position, which is quite informative, and the co-occurrence of **V**- and **C**-elements in a C position – that is, the co-occurrence of sonorancy and noise/stopness – which is extremely informative.

(95)

Max-V&C-in-C Do not delete **V**-elements when occurring in the same phonological expression with **C**-elements in a C position.

(96)

Max-V&C-in-C > Max-V-in-C

The ranking in (96) states that faithfulness to the (unexpected) presence of both **V**-elements and **C**-elements in a consonant is ranked higher than faithfulness to the presence of **V**-elements in a consonant. In other words, faithfulness to stridents and nasals is ranked higher than faithfulness to liquids, since the latter do not contain **C**-elements (**h**, **?**)[31].

Tableau 17: *televisore,* careful speech

/televizore/	Max-V &CinC	Max VinC	Max-C	*?,h V_V	*Complex Melody V_V
☞a) [televizore]				**	***
b) [teɾiɥizɔɚ]			*!	*	**
c) [təlyvzɔwe]		*!			*
d) [televiore]	*!	*	*	*	**

---

[31] As stated earlier, it might be a language-specific matter as to whether nasals and liquids contain the occlusion element. In Italian liquids are likely not to be phonologically specified for occlusion, since they typically form clusters with stops, whereas nasals, which do not, might instead contain **?**. However, it might also be the case that /l/ contains occlusion while /r/ does not, cf. (100).

In the careful speech variant of *televisore*, faithfulness dominates markedness. Candidate (b) fatally violates MAX-C, since intervocalic /r/ is missing (rhoticity is borne by the schwa), while candidate (c) is ruled out because /r/ is replace by [w], meaning that **V**-elements (**A**, **R** or both) are deleted from a C position, thus violating MAX-VINC. Candidate (d) cannot win under any ranking since it violates the highest ranked constraint, MAX-V&CINC, given that /s/ (which contains **V**-elements associated with a **C**-element) is missing. The winning candidate is therefore (a), which only violates the lowest ranked markedness constraints.

Tableau 18: *televisore,* moderate speech

/televizore/	MAX-V &CINC	MAX VINC	*?,h V_V	*COMPLEX MELODY V_V	MAX-C
a) [televizore]			**!	***	
☞b) [teɾiɥizɔɚ]			*	**	*
c) [təlyvzɔwe]		*!		*	
d) [televiore]	*!	*	*	**	*

In tableau 18 the markedness constraints *?,hV_V and *COMPLEXMELODYV_V dominate MAX-C. Under this ranking, (b) is the winner. Candidate (a) violates *?,h V_V twice, while (b) does so only once, and candidate (c) violates MAX-V&CINC, which is still ranked higher than markedness constraints. Candidate (b) also displays the rhoticization of the final schwa, which is likely to be triggered by a constraint of the type SPREAD-R ("spread rhoticity") but since such a constraint plays no role in the selection against the other relevant candidates, it was not included in the tableau.

Tableau 19: *televisore,* casual speech

/televizore/	MAX-V &CINC	*?,h V_V	*COMPLEXMELODY V_V	MAX VINC	MAX-C
a) [televizore]		**!	***		
b) [teɾiɥizɔɚ]		*!	**		*
☞c) [təlyvzɔwe]			*	*	
d) [televiore]	*!	*	**	*	*

Under the ranking in tableau 19, candidate (c) is the winner. I considered (c) to be more casual than (b) because it displays a [vz] sequence which would not be acceptable according to Italian phonotactic rules. Nevertheless, it emerges as the winning candidate because *?,h V_V and *COMPLEXMELODYV_V dominate both MAX-VINC and MAX-C. Only *?,h V_V here is crucial for the evaluation, but *COMPLEXMELODYV_V is maintained in the tableau to explain the lenition of /l/. Both candidates (a) and (b) violate *?,h V_V and are consequently ruled out. Candidate (c) violates *COMPLEXMELODYV_V only once (/l/ occurs between two vowels and contains two melodic elements, **A** and **I**) since the sibilant does not appear in the intervocalic position, and is instead preceded by a consonant, and /r/, which has the same melody as /l/, is replaced by [w].

5.4 The interpretation of V and C

In order to capture the fact that certain elements pattern together because they are more likely to occur in a vocalic slot whilst others prefer to belong to a consonantal slot, I chose to employ the label **V**-elements and **C**-elements. My stance is not original, since there are at least two phonological theories which have massively reduced the number of elements conflating several primes into the **C** and **V** labels. First of all, DP (Anderson & Ewen 1987) describe all types of segments as consisting of |C|, |V| or a combination of the two. |C| and |V can enter in a relationship of dependence, in which either one of the two is the head or they are mutually dependent. In their notation, X;Y means "Y is dependent on X" and X:Y "X and Y are mutually dependent". Their representation of segment types is presented in table 35.

Importantly, the sonority scale from |C| to |V| is not analyzed as a decrease or increase in complexity, since intermediate steps are more complex than both ends of the scale. Instead, the role of |V| becomes more and more important from the second step of the scale onwards. In a voiceless stop, |V| is completely absent. Then, it appears in voiced stops as dependent on |C|, in fricatives as mutually dependent, in nasals and liquids as the head and in vowels it is |C| which is absent. Comparing DP to ET, it appears quite clear that the former better captures the relationship between, e.g., stops and fricatives than the latter. In ET, **ʔ** and **h** are simply two different elements, which happen to occur together in all stops but that are fundamentally two separate entities. Instead, in DP, the difference between stops and fricatives is that the latter are slightly more vocalic than the former and this increase in vocalicness proceeds towards the end of the sonority scale. However, the apparent economy of DP is lost when locational elements are introduced and they are far more numerous than those of ET: besides |a, i, u|, DP employs |@| "centrality", |T| "advanced tongue root", |l| "linguality", |t| "apicality", |d| "dentality", |r| "retracted tongue root", |L| "laterality". As noted by van der Hulst (1994:446) "[t]he DP proposals for locational properties are somewhat arbitrary. The heart of the system is formed by the three elements |a|, |i| and |u|, but when the discussion goes beyond fairly simple vowels and consonant systems the number of elements is rapidly expanded"

Table 35: DP elements and their interpretation

Element	**Interpretation**	**Comment**
|C|	Voiceless stop	The element |C|, by itself, is interpreted as the basic consonant, i.e., a voiceless stop.
|C;V|	Voiced stop	|C| is the head, |V| is dependent on |C|. In this case |C| is prominent and is interpreted as a stop, while |V| as a dependent stands for voicing.
|V:C|	Voiceless fricative	|C| and |V| are mutually dependent. |C| stands for obstruency, while |V| translates into continuancy.
|V:C;V|	Voiced fricative	|V| is dependent on |V| and |C|, which are mutually dependent.

|V;C|	Nasal stop	|C| is dependent on |V|. Headed |V| translates into sonorancy, while |C| conveys stopness.
|V;V:C|	Liquid	|V| and |C| are mutually dependent and both are dependent on |V|. The role of |V| here is greater than in nasals and |V:C| translates into continuancy.
|V|	Vowel	Quite obviously, the element |V| by itself is interpreted as a vowel.

Because of this unnecessary proliferation of elements, van der Hulst feels the necessity to re-elaborate the theory, under the label of Radical CV Phonology (RCVP). In his view, a phonological expression may be either C-headed or V-headed. C and V can dominate another element, but they can never dominate an element identical to themselves, i.e., C cannot be dependent on C and V cannot be dependent on V. There are two types of dependency relations: sister dependency and daughter dependency. The former is represented in (97) and the latter in (98).

(97) *Sister-dependency*

|C|	C
|C;V|	C_v
|V;C|	V_c
|V|	V

(98) *Daughter dependency*

(a) C-headed

C C C_v C_v

| | | |

V V_c V V_c

(b) V-headed

V_c V_c V V

| | | |

C C_v C C_v

The system proposed in van der Hulst (1994, 1995) is very complex, especially when it comes to locational gestures (grossly corresponding to place elements and to height and backness features). As for the categorial gesture (manner), he proposes the following correspondences (van der Hulst 1995:97):

(99)

	Tone	*Stricture*	*Phonation*
C	high tone	stop	constricted glottis
C_v	low tone	continuant	spread glottis
V_c	high register	sonorant	nasal voice
V	low register	vowel	oral voice

Leaving aside the intricacies of DP and RCVP, I will now propose a new theory of representation of the internal structure of segments, with repercussions for the formulation of faithfulness constraints in OT.

5.4.1 Level-1 and level-2 elements

First of all, I will argue for the necessity of at least two levels of representations of elements. The first one is very abstract and may be placed between the UF and the SF. Here segments are represented as composed of solely C and V units, either in isolation or combined. C and V are arranged on three tiers: the skeleton, the manner tier and the melody tier. Each of the two elements receives a different interpretation depending on the tier where it appears. While on the skeleton C and V simply indicate syllabic positions linked to temporal units, on the manner tier and the melody tier they are normally in a dependency relationship with another element, i.e., both can be the head or the operator of a relationship. Headedness is indicated by underlining.

(100) Level-1 representation[32]

(a) Position: consonantal slot in the Skeleton, Manner tier

Element	*Interpretation*
$\underline{C}$	Voiceless stop
$\underline{C}$,V	Voiced stop
C	Voiceless fricative
C,V	Voiced fricative
$\underline{V}$,C	Nasal

---

[32] The comma stands for co-occurrence of two elements on the same tier.

V,C	Lateral
V	Rhotic
$\underline{V}$	Glide

(b) Position: vocalic slot in the skeleton, Manner tier

Element	*Interpretation*
$\underline{V}$	Vowel
V	Rhotic vowel
$\underline{V}$,C	Nasal vowel
C	Voiceless vowel
$\underline{C}$	Glottalized vowel

(c) Melody tier (both consonantal and vocalic slots)

Element	*Interpretation*
$\underline{C}$	Labial/Palatal
C	Dorsal/Coronal
$\underline{V}$	Low/loud
V	Central

Just like in DP, the difference in sonority is not represented by a difference in complexity but by the prominence of C and V. For instance, consonantal lenition consists of a loss of headedness (stop → fricative = $\underline{\mathbf{C}}$ → **C**) and/or the loss of consonantality/acquisition of vocalicness (fricative → approximant = **C** → **V,C**). Since manner is always specified, even when redundant, and laryngeal specifications are subsumed by manner specifications, intervocalic voicing is easily explained by this model. A voiceless consonant occurring between two vowels, which are specified for **V̲** on their manner tier, acquires **V** on its own manner tier, thus becoming voiced. (100b) must be taken cautiously, since it is only tentative. Vowels with elements other than **V** on their manner tier are relatively rare, therefore it is hard to model their structure. However, since in casual speech there are often cases of rhoticity spreading from a consonant to a vowel, I assumed that the representation of rhoticity must be the same for both consonants and vowels, i.e., **V** on the manner tier. In (100c) it is claimed that, on the melody tier, **C̲** translates into labiality/palatality and **C** into dorsality/coronality. Put differently, my assumption is that labials and palatals are somewhat more consonantal than dorsals and coronals. As a matter of fact, labial and palatal consonants prefer to appear in strong positions, such as word-initially, whereas coronals are the default epenthetic consonants between two vowels and dorsals are attracted to the coda position. Both the intervocalic and the final position can be considered to be less consonantal than the word-initial one. **V̲** is uncontroversially interpreted as lowness in vowels and as loudness in consonants (for instance, it characterizes stridents and coronal sonorants), while **V** is a weakened version of **V̲**, which could tentatively identify reduced (schwa-like) vowels and flaps/taps.

If the level-1 representation is somewhere between the UF and the SF, in order to be fed to the AF/ArtF (and be phonetically interpreted), elements are further specified in the SF and acquire the labels that we are already familiar with. For example, **V,C** on the manner tier is translated into **N** ("nasality"), while **C̲** and **C** on the melody tier are split into either **I** or **U**, where **C̲** = **I̲** or **U̲** and **C** = **I** or **U**. The unity of **I** and **U** is proved by the fact that they display a similar behavior, e.g., consonants preferably contain either **I** or **U** or a combination of the two on their melody tier (as in /ʃ, ʒ/, which in a number of languages are often described as rounded), whereas the occurrence of **A** on the melody tier of consonants is more marked. Conversely, low vowels (those characterized by **A**) are less marked than high vowels (those containing **I** or **U**) and the co-occurrence of **I** and **U** in a vowel is highly marked (as in the relatively rare phonemes /y, ø/).

(101) Translation between Level-1 and Level-2

Skeleton	*Tier*	*Level-1*	*Level-2*	*Phonetics*
C	Manner	$\underline{C}$	?-h	stopness
C	Manner	C	h	noise
C	Manner	$\underline{V}$,C	N	nasality
C	Manner	V	R	low F3
C/V	Melody	$\underline{V}$	$\underline{A}$	aperture
C/V	Melody	V	A	centrality
C/V	Melody	$\underline{C}$	$\underline{I}$ or $\underline{U}$	labiality/palatality
C/V	Melody	C	I or U	coronality/dorsality

(102) Representations and elements

|UF|
↓
Level-1 Elements: **C**, **V**
/SF/
Level-2 Elements: **?, h, I, U, A, N, R**…
↓
[AF] (noise, periodicity, formant values, etc.)
[ArtF] (occlusion, lips/tongue movements, etc.)

5.4.2 Constraints formulation

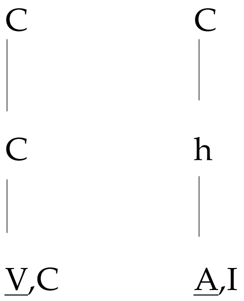
At this point, the following predictions can be made: (1) informative combinations of elements or occurrences of elements in certain positions will tend to be preserved by faithfulness constraints, and, (2), markedness constraints will promote the occurrence of default projections. In other words, given a C position on the syllabic skeleton, the default projection of C on the manner tier will be $\underline{\textbf{C}}$ or **C**, i.e., an obstruent, and the default projection on the melody tier will be $\underline{\textbf{C}}$ or **C**, i.e., any of the four major places of articulation for obstruents (labial – coronal – palatal – dorsal)[33]. Markedness constraints militate for a consonantal position to be occupied by an obstruent (rather than by a sonorant or a glide), while other markedness-related factors will select the place of articulation. For instance, it is less marked for a word-initial consonant to be a labial or a palatal rather than a coronal or a dorsal. Similarly, it is less marked for an

---

[33] The glottal stop would then be interpreted as the realization of a consonant with an empty melody tier.

intervocalic or final consonant to be coronal or dorsal rather than labial or palatal. While markedness constraints favor this state of affairs, faithfulness constraints preserve the unexpected. As stated before, the co-occurrence of **V**-elements with **C**-elements in a consonant is very informative, since it is unexpected (not projected by default), and therefore strident stops and fricatives are rarely deleted. I argue that constraints can refer to both level-1 and level-2 elements, therefore a very general constraint such as Max-V&CinC will protect any occurrence of **V**-elements together with **C**-elements in a consonant, regardless of the tier considered, whereas a more specific constraint such as Max-MannerC&MelodyVinC will specifically target stridents, and might also be translated into Max-h&AinC.

(103) Level-1 and Level-2 representation of /s/



5.5 Universality and word-specificity of the invariant

It is now time to answer the question: do we need a Max-Invariant constraint? The answer is both yes and no. In other words, we do need it but as an umbrella name for a group of constraints preserving the most informative units composing the structure of a word. These units can be informative for several reasons, listed below.

5.5.1 Positional factors

Constraints of the type MAX-O, R_ or ONSETC[PLACE] ("Do not delete an obstruent occurring after a sonorant" and "Onset consonants must be specified for place") refer to structural characteristics, i.e., syllabic positions, without referring to specific elements. They are bound to exert an important role universally, i.e., they are likely to be ranked high in every language. Here the salience of consonants is not decided by their inherent characteristics but by the fact that they occur in positions where the cues for their recognition are rich (phonetically speaking) and/or where they are licensed (phonologically speaking).

5.5.2 Inherent characteristics

Some sounds are simply more acoustically salient than others, and acoustic salience often translates phonologically into headedness. Faithfulness constraints preserving headed elements are, for theory-internal reasons, always ranked higher than constraints preserving the unheaded counterpart of the same elements. Moreover, elements have an inherent loudness, e.g., **A** is louder than **I** and **U** (level-2) and **V** is louder than **C** (level-1).

5.5.3 Informativeness

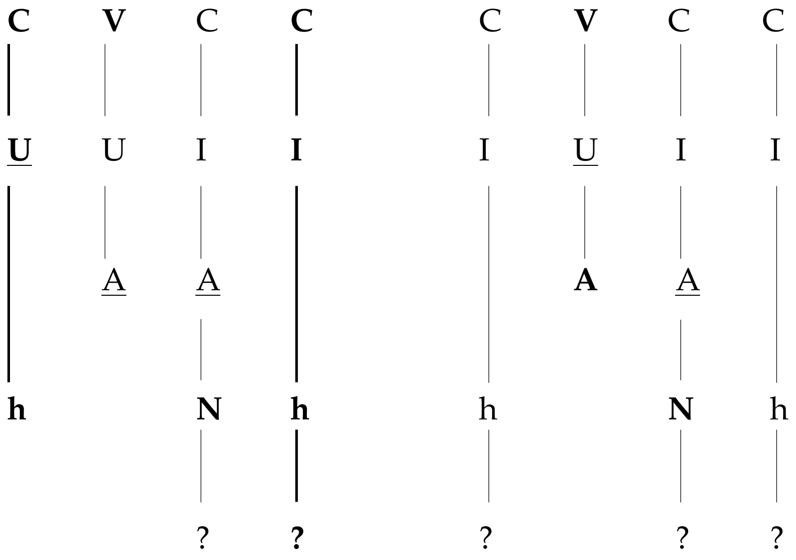
Since in a C position the occurrence of **C** on the manner tier and on the melody tier is to be expected, and similarly in a V position the occurrence of **V** on the manner tier and on the melody tier is to be expected, any deviation from this pattern is informative. Faithfulness constraints preserving the unexpected occurrence of elements in certain position are ranked higher than constraints preserving the default. In a similar vein, since melody is more informative than structure, faithfulness to the syllabic structure and to the distribution of elements on the skeleton can be violated more easily than faithfulness to melody.

5.5.4 Unevenness in the representation

Not all words are treated in the same way, more frequent words have a higher range of variation than less frequent words, even when they are structurally and melodically very similar. For instance, in English *don't* can be reduced to a much greater extent than *font*. I proposed that this fact be explained by unevenness in the representation. This difference between *font* and *don't* is mainly due to two factors: (1) *font* is a content word, while *don't* is a function word, and (2), *don't* is more frequent than *font.* However, I do not expect phonology to have access to this kind of information. The reasons why *don't* can reduce more significantly than *font* are extra-phonological (morphosyntactic, semantic, pragmatic) and all phonology can do is model their representations according to their pronunciation variants. If *don't* is realized with more variation than *font*, then its SF will be more uneven. For explanatory purposes, let us assume that *font* in spoken English can be reduced, at most, to [fn̩t], whereas, as presented in Bybee (2006a) and shown in (8), the reduction of *don't* can go as far as to reach [ə̃].

(104) *font* vs. *don't* (level-2 representation)

(a) *font* (b) *don't*



(104a) shows that the invariant of *font* consists of the full initial and final consonants, a vocalic slot and a nasal element, whereas the invariant of *don't* in (104b) is simply formed by a vowel position linked to **A** and a floating nasality[34]. Without positing uneven surface representations, it is impossible to account for the difference between the two words, since no constraint ranking can explain why the final consonant is never deleted in *font* and may or may not be deleted in *don't.* I am intentionally ignoring the fact that /f/, being labial and continuant, is more likely to be preserved than /d/, which is a coronal stop, and I will focus solely on the fact that they both appear in a position – the beginning of the word – where deletion does not normally occur in English, even assuming that in running speech that position becomes intervocalic. The same set of

---

[34] The actual pronunciation of *don't* is [dəʊnt/doʊnt] and therefore its representation would require a more complex structure, but for the sake of the analysis I am sticking to Bybee's notation, i.e., [dont].

relevant constraints is employed for both words: MAX-N ("Preserve nasality"), MAX ("Do not delete any segment"), ONSET ("Syllables must have onsets"), *COMPLEXCODA ("The coda position cannot host more than one consonant") and NOCODA ("Syllables must not have codas").

Tableau 20: *font,* unreduced

/fɒnt/	MAX-N	MAX	ONSET	*COMPLEX CODA	NOCODA
☞a) [fɒnt]				*	*
b) [fn̩t]		*!			*
c) [ə̃]		***	*		
d) [fɒt]	*!	*			*

In tableau 20 the winner is the unreduced form, [fɒnt], since all the relevant faithfulness constraints dominate *COMPLEXCODA and NOCODA, which are markedness constraints. (b) is ruled out because /ɒ/ is deleted and its nuclear position is occupied by the nasal, whereas (c) is precluded because it fatally violates MAX three times (one violation for every input segment missing). Candidate (d) avoids the violation of *COMPLEXCODA by deleting /n/, but fatally violates the highest ranking constraint MAX-N.

Tableau 21: *font*, reduced

/fɒnt/	Max-N	Onset	*Complex Coda	Max	NoCoda
a) [fɒnt]			*!		*
☞b) [fn̩t]				*	*
c) [ə̃]		*!		***	
d) [fɒt]	*!	*			*

In tableau 21, by promoting Onset and *ComplexCoda over Max, the winning candidate becomes (b), because it preserves both the onset consonant and the element **N** and avoids violating *ComplexCoda. Since the relative ranking of markedness constraints cannot be manipulated, there is no ranking under which candidate (c) would be the winner (*ComplexCoda cannot dominate Onset). Conversely, the hypothetical ranking *ComplexCoda > MaxN, Max would select candidate (d) as the winner, although to my knowledge [fɒt] would hardly be an attested pronunciation of /fɒnt/. Assuming that nasality forms part of the invariant of *font*, Max-N could not be demoted and candidate (d) would never emerge as the winner.

Tableau 22: *don't*, unreduced

/dont/	MAX-N	MAX	ONSET	*COMPLEX CODA	NOCODA
☞a) [dont]				*	*
b) [ɾõt]		*!			*
c) [ə̃]		***!	*		
d) [dot]	*!	*			*

Tableau 22, showing the victory of the unreduced variant of *don't*, has the same ranking as Tableau 20. Possessing an almost identical structure, *font* and *don't* behave similarly.

Tableau 23: *don't*, partially reduced

/dont/	MAX-N	ONSET	*COMPLEX CODA	MAX	NOCODA
a) [dont]			*!		*
☞b) [ɾõt]				*	*
c) [ə̃]		*!		***	
d) [dot]	*!	*			*

As in tableau 21, if *COMPLEXCODA is promoted over MAX, (b) emerges as the winning candidate, and [ɾõt] is actually an attested pronunciation of *don't*, as reported by Bybee (2006a). I will not deal with the flapping of the initial consonant here, since it is not relevant for the analysis. Candidate (b) is still preferable to (c) because it has an onset and to (d) because it preserves nasality.

Tableau 24: *don't*, heavily reduced

/dont/	Onset	*Complex Coda	NoCoda	Max-N	Max
a) [dont]		*!	*		
💣b) [ɾõt]			*		*
☞c) [ə̃]	*!				***
d) [dot]			*	*!	*

Tableau 24 fails to predict that the heavily reduced variant of *don't* is [ə̃] and the bomb symbol indicates that candidate (b) is wrongly selected instead. Even by promoting all markedness constraints to the top of the hierarchy, there is no way that an onsetless word can be considered to be better formed than one with an onset, unless we resort to the representation of *don't* given in (104b). Given that [ə̃] is the pronunciation variant arising in allegro/allegrissimo style, it might be that speakers feel the necessity to get rid of as much phonetic material as possible, in order to reduce the number of articulatory gestures. Or, if one wants to abstract away from phonetics, to reduce complexity as much as possible. I propose that in allegro style faithfulness to the invariant is ranked higher than faithfulness to the full representation. As proposed in 2.3, a *Weak(Invariant) constraint could be at play, deleting everything that does not receive a "special" protection in the SF. *Weak(Invariant) would be ranked very low in more careful styles and would be promoted in allegro style, deleting every phonological unit which is not part of the invariant.

Tableau 25: *don't*, heavily reduced

/dont/ Invar: **V-A-N**	*WEAK (INVARIANT)	ONSET	*COMPLEX CODA	NO CODA	MAX-N	MAX
a) [dont]	*!**…		*	*		
b) [ɾõt]	*!*…			*		*
☞c) [ə̃]		*				**
d) [dot]	*!...			*	*	*

Tableau 25 shows that candidate (c) can emerge as the winner if *WEAK(INVARIANT) is undominated in the hierarchy, having been promoted over both faithfulness and markedness constraints. I propose considering *WEAK(INVARIANT) as belonging to neither of the two constraint families, since it is not properly categorizable as either a markedness or a faithfulness constraint. Like markedness constraints, its effect is to delete material, but like faithfulness constraints, it needs an input and an output to refer to. The invariant of *don't* is represented in the tableau by **V-A-N**, i.e., a vocalic slot associated with unheaded **A** and nasality[35]. Candidates (a, b, d), which preserve units not present in the invariant, are immediately ruled out. As for candidate (d), if **N** forms part of the invariant of *don't*, markedness constraints would not be able to dominate MAX-N and therefore [dot] would never qualify as the winning candidate.

---

[35] Bybee (2006a) reports tokens of *don't* only consisting of (non-nasalized) schwa, but such pronunciation emerges, most of the time, before *know*, which begins with a nasal, suggesting that in those cases **N** may have been absorbed by the following consonant.

5.6 Interaction between informativeness, salience and headedness

The consonants that have been the object of my analysis throughout the thesis have shown to have different degrees of resistance to deletion, reduction and assimilation. They have also proved to differ in perceptibility in plateau clusters. While vulnerability and perceptibility can be explained on a phonetic basis (acoustic and articulatory), I have also proposed that the degree of informativeness of each segment depends on its internal structure and plays a role in determining its phonological salience. I will now give some examples. Since a C slot is expected to project **C̲** on its manner tier, being a stop for a consonant is uninformative. Being a fricative (having **C** on the manner tier) is slightly more informative, and informativeness increases as consonantality is lost and vocalicness is acquired. Following this logic, the most informative consonant, manner-wise, would have **V** or **V̲** on its manner tier (i.e., an approximant or a glide). On the melody tier, it is the degree of similarity with the manner tier which determines informativeness. In other words, overall informativeness depends, on the one hand, on the relation between the skeleton and the manner tier, and on the other hand, on the relation between the manner tier and the melody tier. For instance, in (105), [p] is considered uninformative because every position on each tier is occupied precisely the same element[36]. [n] displays the unexpected occurrence of **V̲** on its manner tier, but then its informativeness is lowered since the melody tier contains exactly the same elements as the manner tier. [s] is highly informative, having **C** (and not **C̲**) on its manner tier and **V̲** on its melody tier.

---

[36] C and V on the skeletal tier projects, by default, **C̲** and **V̲** on the lower tiers. Therefore, skeletal C and V must be considered as always headed.

(105)

Skeleton	*Manner*	*Melody*	*Example*	*Informativeness*
C	$\underline{C}$	$\underline{C}$	[p]	very low
C	C	$\underline{C}$	[f]	low
C	C	$\underline{V}$,C	[s]	high
C	$\underline{V}$,C	$\underline{V}$,C	[n]	medium
C	V	$\underline{V}$,C	[r]	medium

In the representation I propose, the manner tier is closer to the skeleton than the melody tier. As a matter of fact, manner is primary when it comes to sound categorization. If manner is informative – a consonant is **V**-headed or contains only **V** on its manner tier – then it becomes less important to preserve melody. As a result, obstruents tend to preserve their melody to a greater extent than sonorants. Nasals contain $\underline{\mathbf{V}}$,**C** on their manner tier, and liquids contain either $\underline{\mathbf{V}}$ or **V**. Accordingly, nasals typically agree in place with the following consonant and place is rarely distinctive in liquids crosslinguistically. Conversely, when melody is more informative than manner, as in the case of sibilants, it is less likely for it to be dispensed with.

Informativeness interacts with acoustic salience and markedness, and while the role of the former is likely to affect perception in all languages, the role of the latter depends on the language-specific constraint ranking. However, by way of conclusion, it might be useful to give an overview of the degree of acoustic salience, vulnerability and informativeness of the consonants under analysis. The aspects considered are the following: does the internal structure of the consonants contain unexpected combinations?

Is the consonant acoustically salient? Is the consonant phonologically salient (is there a headed element on its melody tier)? Is the consonant easily deleted, reduced or assimilated?

5.6.1 /s/

Informativeness. /s/, like all fricatives, is more informative than stops, because it hosts **C** (and not **C̲**) on its manner tier. However, most of its informativeness comes from the melody tier, where the head is **V̲**.

Acoustic salience. /s/ is uncontroversially highly perceptible, even when occurring in syllabic positions lacking vocalic support (e.g., word-finally, pre-consonantally).

Phonological salience. /s/ is **V**-headed on the melody tier, it is therefore phonologically salient.

Vulnerability. Because of its high informativeness and salience (both acoustic and phonological), /s/ is practically invulnerable in Italian and presumably in many other languages.

5.6.2 /ʃ/

Informativeness. The postalveolar fricative is still more informative than a stop, but less than /s/, since it is melodically **C**-headed (yet containing **V**).

Acoustic salience. The spectral characteristics of /ʃ/ are similar to those of /s/.

Phonological salience. /ʃ/ is **C**-headed on the melody tier, it is therefore phonologically salient.

Vulnerability. /ʃ/ appears to be very resistant, but its distribution is more limited than that of /s/.

5.6.3 /f/

Informativeness. /f/ is the least informative of the three fricatives under analysis, since it contains **C̲** on its melody tier (and arguably not **V**, see (82)).

Acoustic salience. Acoustically, /f/ is not easily audible in adverse environments.

Phonological salience. /f/ is **C**-headed on the melody tier, it is therefore phonologically salient.

Vulnerability. /f/ appears to be very resistant, but its distribution is more limited than that of /s/.

5.6.4 /k, p, t/

Informativeness. Mannerwise, stops are less informative than fricatives and sonorants, since they are the default consonants. /k, p, t/ are also melodically uninformative, since they are all characterized by **C** (headed or not).

Acoustic salience. It is hard to establish which stop is absolutely more salient than the two others. /k/ appears to be highly dependent on the vocalic context and /p/ on its syllabic position. There is general agreement that /k, p/ are more salient than /t/, although some experimental results contradict this claim.

Phonological salience. /p/ should be stronger than both /k/ and /t/, since /p/ is **C**-headed and /k/ and /t/ are headless.

Vulnerability. Generally, /k/ is considered to be the most vulnerable of the three, followed by /t/ and then by /p/.

5.6.5 /m/ and /n/

Informativeness. Nasals are more informative than obstruents because they contain $\underline{\mathbf{V}}$ on their manner tier (on which **C** depends). Melodically, they are almost equivalent, since /m/ consists of $\underline{\mathbf{C}}$,**V** and /n/ of $\underline{\mathbf{V}}$,**C**.

Acoustic salience. While nasality is *per se* very audible, /m/ is deemed to be "more nasal" than /n/, and therefore more salient.

Phonological salience. Both contain a headed element, but /n/ is more likely to lose its place specification (since place is indicated by $\underline{\mathbf{C}}$ in /m/ and by unheaded **C** in /n/).

Vulnerability. /n/ is deleted and reduced more systematically than /m/.

5.6.6 /r/ and /l/

Informativeness. Liquids should be very informative, since they are **V**-mannered consonants. Melodically, however, they both contain $\underline{\mathbf{V}}$, which in this case is expected, and therefore, uninformative.

Acoustic salience. Both display vowel-like formants and are therefore quite loud. /r/, in its prototypical realization, is likely to be more audible than /l/, given its higher sonority and greater assimilatory effects on the adjacent sounds.

Phonological salience. Both contain a headed element.

Vulnerability. While /r/ proved to be slightly more resistant than /l/ in this study, languages with a different phonetic realization of the rhotic might exhibit the opposite pattern.

6

CONCLUSION

The current study started out with many questions and ends, hopefully, with some answers. One of the main challenges consisted of combining together three theoretical approaches to the *problem of phonetic variation*. While OT typically employs SPE-style representations (strings of segments – each of the segments standing for a bundle of binary features), I argued for the superiority of multilinear representations, displaying privative elements attached to syllabic positions in the skeleton. Following Boersma (2011), I assumed the existence of (at least) four levels of representation which are relevant and necessary for phonology: UF, SF, AF and ArtF. The locus of phonetic variation is the mapping between the SF and the AF, where the SF is still abstract (although less abstract than the UF) and the AF is the realization of abstract cognitive elements and configurations. In addition, I borrowed the concept of entrenchment from usage-based linguistics arguing that some of the units composing the SF are entrenched more deeply than others, i.e., they constitute the "core essence" of the phonological word. Since such a model risks being highly stipulative, it was necessary to explain why these units have a stronger status in the representation, and I proposed that elements occurring in all or most of the pronunciation variants of a word have higher chances of forming part of its *invariant*. Obviously, the more frequent a word, the more *uneven* its representation. But then the question became: why are certain elements likely to be preserved in every variant and others easily dispensed with? My guess was that a significant role is played by *acoustic salience*. Sounds with strong acoustic cues are more likely to impress the listener and therefore become essential for word recognition. This tendency to preserve salient sounds often comes into conflict with articulatory ease, since highly salient sounds also tend to be articulatorily costly. In order to test my prediction, i.e., that in casual speech speakers tend to preserve more salient segments and to dispense with less salient ones, the results of a perception experiment on plateau clusters were presented, together with data from consonantal reduction in spoken Italian. The following patterns clearly emerged:

- In obstruent clusters, stridents, the dorsal stop and the coronal stop are easily identified by listeners, whereas the cues of [f] and [p] are more likely to be obscured. Compared against typological data on the legality of word-initial and word-final plateau clusters in a number of languages, these results suggest that the preservation of a plateau cluster depends (among other factors) on the perceptibility of a consonant in pre-consonantal or word-final context. In the specific case of [t], it is hard to determine whether listeners identified it with ease because of its perceptibility or because of its frequency. Considering that, unlike other obstruents, [t] was sometimes hallucinated by the participants to the experiment, I suggest that frequency (and therefore, *predictability*) might have played a role. As for nasal and liquid clusters, [m] and [r] were identified correctly more often than [n] and [l], respectively. I interpret these results as evidence of the richer cues of the labial nasal and the rhotic (intended here as an apical alveolar trill).

- In spoken Italian, it was observed that, generally speaking, consonantal reduction is not such a pervasive phenomenon as in, say, Germanic languages, probably because of the syllable-timed nature of Italian (as opposed to stress-timed languages). However, Warner (2011:1872) suggests that the Map Task may produce speech that is still relatively careful, compared to actual spontaneous speech. The most significant result of the analysis is that, among obstruents, /s/ and /t/ are much more frequent than /k, p/, which are, in turn, more frequent than /ʃ/ and /f/. Among sonorants, /n/ is more frequent than /m/ and /r/ than /l/. Unsurprisingly, fricatives appear particularly resistant to lenition, especially if compared to stops. Of the three stops, /p/ is the most likely to be reduced, /t/ the most likely to be deleted and /k/ the most likely to be assimilated. Of the two nasals, /m/ is generally preserved more often than /n/, but is also more likely to be deleted, whereas /n/ is more prone to assimilation and reduction. The two liquids behave similarly but /r/ is slightly more resistant than /l/, which in turn is clearly more likely to undergo assimilation than the rhotic. According to the Two-Way ANOVA test, the following correlations were significant: the full realization of liquids depended on liquid type (/r/ vs. /l/, $p < .05$) and on position ($p < .001$), the full realization of nasals depended on nasal type (/m/ vs. /n/, $p < .05$) but not on position, whereas both the full and the reduced realization of obstruents depended on position ($p < .05$) but not on obstruent type. These results suggest that obstruents behave similarly in Italian and their actual phonetic realization depends more on the syllabic position in which they occur than on their inherent properties.

- The collected data indicate that salience does play a role (/r/, /m/ and fricatives are indeed preserved more often than /l/, /n/ and stops respectively) but we also need to characterize salience from a phonological point of view. The case of /t/ is particularly useful. Several experiments have brought contradictory results as to whether /t/ is more or less salient than other stops. My study has revealed that /t/ is easily perceived in consonant clusters, both word-initially and word-finally, but is also deleted to a considerably great extent. I therefore suggest that *phonological salience* derives from the interaction of *informativeness* and melodic *headedness*. /t/ is an uninformative segment, since it occupies a consonantal slot in the skeleton (C), contains **C** on its manner tier and **C** on its melody tier. /t/ is also melodically weak, since it has no headed elements on its melody tier. Conversely, /s/ is both informative, since its manner and melody tiers do not agree (**C** on the manner tier and **V**,**C** on the melody tier) and the melody tier contains a headed element, which also happens to be **V**, standing for loudness. The difference between /s/ and /t/ is particularly relevant since they are the two most frequent obstruents in the dialogues analyzed but, while the former is preserved most of the time, regardless of context, the latter is prone to deletion, meaning that here the role of frequency is not explanatory.

At this point, it is arguable that the invariant of a word consists of *contextually salient* segments (e.g., non-intervocalic onset consonants, stressed vowels) and *inherently salient* segmental and subsegmental units, both from an acoustic and a phonological point of view (e.g., strident consonants, nasality, etc.). It must be admitted, however, that variation is often *word-specific*, since highly frequent words display a wider range of variation. Consequently, I propose that words are stored together with information about their invariant, or at least *the part of the invariant that cannot be deduced by contextual or inherent salience*. Assuming that this information is present in the representation, it is possible to postulate a family of constraints referring to the invariant – which is specified on a word-specific basis each time. Such constraints could be, tentatively, Max-Invariant and *Weak(Invariant). The former would always be, by definition, undominated in the hierarchy, whereas the latter would be low-ranked in careful/lento style and would be promoted in casual/allegro style, triggering the deletion of all the segments and elements that do not form part of the invariant. I argue that it is desirable that the role of this constraint family be minimal, since most casual speech phenomena can be explained through constraints of the type Max-*x*-in-*y*, where *x* is an element on the manner tier or on the melody tier and *y* a syllabic position. The constraints exhibiting this template are presumably universally ranked, as the least predictable tend to be preserved at the expense of the most predictable, e.g., Max-V̲(manner)-in-C is ranked higher than Max-C̲(manner)-in-C. In other words, faithfulness to sonorancy in consonants is normally ranked higher than faithfulness to occlusion in consonants. Nevertheless, stops are more frequent than sonorants crosslinguistically, but this fact is easily explained by the action of markedness constraints, which promote the occurrence of **C**-elements in consonants and of **V**-elements in vowels. MAX-*x*-IN-*y* constraints interact with *cue constraints* (*à la* Steriade) that preserve consonants occurring in positions where their cues are easily perceptible, such as MAX-O, R_ ("Do not delete obstruents occurring after a sonorant"). Assimilation phenomena typical of connected speech are explained by SPREAD constraints, which promote the sharing of the same element between two or more syllabic positions, e.g., SPREAD-N, SPREAD-I, etc. Finally, reductions and deletions are triggered by sensorimotor constraints, aiming at reducing articulatory gestures, and (structural) markedness constraints, driving towards an unmarked syllable structure (e.g., NOCODA).

6.1 Residual issues

6.1.1 On markedness

I already tackled the question of whether the concept of informativeness is necessary, as it might be subsumed by markedness, a more familiar label to phonologists. However, I wish to maintain this terminology for two reasons. Firstly, as I have showed earlier, markedness and informativeness do not always make the same predictions. For instance, /s/ appears to be, at the same time, more informative and less marked than /f/ and /x/. Secondly, even if it were the case that the two concepts – markedness and informativeness – are overlapping, then informativeness is to be welcomed as a straightforward, theory-internal, falsifiable way to define markedness. The theory sketched in this thesis clearly states what is expected and what is not expected to occur in a given syllabic position, and any deviation from this pattern is informative. The degree of informativeness of a segment is clearly determinable by calculating how much its internal structure deviates from the default.

6.1.2 Faithfulness to V in C

According to the current proposal, since the occurrence of a **V**-element in a consonantal slot is more informative than the occurrence of a **C**-element in the same position, faithfulness should preserve **V**-elements in consonants. However, the case of *televisore* showed that reduction and assimilation can affect /r/ as well, which in a token becomes a labiovelar glide. This process implies the loss of a **V**-element (or **A**) and the acquisition of a **C**-element (or **U**), thus contradicting the prediction that **V**-elements should be preserved in consonants. The issue is easily solved if one assumes a ranking of the type: MAX-C(MANNER)-V(MELODY >> MAX-V(MANNER)-V(MELODY), i.e., it is more important to preserve **V**-elements when they occur on the melody tier of a consonant which contains **C**-elements on its manner tier, rather than when **V**-elements occur on both tiers. In other words, **V** is more informative in obstruents and nasals than in rhotics.

6.1.3 About U and I

In the fifth chapter I proposed that there exist two levels for elements, a more abstract one (Level-1), where they can only be either C or V and they are distinguished by their position and their prominence (headed vs. non-headed), and a less abstract one (Level-2), where they “translate” into a greater number of units which can be read off by other modules, such as the acoustic and the articulatory one. One pending question is how the element **C** on the melody tier is split into **I** and **U**. Tentatively, it can be said that with regards to the constraint ranking, what actually defines **I** and **U** is that they are non-**A**. It can therefore be expected that MAX-A-IN-C will be ranked higher than MAX-I-IN-C and MAX-U-IN-C. From the data in my possession, it is not possible to establish whether it is **I** or **U** which tends to be preserved more often. Unsurprisingly, **I**-headed and **U**-headed consonants (palatals and labials) are more resistant than their headless equivalent (coronals and velars). However, when it comes to vowel reduction and deletion, it is important to keep in mind that markedness constraints play a prominent role. Undoubtedly, the prototypical reduced vowel is [ə], which I assume to be the expression of non-headed A. If /o/ and /e/, i.e., combinations of **A** and **U** and **A** and **I**, respectively, are frequently reduced to simple **A**, it means that **A** is preserved more often than other elements in V. The occurrence of **A** in a V slot is less marked than the occurrence of other elements, therefore markedness selects [ə] over [o, e] in unstressed positions and in running speech.

6.1.4 /s/-debuccalization

Finally, although my model predicts /s/ to be practically untouchable, in a number of languages /s/ undergoes debuccalization, which consists in the loss of **A** (or the **V**-element) and **I**. The two most famous examples are Caribbean Spanish and Proto-Greek. In the former, /s/ is realized as [h] when occurring in the coda, even when followed by a word beginning with a vowel. Similarly, the prefix _des-_ is pronounced [deh-] even when it precedes a vowel-initial stem (Shepherd 2003).

(106)

	European Spanish	Caribbean Spanish	Gloss
las alas	[lasalas]	[lahalah]	‘the wings’
deshacer	[desaθer]	[dehaser]	‘to undo’

The debuccalization of /s/ has been even more pervasive in Proto-Greek, which, at some point of its history, turned /s/ into /h/ word-initially and between sonorants, e.g., PIE _*septm̥, *sal_ → Ancient Greek _heptá_ ‘seven’, _hals_ ‘salt’, PIE _*genesos, *h$_1$esmi_ → Proto-Greek _*genehos_ ‘of a race’, _*ehmi_ ‘I am’ (Sommerstein 1973).

How can my model account for these phenomena? It is important to point out that, at least to my knowledge, there are no language where /s/ undergoes debuccalization in certain position and other fricatives do not. In the case of Spanish, /s/ is the only obstruent (with few exceptions) allowed to occur in the coda, so obviously debuccalization can only apply to it[37]. Moreover, in the history of Spanish another obstruent had turned into [h], namely /f/, but that occurred even word-initially. As for Proto-Greek, its phonological inventory is not reconstructable with certainty, but it appears likely that /s/ was the only fricative. Moreover, while I argued for a (supposedly) universal ranking of faithfulness constraints, based on informativeness, it might be that the relative ranking of markedness constraints is language-specific and markedness penalizes the occurrence

---

[37] Other obstruents, when occurring in the coda, tend to be realized as approximants.

of **V̲** in a consonantal position. In sum, the fact that in some languages /s/ undergoes debuccalization is not problematic, given that it can be explained ranking the relevant markedness constraints higher than faithfulness constraints.

6.2 Further research

The focus of this thesis was on Italian casual speech phenomena, although the participants in the perception experiments were also Dutch. The most obvious next step would be to put phonetically-annotated spoken corpora of Dutch under scrutiny and look for a relationship between acoustic reduction and perceptibility of segments, as has been done for Italian in the current study. Subsequently, the same approach could be applied to other languages, perhaps starting with the ones whose corpora are already available, such as English, French, Spanish, Danish, German, etc. Ending up with a general overview of the connection between acoustic prominence and resistance to reduction/deletion would help phonologists understand what processes are universal (and thus arguably based on human perception and articulation) and what processes are dependent on the language-specific phonology (e.g., languages with smaller inventories might display greater allophonic variation than languages with bigger inventories).

With regards to the theoretical approach, I demonstrated that it is possible to fruitfully combine different theories, such as OT, ET, DP, etc. in order to obtain a more realistic description of phonetic and phonological phenomena. My stance here is close to that of Aronoff & Fudeman (2011), who claim to follow an “anything-goes approach”. They go on by explaining that “[they] take a noholds-barred approach to linguistics, [they]’ll use any tool or method that will tell [them] how language works”. Basically, since every theory is obviously good for something, otherwise it would have no practitioners, I tried to the get the best that I could from different approaches, depending on the phenomena that needed to be analyzed. Whether such a “spurious” method could be successfully applied also to other linguistic phenomena is another question left for future research.

REFERENCES

ALBANO LEONI F., CUTUGNO, F. & LAUDANNA, A. 1999. L'attivazione di rappresentazioni fonemiche durante il riconoscimento del parlato: una risorsa metalinguistica? In MIONI, A., BENINCÀ, P. & VANELLI, L. (eds.), *Fonologia e Morfologia dell'Italiano e dei dialetti d'Italia,* Atti del XXXI Congresso della Società di Linguistica Italiana, pp. 35-51. Roma: Bulzoni.

ANDERSON, J. M. & EWEN, C. J. 1987. *Principles of Dependency Phonology.* Cambridge: Cambridge University Press.

ANDERSON, G. D. S. 2011. The Velar Nasal. In DRYER, M. S. & HASPELMATH, M. (eds.), *The World Atlas of Language Structures Online*. Chapter 9. Munich: Max Planck Digital Library. Available online at http://wals.info/chapter/9. Last accessed on 23 July 2013.

ANTTILA, A. 1997. Deriving variation from grammar. In HINSKENS, F., VAN HOUT, R. & WETZELS, L. (eds.), *Variation, Change and Phonological Theory,* pp. 35-68. Amsterdam: John Benjamins.

ARONOFF, M. & FUDEMAN, K. 2011 (2nd edition). *What is morphology?* Malden, MA: Wiley-Blackwell.

ARTUSO, C. 1998. *Noogom gaa-izhi-anishinaabemonaaniwag: Generational Differences in Algonquin.* Master thesis, University of Manitoba.

BACKLEY, P. 2011. *An Introduction to Element Theory.* Edinburgh: Edinburgh University Press.

BARD, E. G., SOTILLO, C., KELLY, M. L. & AYLETT, M. P. 2001. Taking the hit: Leaving some lexical competition to be resolved post-lexically. *Language and Cognitive Processes* 16. 731-737.

BARKER, M. A. R. 1964. *Klamath grammar*. University of California Publications in Linguistics volume 32. Berkeley & Los Angeles: University of California Press.

BARONI, A**.** 2011. Alphabetic vs. non-alphabetic writing: Linguistic fit and natural tendencies. *Italian Journal of Linguistics/Rivista di Linguistica.* 23:2. 127-159.

BARONI, A. 2012a. A Beats-and-Binding Account of Italian Phonotactics. *Padua Working Papers in Linguistics* 5. 45-72.

BARONI, A. 2012b. Romance glides revisited: An OT analysis of high vowels/glides alternation in French and Italian. Talk presented at the *International Conference on Romance Languages – In Memoriam Alf Lombard*, Lund University, Sweden, 7-8 November 2012.

BARONI, A. 2013. Eye Dialect and Casual Speech Spelling – Orthographic Variation in OT. *Writing Systems Research* 5:1. 24-53.

BARONI, A. (in press). On the importance of being noticed: the role of acoustic salience in phonotactics and casual speech. In CALDERONE, B. & CELATA, C. (eds.) *Theory and evidence in the study of (mor)phonotactics* (special issue of *Language Sciences*).

BARONI, A. & SIMONOVIĆ, M. 2013a. Articulating the Inarticulate. When the stimulus is in the read. Talk presented at the *39th Incontro di Grammatica Generativa*, Modena & Reggio Emilia, 21 – 23 February 2013.

BARONI, A. & SIMONOVIĆ, M. 2013b. A Perception Based Account of Variation. Talk presented at *Approaches to Phonology and Phonetics* (APAP), Lublin, 21-23 June 2013.

BARONI, M. 1999. The representation of prefixed forms in the Italian lexicon: evidence from the distribution of intervocalic [s] and [z] in Northern Italian. *Yearbook of Morphology 1999*. 121-152.

BASBØLL, H. 2005. *The Phonology of Danish*. Oxford: Oxford University Press.

BASBØLL, H. 2012. Syllable-word interaction: Sonority as a foundation of (mor)phonotactics. Talk presented at *Phonotactic Grammar: Theories and Models*, Scuola Normale Superiore, Cortona (AR), Italy, 28-29 September 2012.

BEA Hungarian spontaneous speech database. 2013. Available at http://www.nytud.hu/adatb/bea/index.html. Last accessed on 13 December 2013.

BERTINETTO, P. M. 1989. Reflections on the dichotomy 'stress- vs. syllable-timing'. *Revue de Phonétique Appliquée* 91/93. 99-130.

BERTINETTO, P. M. & LOPORCARO, M. 2005. The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association* 35:2. 131-151.

BIČAN, A. 2011. *The Phonotactics of Czech*. PhD dissertation, Masaryk University, Brno.

BLAHO, S. 2008. *The Syntax of Phonology. A Radically Substance-free Approach*. PhD dissertation, University of Tromsø.

BLEVINS, J. 2004. *Evolutionary Phonology. The Emergence of Sound Patterns*. New York: Cambridge University Press.

BLEVINS, J. & GARRETT, A. 2004. The evolution of metathesis. In HAYES, B., KIRCHNER, R. & STERIADE, D. (eds.), *Phonetically based Phonology*, pp. 117-156. Cambridge, MA: Cambridge University Press.

BOAS, F. & DELORIA, E. 1976. *Dakota Grammar*. New York: AMS Press.

BOERSMA, P. & HAYES, B. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32. 45-86.

BOERSMA, P. 1997. How we learn variation, optionality and probability. In VAN SON, R. J. J. H. (ed.), *IFA Proceedings* 21, pp. 43-58.

BOERSMA, P. 1998. *Functional Phonology: Formalizing the Interaction Between Articulatory and Perceptual Drives*. PhD dissertation, University of Amsterdam.

BOERSMA, P. 2009. Cue constraints and their interactions in phonological perception and production. In BOERSMA, P. & HAMANN, S. (eds.), *Phonology in Perception*, pp. 55-110. Berlin: Mouton de Gruyter.

BOERSMA, P. 2011. A programme for bidirectional phonology and phonetics and their acquisition and evolution. In BENZ, A. & MATTAUSCH, J. (eds.), *Bidirectional Optimality Theory*, pp. 33-72. Amsterdam: John Benjamins.

BOERSMA, P., BENDERS, T., & SEINHORST, K. 2012. *Neural network models for phonology and phonetics.* Unpublished manuscript. Available online at http://www.fon.hum.uva.nl/paul/papers/BoeBenSei21.pdf. Last accessed on 24 January 2014.

BOERSMA, P., & WEENINK, D., 2013. *Praat: doing phonetics by computer* [Computer program]. Version 5.3.52, retrieved 12 June 2013 from http://www.praat.org/.

BOTMA, B. 2004. *Phonological Aspects of Nasality. An Element-Based Dependency Approach.* Utrecht: LOT.

BOTMA, B. 2009. Transparency in nasal harmony and the limits of reductionism. In NASUKAWA, K. & BACKLEY, P. (eds.), *Strength Relations in*

Phonology, pp. 79-112. Berlin & New York: Mouton de Gruyter.

BOTMA, E. D., KULA, N. C. & NASUKAWA, K. 2011. Features. In KULA, N. C., BOTMA, B., & NASUKAWA, K. (eds.), *The Continuum Companion to Phonology*, pp. 33-63. London & New York: Continuum.

BRANDÃO DE CARVALHO, J. & TRIFIT, A., 2013. Velar/coronal asymmetry in phonemic patterns and historical changes: a unified account. Talk presented at *GLOW 36*, University of Lund, Sweden, 2 - 6 April 2013.

BREIT, F. 2013. How many segments? Theories of segmental representation and their generative capacity. Talk presented at *Approaches to Phonology and Phonetics* (APAP), Lublin, 21-23 June 2013.

BROUWER, S., MITTERER, H., HUETTIG, F. 2012. Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes* 27. 539-571.

BROWMAN, C. & GOLDSTEIN, L. 1990. Tiers in Articulatory Phonology, with Some Implications for Casual Speech. In KINGSTON, J. C. & BECKMAN, M. E. (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, pp. 341-376. New York: Cambridge University Press.

BROWN, G. 1977. *Listening to Spoken English*. London: Longman.

BUCKLEY, E. 2011. Metathesis. In OOSTENDORP, M. VAN, EWEN, C. J., HUME, E. V. & RICE, K. (eds.), *The Blackwell Companion to Phonology*, vol. III, pp. 1380-1406. Oxford: Blackwell.

BÜRKI, A., ERNESTUS, M. & FRAUENFELDER, U. H. 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. 421-437.

BUTSKHRIKIDZE, M. 2002. *The consonant phonotactics of Georgian*. Utrecht: LOT.

BYBEE, J. 2001. *Phonology and Language Use.* Cambridge: Cambridge University Press.

BYBEE, J. 2006a. *Frequency of Use and the Organization of Language.* Oxford: Oxford University Press.

BYBEE, J. 2006b. From usage to grammar: The mind's response to repetition. *Language* 82. 711-733.

BYBEE, J. 2010. Markedness, iconicity, economy and frequency. In JUNG SONG, J. (ed.), *Handbook of Linguistic Typology*. Oxford: Oxford University Press.

CARR, P. 2005. Salience, Headhood and Analogies. In CARR, P., DURAND, J. & EWEN, C. J. (eds.), *Headhood, Elements, Specification and Contrastivity*, pp. 15-30. Amsterdam: John Benjamins.

CHARETTE, M. 1991. *Conditions on Phonological Government*. Cambridge: Cambridge University Press.

CHITORAN, I. 1994. Acoustic Investigation of Georgian Harmonic Clusters. *Working Papers of the Cornell Phonetics Laboratory* 9. 27-65.

CHOMSKY, N. & HALLE, M. 1968. *The Sound Pattern of English.* New York: Harper & Row.

CHOMSKY, N. 1957. *Syntactic Structures.* The Hague & Paris: Mouton.

CHOMSKY, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

CLEMENTS, G. N. & SEZER, E. 1982. Vowel and consonant disharmony in Turkish. In VAN DER HULST, H. & SMITH, N. (eds.), *The Structure of Phonological Representation – Part II*, pp. 213-255. Dordrecht: Foris.

CLEMENTS, G. N. 1977. The autosegmental treatment of vowel harmony. In DRESSLER, W. U. & PFEIFFER, O. (eds.), *Phonologica 1976*, pp. 111-119. Innsbruck: Universität Innsbruck.

CLOPPER, C. G., & PISONI, D. B. 2004. Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics* 32. 111-140.

COLE. J. & HUALDE, J. I. 2011. Underlying Representations. In OOSTENDORP, M. VAN, EWEN, C. J., HUME, E. V. & RICE, K. (eds.), *The Blackwell Companion to Phonology*, vol. I, pp. 1-26. Oxford: Blackwell.

COLE, R. A., COLTHEART, M. & ALLARD, F. 1974. Memory of a speaker's voice: Reaction time to same-or-different-voiced letters. *The Quarterly Journal of Experimental Psychology* 26. 1-7.

CRAWFORD, J. M. 1973. Yuchi phonology. *International Journal of American Linguistics* 39:3. 173-179.

CUTLER, A., WEBER, A., SMITS, R. & COOPER, N. 2004. Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustic Society of America* 116:6. 3668-3678.

CYRAN, E. 2003. *Complexity Scales and Licensing Strength in Phonology.* Lublin: KUL.

DALBY, J. 1986. *Phonetic Structure of Fast Speech in American English.* Bloomington: Indiana University Linguistics Club.

DANIELS, P. T. & BRIGHT, W. 1996. *The World's Writing Systems.* Oxford: Oxford University Press.

DART, S. N. 1991. *Articulatory and Acoustic Properties of Apical and Laminal Articulations*. PhD dissertation, *UCLA Working Papers in Phonetics* 79.

DAVIS, M. H., MARSLEN-WILSON, W. D., & GASKELL, M. G. 2002. Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 28. 218-244.

DEPREZ, V. 1988. Georgian Complex Segments. *Proceedings of NELS* 16. 109-123.

DONEGAN, P. & STAMPE, D. 1979. The study of natural phonology. In DINNSEN, D. (ed.), *Current Approaches to Phonological Theory*, pp. 126-173. Bloomington, IN: Indiana University Press.

DONEGAN, P. 1978. *On the Natural Phonology of Vowels*. PhD dissertation, Ohio State University.

DRESSLER, W. U. 2009. Natural Phonology as Part of Natural Linguistics. *Poznań Studies in Contemporary Linguistics* 45:1. 32-42.

DU BOIS, J. W., CHAFE, W. L., MEYER, C., THOMPSON, S. A., ENGLEBRETSON, R. & MARTEY, N. 2000/2005. *Santa Barbara corpus of spoken American English, Parts 1-4.* Philadelphia: Linguistic Data Consortium.

DZIUBALSKA-KOŁACZYK, K. 2002. *Beats-and-binding phonology*. Frankfurt am Main: Peter Lang.

DZIUBALSKA-KOŁACZYK, K. 2009. NP extension: B&B phonotactics. *Poznań Studies in Contemporary Linguistics* 45:1. 55-71.

DZIUBALSKA-KOŁACZYK, K. 2013. Are frequent, early and easy clusters also unmarked? Talk presented at the *2nd International Workshop on Phonotactics and Phonotactic Modeling*, Pisa, Italy, 21-22 November.

EIMAS, P. 1974. Auditory and linguistic processing of cues for place of articulation by infants. *Perception and Psychophysics* 16. 513-521.

EIMAS, P., SIQUELAND, E. R., JUSCZYK, P. & VIGORITO, J. 1971. Speech perception in infants. *Science* 171. 303-306.

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

ERNESTUS, M. 2013. Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua* (advance online publication).

FARNETANI, E. & BUSÀ, M. G. 1993. Italian clusters in continuous speech. *Proceedings of the International Conference of Spoken Language Processing 1994*, pp. 359-362. Yokohama, Japan, 18-22 September.

FARNETANI, E. 1995. The spatial and temporal dimension of consonant reduction in conversational Italian. *Proceedings of the 4th European Conference on Speech Communication and Technology, EUROSPEECH 1995*, Madrid, Spain, 18-21 September.

FERLUS, M. 1982. Spirantisation des obstruantes médiales et formation du système consonantique du vietnamien. *Cahiers de linguistique – Asie Orientale* 11. 83-106.

FERRAZ, L. I. 1987. The liquid in the Gulf of Guinea Creoles. *African Studies* 46:2. 287-295.

FIKKERT, P. & LEVELT, C., 2008. How does Place fall into place? The lexicon and emergent constraints in children's developing phonological grammar. In AVERY, P., DRESHER, B. E. & RICE, K. (eds.), *Contrast in Phonology. Theory, perception and acquisition*, pp. 231-268. Berlin: Mouton de Gruyter.

FIKKERT, P., VAN HEUGTEN, M., OFFERMANS, P. & ZAMUNER, T. S. 2004. Rhymes as a window into grammar. *Proceedings of the 29th Annual Boston University Conference on Language Development*, pp. 204-215. Somerville, MA: Cascadilla Press.

FIRTH, J. R. 1957. *Papers in Linguistics 1934-1951*. London: Oxford University Press

FLETCHER, H. 1972. *Speech and hearing in communication*. Huntington, NY: Robert E. Krieger.

FRISCH, S. A. 2004. Language processing and segmental OCP effects. In HAYES, B., KIRCHNER, R., & STERIADE, D. (eds.), *Phonetically-Based Phonology*. Cambridge, MA: Cambridge University Press. 374-371.

FRISCH, S. A., & ZAWAYDEH, B. 2001. The psychological reality of OCP-place in Arabic. *Language* 77. 91-106.

GAHL, S. & YU, A. C. L. 2006. Introduction to the special issue on exemplar-based models in linguistics. *The Linguistic Review* 23. 213-216.

GIBSON, L. F. 1956. Pame (Otomi) phonemics and morphophonemics. *International Journal of American Linguistics* 22. 242-265.

GNERRE, M. 1976. Le velocità di pronuncia e le loro implicazioni per la teoria fonologica generativa. In SIMONE, R., VIGNUZZI, U. & RUGGIERO, G. (eds.), *Studi di fonetica e fonologia. Atti del Congresso internazionale della Società di Linguistica Italiana - Padua, 1-2 October 1973*, pp. 285-296. Rome: Bulzoni.

GODDARD, I. 2001. The Algonquian languages of the Plains. In MALLIE, R. J. DE (ed.), *Handbook of North American Indians vol. 13 (Plains)*, pp. 71-77. Washington D.C.: Smithsonian Institution.

GOLDINGER, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105. 251-279.

GOLDINGER, S. D. 2007. A complementary-systems approach to abstract and episodic speech perception. *Proceedings of the 16th International Congress of Phonetic Sciences*, pp. 49-54. Saarbrücken.

GOLDINGER, S. D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22. 1166-1183.

GOLDINGER, S. D., PISONI, D. B. & LOGAN, J. S. 1991. On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory and Cognition* 17. 152-162.

GOLDSMITH, J. A. 1974. An autosegmental typology of tone: And how Japanese fits in. In KAISSE, E. & HANKAMER, J. (eds.), *Proceedings of NELS 5*, pp. 172-182. Cambridge, MA: Harvard University Linguistics Department.

GORDON, M., BARTHMAIER, P. & SANDS, K. 2002. A cross-linguistic acoustic study of fricatives. *Journal of the International Phonetic Association* 32. 141-174.

GREENBERG, J. H., 1978. Some generalizations concerning initial and final consonant clusters. In GREENBERG, J. H. (ed.), *Universals of Human Language vol. 2: Phonology*, pp. 243-279. Stanford, CA: Stanford University Press.

GREENLEE, M. & OHALA, J. J., 1980. Phonetically motivated parallels between child phonology and historical sound change. *Language Sciences* 2. 283-308.

GRØNNUM, N. 2006. DanPASS: A Danish Phonetically Annotated Spontaneous Speech Corpus. In CALZOLARI N., CHOUKRI, K., GANGEMI, A., MAEGAARD, B., MARIANI, J., ODIJK, J. & TAPIAS, D. (eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation*, Genova, Italy, 24-26 May.

GRØNNUM, N. & TØNDERING, J. 2007. Question intonation in non-scripted Danish dialogues. *Proceedings of the XVIth International Congress of Phonetic Sciences 2007*, pp. 1229-1232. Saarbrücken: Saarland University.

GUSSMAN, E. 2007. *The Phonology of Polish.* Oxford: Oxford University Press.

HALE, M. & REISS, C. 2000. Phonology as cognition. In BURTON-ROBERTS, N., CARR, P. & DOCHERTY, G. (eds.), *Phonological Knowledge. Conceptual and Empirical Issues*, pp. 161-184. Oxford: Oxford University Press.

HALE, M. & REISS, C. 2003. The Subset Principle in phonology: why the *tabula* can't be *rasa*. *Journal of Linguistics* 39. 219-244.

HALE, M. & REISS, C. 2008. *The Phonological Enterprise*. Oxford: Oxford University Press.

HALE, M., REISS, C. & KISSOCK, M. J. 2007. Microvariation, variation, and the features of Universal Grammar. *Lingua* 117. 645-665.

HALL, T. A. & HAMANN, S. 2010. On the cross-linguistic avoidance of rhotic plus high front vocoid sequences. *Lingua* 120: 1821-1844.

HALL, T. A. 1992. Syllable Structure and Syllable-Related Processes in German. *Linguistische Arbeiten* 276.

HALLE, M. 1959. *The Sound Pattern of Russian.* The Hague: Mouton.

HALLE, M., VAUX, B., WOLFE, A., 2000. On Feature Spreading and the Representation of Place of Articulation. *Linguistic Inquiry* 31:3. 387-444.

HANIQUE, I. & ERNESTUS, M. 2012. The processes underlying two frequent casual speech phenomena in Dutch: A production experiment. *Proceedings of INTERSPEECH 2012: 13th Annual Conference of the International Speech Communication Association*, 2011-2014.

HARAGUCHI, S. 1984. Some tonal and segmental effects of vowel height in Japanese. In ARONOFF, M. & ORHRLE, R. T. (eds.), *Language sound structure: studies in phonology presented to Morris Halle by his teacher and students*, pp. 145-156. Cambridge, MA: MIT Press.

HARRIS, JAMES 1983. *Syllable Structure and Stress in Spanish*. Cambridge, MA: MIT Press.

HARRIS, JOHN 1990. Segmental complexity and phonological government. *Phonology* 7. 255-300.

HARRIS, JOHN 1994. *English Sound Structure*. Oxford: Blackwell.

HARRIS, JOHN 1997. Licensing inheritance: An integrated theory of neutralisation. *Phonology* 14. 315-370.

HARRIS, JOHN 2006. The phonology of being understood: Further arguments against sonority. *Lingua* 116. 1483-1494.

HARRIS, K. S., 1958. Cues for the Discrimination of American English Fricatives in Spoken Syllables. *Language & Speech* 1:1. 1-7.

HASPELMATH, M. 1993. *A Grammar of Lezgian*. Berlin: Mouton de Gruyter.

HAWKINS, S. & SLATER, A. 1994. Spread of CV and V-to-V coarticulation in British English: Implications for the intelligibility of synthetic speech. *Proceedings of the 1994 International Conference on Spoken Language Processing,* pp. 57-60. Yokohama.

HAWKINS, S. & SMITH, R. 2001. Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics/Rivista di Linguistica* 13. 99-188.

HAY, J., JANNEDY, S. & MENDOZA-DENTON, N. 1999. Oprah and /ay/: Lexical frequency, referee design and style. *Proceedings of the 14th International Congress of Phonetic Sciences*, pp. 1389-1392. Berkeley University, California.

HAYES, B. 1986. Assimilation as spreading in Toba Batak. *Linguistic Inquiry* 17. 467-500.

HEID, S. & HAWKINS, S. 2000. An acoustical study of long-domain /r/ and /l/ coarticulation. *Proceedings of the 5th Seminar on Speech Production: Models and Data* (ISCA), pp. 77-80. Kloster Seeon, Bavaria, Germany.

HENDERSON, E. J. A., 1976. Khasi Initial Clusters. In JENNER, P. N., THOMPSON, L. C. & STAROSTA, S. (eds.), *Austroasiatic studies*, pp. 523-538. Honolulu: The University Press of Hawaii.

HENKE, E., KAISSE, E. M., & WRIGHT, R., 2012. Is the Sonority Sequencing Principle an epiphenomenon? In PARKER, S. (ed.), *The Sonority Controversy,* pp. 65-100. Berlin: Mouton de Gruyter.

HINSKENS, F. 2011. Lexicon, phonology and phonetics. Or: Rule-based and usage-based approaches to phonological variation. In SIEMUND, P. (ed.) *Linguistic Universals and Language Variation*, pp. 416-456. Berlin: Mouton de Gruyter.

HOCKETT, C. F. 1955. *A Manual of Phonology*. Baltimore, MD: Waverly Press.

HODGE, C. 1946. Phonemes of Serbo-Croatian. *Language* 22. 112-20.

HOFFMAN, C. 1963. *A Grammar of the Margi Language*. London: Oxford University Press for International African Institute.

HONGKAI, S., 1986. *Notes on Tibeto-Burman Consonant Clusters. Linguistics of the Tibeto-Burman Area.* Beijing: Chinese Academy of Seoul Sciences.

HUALDE, J. I. 1991. *Basque Phonology*. London: Routledge.

HUBER, D. 2009. On the interaction of velars and labials. *International Journal of Basque Linguistics and Philology* 41:2. 145-161.

HULST, H. VAN DER 1994. An introduction to Radical CV Phonology. In SHORE, S. & VILKUNA, M. (eds.), *SKY 1994: Yearbook of the linguistic association of Finland*, pp. 23-56. Helsinki.

HULST, H. VAN DER 1995. Radical CV Phonology: the categorial gesture. In DURAND, J. & KATAMBA, F. (eds.), *Frontiers of phonology,* 80-116. Essex: Longman.

HULST, H. VAN DER & WEIJER, J. VAN DER 1995. Vowel Harmony. In GOLDSMITH, J. A. (ed.), *The Handbook of Phonological Theory*, pp. 495-534. Oxford: Blackwell.

HUME, E. V. 1998. The role of perceptibility in consonant/consonant metathesis. In BLAKE, S., EUK-SOOK K. & SHAHIN, K. (eds.), *WCCFL XVII Proceedings*. Standford: CSLI.

HUME, E. V. 2011. *Markedness.* In OOSTENDORP, M. VAN, EWEN, C. J., HUME, E. V. & RICE, K. (eds.), *The Blackwell Companion to Phonology*, vol. I, pp. 79-106. Oxford: Blackwell.

HUME, E. V., JOHNSON, K., SEO, M. & TSERDANELIS, G., 1999. A cross-linguistic study of stop place perception. *Proceedings of the International Congress of Phonetic Sciences*, pp. 2069-2072.

INGRAM, D. 1974. Phonological rules in young children. *Journal of Child Language* 1. 29-64.

ITÔ, J. & MESTER, R. A. 1995. The core-periphery structure of the lexicon and constraints on reranking. In BECKMAN, J., WALSH DICKEY, L., & URBANCZYK, S. (eds.), *Papers in Optimality Theory*, pp. 181-209. University of Massachusetts Occasional Papers in Linguistics 18.

IVERSON, G. K. 1990. The stipulation of extraprosodicity in syllabic phonology. *Language Research* 26. 515-552.

JACQUES, G. 2013. The sound change **s- > n-* in Arapaho. *Folia Linguistica Historica* 34. 43-58.

JACQUES, G. 2011. A panchronic study of aspirated fricatives, with new evidence from Pumi. *Lingua* 121:9. 1518-1538.

JACQUES, G., 2004. The laterals in Tibetan. Talk presented at the *10th Himalayan Languages Symposium*, Thimbhu, Bhutan, 1 - 3 December 2004.

JAEGER, J. 1980. *Categorization in Phonology: An Experimental Approach.* PhD dissertation, University of California, Berkeley.

JAEGER, J. J. & VALIN R. D. VAN 1982. Initial Consonant Clusters in Yateé Zapotec. *International Journal of American Linguistics* 48:2. 125-138.

JAKOBSON, R. 1962 [1971-1985]. Phonological Studies. In Rudy, S. (ed.), *Selected Writings*, vol. I. The Hague: Mouton.

JAKOBSON, R. & HALLE, M. 1956. *Fundamentals of Language.* The Hague: Mouton.

JANSE, E., NOOTEBOOM, S. G. & QUENÉ, H. 2007. Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. *Language and Cognitive Processes* 22. 161-200.

JASKUŁA, K. 2013. Decadence of labials in Polish: phonetics or phonology? Talk presented at *Approaches to Phonology and Phonetics* (APAP), Lublin, 21-23 June 2013.

JOHNSON, K. 1997. Speech perception without speaker normalization: An exemplar model. In JOHNSON, K. & MULLENMIX, W. (eds.), *Talker Variability in Speech Processing*. New York: Academic Press. 143-165.

JOHNSON, K., FLEMMING, E. & WRIGHT, R. 1993. The Hyperspace Effect: Phonetic Targets are Hyperarticulated. *Language* 69:3. 505-528.

JUN, J. 1995. *Perceptual and articulatory factors in place assimilation: an Optimality Theoretic approach.* PhD dissertation, University of California, Los Angeles.

JUSCZYK, P. & ASLIN, R. N. 1995. Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology* 29:1. 1-23.

KAGER, R. 1999. *Optimality Theory*. Cambridge, MA: Cambridge University Press.

KAHN, D. 1976 [1980]. *Syllable-Based Generalizations in English Phonology.* PhD dissertation, MIT. New York, NY: Garland Press.

KAISSE, E. M. 1985. *Connected speech: the interaction of syntax and phonology*. Orlando, FL & London: Academic Press.

KAYE, J. D. & PÖCHTRAGER, M. 2009. GP 2.0. Talk presented at the *Government Phonology Roundtable*, 25 April 2009, Piliscsaba, Hungary.

KAYE, J. D. 1990. Government in phonology: The case of Moroccan Arabic. *The Linguistic Review* 6. 131-159.

KAYE, J. D. 2000. *A User's Guide to Government Phonology*. Unpublished manuscript, University of Ulster.

KAYE, J. D. 2005. GP, I'll have to put your flat feet on the ground. In BROEKHUIS, H., CORVER, N., HUYBREGTS, R., KLEINHENZ, U. & KOSTER, J. (eds.) *Organizing Grammar. Studies in Honor of Henk van Riemsdijk*, pp. 283-288. Berlin: Mouton de Gruyter.

KAYE, J., LOWENSTAMM J. & VERGNAUD, J.-R. 1985. The internal structure of phonological representations: a theory of Charm and Government. *Phonology Yearbook* 2. 305-328.

KAYE, J., LOWENSTAMM J. & VERGNAUD, J.-R. 1990. Constituent Structure and Government in Phonology. *Phonology* 7. 193-231.

KEATING, P. 1998. Word-level phonetic variation in large speech corpora. In ALEXIADOU, A. (ed.), *ZAS Papers in Linguistics* 11. 35-50.

KELLY, J. & LOCAL, J. K. 1986. Long-domain resonance patterns in English. *International Conference on Speech Input/Output; Techniques and Applications*, pp. 304-309. London, Institution of Electrical Engineers, Conference Publication n. 258.

KEMP, R. L. 2011. *The Perception of German Dorsal Fricatives by Native Speakers of English*. Master thesis, University of Georgia.

KEMPS, R., ERNESTUS, M., VAN HOUT, R. & BAAYEN, R. H. 2004. Processing reduced word forms: the suffix restoration effect. *Brain and Language* 90. 117-127.

KENSTOWICZ, M. 1994. *Phonology in Generative Grammar*. Cambridge, MA & Oxford: Blackwell.

KIRCHNER, R. 2001. *An Effort Based Approach to Consonant Lenition*. New York, NY & London: Routledge.

KLAUSENBURGER, J. 2012. French prosodics and phonotactics: an historical typology. *Beihefte zur Zeitschrift für romanische Philologie* 124.

KOESTERS GENSINI, S. E. 2000. Phonetic and morphological factors in spoken German. *Lingua e stile* 1:35. 61-77.

KOHLER, K. J. 1990. Segmental reduction in connected speech in German: phonological facts and phonetic explanations. In HARDCASTLE, W. J. & MARCHAL, A. (eds.), *Speech Production and Speech Modelling*, pp. 69-92. Dordrecht: Kluwer Academic Press.

KOHLER, K. J. 1995. *Einführung in die Phonetik des Deutschen* (2 volumes). Berlin: Erich Schmidt Verlag.

KOHLER, K. ET AL. 1995/1997. *The Kiel Corpus of Spontaneous Speech* (3 volumes). Institute of phonetics and digital speech processing, University of Kiel, Germany.

KOHLER, K. J. 1999. Articulatory prosodies in German reduced speech. *Proceedings of the XIV ICPhS* 1, pp. 89-92. San Francisco, CA.

LACY, P. DE 2006. *Markedness. Reduction and Preservation in Phonology.* Cambridge, MA: Cambridge University Press.

LADEFOGED, P. & MADDIESON, I. 1986. Some of the Sounds of the World's Languages (preliminary version). *UCLA Working Papers in Phonetics* 64. 1-137.

LADEFOGED, P. & MADDIESON, I. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

LADEFOGED, P. 1964. *A Phonetic Study of West African Languages*. London: Cambridge University Press.

LANDI, R. & SAVY, R. 1996. Durata vocalica, struttura sillabica e velocità di eloquio nel parlato connesso. *Proceedings of XXIV Convegno Nazionale dell'Associazione Italiana di Acustica*, pp. 65-70. Trento, Italy, 12-14 June 1996.

LAWRENCE, E. 1977. *Haida dictionary*. Fairbanks: Alaska Native Language Center.

LEBEN, W. 1973. *Suprasegmental Phonology.* PhD dissertation, MIT.

LENTZ, T. 2011. *Phonotactic Illegality and Probability in Speech Perception. Evidence from second language listeners.* Utrecht: LOT.

LI, P. J. 1985. A secret language in Taiwanese. *Journal of Chinese Linguistics* 13:1. 91-121.

LIBERMAN, M. 1975. *The Intonantional System of English*. PhD dissertation, MIT.

LINDAU, M. 1985. The story of r. In FROMKIN, V. A. (ed.), *Phonetic Linguistics*, pp. 157-168. Orlando, FL: Academic Press.

LINELL, P. 1979. *Psychological reality in phonology*. Cambridge, MA: Cambridge University Press.

LOCK, A. 2007. *Phonology Essentials – Abau Language*. Ukarumpa, Papua New Guinea: SIL International.

LOWENSTAMM, J. 1996. CV as the only syllable type. In DURAND, J. & LAKS, B. (eds.), *Current Trends in Phonology. Models and Methods*, pp. 419-441. Salford, Manchester: ESRI.

LOWENSTAMM, J. 1999. The beginning of the word. In RENISON, J. R. & KÜHNHAMMER, K. (eds.), *Phonologika 1996. Syllable !?*, pp. 153-166. The Hague: Thesus.

LYCHE, C. 1979. Glides in French: Questions for Natural Generative Phonology. *Lingua* 49. 315-330.

MACKAY, C. J. 1994. A Sketch of Misantla Totonac Phonology. *International Journal of American Linguistics,* 60:4. 369-419.

MACNEILAGE, P. F. & DAVIS, B. L. 2000. Origin of the internal structure of word forms. *Science* 288. 527-531.

MADDIESON, I. 1981. Unusual Consonant Clusters and Complex Segments in Eggon. *Studies in African Linguistics. Supplement* 8. 89-92.

MADDIESON, I. 1984. *Patterns of Sounds*. Cambridge, MA: Cambridge University Press.

MADDIESON, I. 2011. Consonant-Vowel Ratio. In DRYER, M. S. & HASPELMATH, M. (eds.), *The World Atlas of Language Structure Online,* chapter 3. Munich: Max Planck Digital Library. Available online at http://wals.info/chapter/3. Last accessed on 24 January 2014.

MALÉCOT, A. 1958. The role of releases in the identification of released final stops. *Language* 34. 370-380.

MARLETT, S. A. 1981. *The structure of Seri*. PhD dissertation, University of San Diego, California.

MARLETT, S. A. 1988. The Syllable Structure of Seri. *International Journal of American Linguistics*. 54:3. 245-278.

MARLETT, S. A. & PICKETT, V. 1987. The Syllable Structure and Aspect Morphology of Isthmus Zapotec. *International Journal of American Linguistics*. 53:4. 398-422.

MAROTTA, G. 2008. Lenition in Tuscan Italian (Gorgia Toscana). In BRANDÃO DE CARVALHO, J., SCHEER, T. & SÉGÉRAL, P. (eds.), *Lenition and Fortition*, pp. 235-272. Berlin: Mouton de Gruyter.

MATTYS, S. L. & LISS, J. M. 2008. On building models of spoken word recognition: when there is as much to learn from natural "oddities" as artificial normality. *Perception and Psychophysics* 70. 1235-1242.

MCCARTHY, J. J. 1979. On stress and syllabification. *Linguistic Inquiry* 10. 443-466.

MCLENNAN, C. T. & LUCE, P.-A. 2005. Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition* 31. 306-321.

MCLENNAN, C. T., LUCE, P.-A. & CHARLES-LUCE, J. 2003. Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory and Cognition* 29. 539-553.

MCMAHON, A. 2005. Heads I Win, Tails You Lose. In CARR, P., DURAND, J. & EWEN, C. J. (eds.), *Headhood, Elements, Specification and Contrastivity*, pp. 255-276. Amsterdam: John Benjamins.

MEINHOLD, G. 1973. *Deutsche Standardaussprache: Lautschwächungen und Formstufen*. Jena: Friedrich Schiller Universität.

MILLER, G. A. & NICELY, P. E. 1955. Analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27. 338-352.

MIONI, A. 1976. Fonologia generativa e sociolinguistica. In Simone, R., Vignuzzi, U. & Ruggiero, G. (eds.), *Studi di fonetica e fonologia*, pp. 297-312. Rome: Bulzoni.

MIONI, A. 2000. *Elementi di fonetica.* Padua: Unipress.

MIONI, A. (in preparation). *Lingue del mondo.*

MITTERER, H. 2000. Are connected-speech processes selected for communicative consequences? In BOER, B. DE & VOGT, P. (eds.), *Proceedings of the Benevolang Workshop 2000*, pp. 31-34. Brussels: Vrije Universiteit Brussel.

MITTERER, H. 2008. How are words reduced in spontaneous speech? In BOTINIS, A. (ed.), *Proceedings of the ISCA Tutorial and Research Workshop on Experimental Linguistics*, pp. 165-168. Athens: University of Athens.

MITTERER, H., YONEYAMA, K., & ERNESTUS, M. 2008. How we hear what is hardly there: Mechanisms underlying compensation for /t/-reduction in speech comprehension. *Journal of Memory and Language* 59: 133-152.

MITTERER, H. & ERNESTUS, M. 2006. Listeners recover /t/s that speakers reduce: evidence from /t/-lenition in Dutch. *Journal of Phonetics* 34. 73-103.

MOHANAN, K. P. 1993. Fields of attraction in phonology. In GOLDSMITH, J. A. (ed.), *The last phonological rule: reflections on constraints and derivations*, pp. 61-116. Chicago, IL: University of Chicago Press.

MORELLI, F. 1999. *The Phonotactics and Phonology of Obstruent Clusters in Optimality Theory*. PhD dissertation, University of Maryland at College Park.

NACASKUL, K. 1978. The Syllabic and Morphological Structure of Cambodian Words. *Mon-Khmer Studies* 7. 183-200.

NAGAMMA REDDY, K. 1987. Constraints on Consonant Sequences across some Indian Languages: a Typological View. *Osmania Papers in Linguistics* 13. 39-57.

NAMKUNG, J. 1996. Phonological Inventories of Tibeto-Burman Languages. *STED Monograph Series* 3.

NARAYAN, C. R., 2006. *Acoustic-perceptual salience and developmental speech perception*. PhD dissertation, University of Michigan.

NARAYAN, C. R., WERKER, J. F. & BEDDOR, P. S. 2010. The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science* 13:3. 407-420.

NESPOR, M. 1993. *Fonologia*. Bologna: Il Mulino.

NEWMARK, L. 1998. *Albanian-English Dictionary*. Oxford: Oxford University Press.

NIEBUHR, O. & KOHLER, K. J. 2011. Perception of phonetic detail in the identification of highly reduced words. *Journal of Phonetics* 39. 319-329.

NOCENTINI, A. & PARENTI, A. 2010. *L'etimologico. Vocabolario della lingua italiana*. Florence: Le Monnier.

OAKES, P. J. 2001. A Description of Teribe Phonology. *SIL Electronic Working Papers*.

OHALA, J. J. 1997. Emergent stops. *Proceedings of the 4th Seoul Conference on Linguistics (SICOL)*, pp. 84-91. 11-15 August 1997, Seoul.

OHALA, J. J., & KAWASAKI-FUKUMORI, H. 1997. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In ELIASSON, S.,

& JAHR, E. H. (eds.), *Language And Its Ecology: Essays In Memory Of Einar Haugen. Trends in Linguistics. Studies & Monographs* 100, pp. 343-365. Berlin: Mouton de Gruyter,

OOSTENDORP, M. VAN 1997. Style Levels in Conflict Resolution. In HINSKENS, F. L., VAN HOUT, R. & WETZELS, W. L. (eds.), *Variation, Change and Phonological Theory*, pp. 207-230. Amsterdam: John Benjamins.

OOSTENDORP, M. VAN 2013. Phonology between theory and data. Online paper (http://ling.auf.net/lingbuzz/001872), last accessed on 15 January 2014.

PARADIS, C. & BRUNET, J.-F., (eds.), 1991. *Phonetics and Phonology. The Special Status of Coronals – Internal and External Evidence 2*. San Diego, CA: Academic Press.

PARKER, S. 2012. *Onset consonant clusters database*. Unpublished manuscript. Dallas, TX: Graduate Institute of Applied Linguistics and SIL International.

PASSINO, D. 1999. *Phonological vs Morphological words: a case study on the domain of intervocalic /s/ voicing in Northern Italian*. MA thesis, University of L'Aquila, Italy.

PATE, J. K. & GOLDWATER, S. 2011. Predictability effects in adult-directed and infant-directed speech: Does the listener matter? *Proceedings of the 33rd annual meeting of the Cognitive Science Society.*

PATER, J. 2000. Nonuniformity in English stress: the role of ranked and lexically specific constraints. *Phonology* 17:2. 237-274.

PATER, J. 2010. Morpheme-Specific Phonology: Constraint Indexation and Inconsistency Resolution. In PARKER, S. (ed.), *Phonological Argumentation: Essays on Evidence and Motivation*, pp. 123-154. London: Equinox.

PENZL, H. 1955. *A Grammar of Pashto*. Washington D.C: American Counsil of Learned Societies.

PHARAO, N. 2009. *Consonant reduction in Copenhagen Danish. A study of linguistic and extra-linguistic factors in phonetic variation and change.* PhD dissertation, University of Copenhagen.

PIERREHUMBERT, J. B. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In BYBEE, J. & HOPPER, P. J. (eds.), *Frequency Effects and the Emergence of Linguistic Structure,* pp. 137-157. Amsterdam: John Benjamins.

PIERREHUMBERT, J. B. 2002. Word-specific phonetics. In GUSSENHOVEN, C. & WARNER, N. (eds.), *Laboratory phonology VII,* pp. 101-140. Berlin: Mouton de Gruyter.

PIERREHUMBERT, J. B. 2003. Probabilistic phonology: Discrimination and robustness. In BOD, R., HAY, J. & JANNEDY, S. (eds.), *Probabilistic Linguistics,* pp. 177-228. Cambridge, MA: MIT Press.

PIZZO, P. 2009. *A Typology of Sonority Sequences in Word-Final Consonant Clusters*. BA Thesis, Emory University, Atlanta.

PÖCHTRAGER, M. 2006. *The Structure of Length*. PhD dissertation, University of Vienna.

PÖCHTRAGER, M. 2012. Deconstructing A. Talk presented at the *MFM Fringe Meeting on Segmental Architecture,* 23 May 2012, Manchester, UK.

POLGÁRDI, K. 1998. *Vowel Harmony: An Account in Terms of Government and Optimality.* Utrecht: LOT.

PONS MOLL, C. 2008. The Sonority Scale: Categorical or Gradient? Poster presented at the *CUNY Conference on the Syllable*, 17-19 January 2008, New York, NY.

PRINCE, A. & SMOLENSKY, P. 1993/2004. *Optimality Theory: constraints* *interaction in generative grammar.* Malden, MA: Blackwell. Originally distributed as Rutgers University Center of Cognitive Science Technical Report #2.

REDDEN, J. E. 1965. *Walapai phonology and morphology*. PhD dissertation, Indiana University.

REGNIER, S. 1993. Quiegolani Zapotec Phonology. *Work Papers of the Summer Institute of Linguistics* 37, University of North Dakota.

ROOD, D. 1975. The Implications of Wichita Phonology. *Language* 51:2. 315-337.

ROSE, S. & WALKER, R. 2004. A Typology of Consonant Agreement as Correspondence. *Language* 80:3. 475-531.

RUBACH, J. 2000. Backness Switch in Russian. *Phonology* 17. 39-64.

RUMELHART, D. E., MCCLELLAND, J. L. & THE PDP RESEARCH GROUP (eds). 1986. *Parallel Distributed Processing: Exploration in the Micro-Structure of Cognition* (2 volumes). Cambridge, MA: MIT Press.

SAVY, R. & CUTUGNO, F. 1997. Hypospeech, vowel reduction, centralization, how do they interact in diaphasic variations? *Actes du XVIe Congrès International des Linguists*, pp. 1-13. Oxford: Pergamon-Elsevier.

SAVY, R. & CUTUGNO, F. 1998. Analisi acustica di vocali in monosillabi clitici in parlato spontaneo italiano. *XXVI Convegno Nazionale AIA*, pp. 301-306. 27-29 May 1998, Turin, Italy.

SAVY, R. & CUTUGNO, F. 2009. CLIPS. Diatopic, diamesic and diaphasic variations in spoken Italian. *Proceedings of the 5th Corpus Linguistics Conference*, 20th - 23rd July 2009, Liverpool, UK.

SCHANE, S. 2005. The aperture particle |a|: Its Roles and Functions. In CARR, P., DURAND, J. & EWEN, C. J. (eds.), *Headhood, Elements, Specification and Contrastivity*, pp. 313-338. Amsterdam: John Benjamins.

SCHEER, T. 1999. A theory of consonantal interaction. *Folia Linguistica* 32. 201-237.

SCHEER, T., 2004. *A Lateral Theory of Phonology. What is CVCV, and why should it be?* Berlin: Mouton de Gruyter.

SCHEER, T. 2011. Aspects of the development of generative phonology. In KULA, N. C., BOTMA, B., & NASUKAWA, K. (eds.), *The Continuum Companion to Phonology*, pp. 397-446. London & New York: Continuum.

SCHEER, T. 2012. *Direct Interface and One-Channel Translation. Vol. 2 of a Lateral Theory of Phonology*. Berlin: Mouton de Gruyter.

SCHEER, T. 2013. Spell-out, post-phonological. Talk presented at *Approaches to Phonology and Phonetics* (APAP), Lublin, 21-23 June 2013.

SCHUTTER, G. DE. 1994. Dutch. In KOENIG E. & J. VAN DER AUWERA (eds.), *The Germanic Languages*, pp. 439-477. London & New York, NY: Routledge.

SCHWARTZ, G. 2010. Rhythm and vowel quality in accents of English. *Research in Language* 8: 135-147.

SCRIBE (Spoken Corpus of British English). 1989/90. Available at http://www.phon.ucl.ac.uk/resource/scribe/. Last accessed on 24 January 2014.

SÉGÉRAL, P. & SCHEER, T. 2001. Le Coda-Miroir. *Bulletin de la Société de Linguistique de Paris* 96. 107-152.

SELKIRK, E. O. 1984. On the Major Class Features and Syllable Theory. In ARONOFF, M. & OEHRLE, R. (eds.), *Language Sound Structure,* pp. Cambridge, MA: MIT Press.

SHEPHERD, M. A. 2003. *Constraint interactions in Spanish phonotactics: An Optimality Theory analysis of syllable-level phenomena in the Spanish language.* MA thesis, California State University, Northridge.

SHOCKEY, L. 2003. *Sound Patterns of Spoken English.* Oxford: Blackwell.

SHOCKEY, L. & FARNETANI, E. 1992. EPG of spontaneous speech in English and Italian. *Proceedings of the International Conference of Spoken Language Processing*, pp. 667- 670. 12-16 November 1992, Banff, Alberta, Canada.

SILVERMAN, D. 2011. Usage-Based Phonology. In KULA, N. C., BOTMA, B., & NASUKAWA, K. (eds.), *The Continuum Companion to Phonology*, pp. 369-397. London & New York, NY: Continuum.

ŠIMÁČKOVA, S. 2009. Variable quality of the Czech lateral liquid: A perception experiment with young listeners. In KÜGLER, F., FERY, C. & VAN DE VIJVER, R. (eds.), *Variation and Gradience in Phonetics and Phonology,* pp. 125-139. Berlin: Mouton de Gruyter.

SIMONOVIĆ, M. 2013. Nepostojano ili nepostojeće?: Generativni pristup nepostojanom a u srbijanskom. Draft version, to appear in: ARSENIJEVIĆ, B. & HALUPKA, S. (eds.), *Srpski jezik u savremenoj lingvističkoj teoriji.*

SIPTÁR, P. 1991. Fast-speech processes in Hungarian. In GÓSY, M. (ed.), *Temporal Factors in Speech: a Collection of Papers,* pp. 27-62. Budapest: A Magyar Tudományos Akadémia Nyelvtudományi Intézete.

SKORUPPA, K., MANI, N. & PEPERKAMP, S. 2013. Toddlers' processing of phonological alternations: Early compensation for assimilation in English and French. *Child Development* 84:1. 313-330.

SMITH, R. & HAWKINS, S. 2012. Production and perception of speaker-specific phonetic detail at word boundaries. *Journal of Phonetics* 40:2. 213-233.

SMOLENSKY, P. 1987. Connectionist AI, symbolic AI, and the brain. *Artificial Intelligence Review* 1. 95-109.

SOMMERSTEIN, A. H. 1973. *The sound pattern of Ancient Greek*. Oxford: Blackwell.

STAMPE, D. 1969. The acquisition of phonetic representation. In BINNICK, R. (ed.), *Papers from the fifth regional meeting of the Chicago Linguistic Society*, pp. 443–454. Chicago: Chicago Linguistic Society.

STERIADE, D. 1982. *Greek Prosodies and the Nature of Syllabification*. PhD dissertation, MIT, Cambridge, Massachusetts.

STERIADE, D. 2001/2008. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In HANSON, K. & INKELAS, S. (eds.), *The nature of the word*, pp. 151-179. Cambridge, MA: MIT Press. Originally circulated in 2001.

STERNER, J. K. 1975. Sobei phonology. *Oceanic Linguistics* 14. 146-167.

SWANTON, J.R. 1910. *Haida, an Illustrative Sketch*. Bureau of American Phonology, Bullettin 40, pt I. 205-282.

SZIGETVÁRI, P. 1999. *VC Phonology: a theory of consonant lenition and phonotactics*. PhD dissertation, Eötvös Loránd University, Budapest.

SZIGETVÁRI, P. 2002. Dismantling syllable structure. *Acta Linguistica Hungarica* 48. 155-181.

TANKEVICIUTE, M. & STRIMAITIENE, M. 1990. Initial Consonant Clusters in Prussian. *Baltistica* 26:2. 105-110.

TARPENT, M. L. 1989. *A Grammar of the Nisgha Language*. PhD dissertation, University of Victoria.

TÖRKENCZY, M. 2006. *The Phonotactics of Hungarian*. DSc dissertation, Eötvös Loránd University, Budapest.

TORRE, E. J. VAN DER 2003. *Dutch sonorants. The role of place of articulation in phonotactics.* Utrecht: LOT.

TORREIRA, F. & ERNESTUS, M. 2010. The Nijmegen corpus of casual Spanish. IN CALZOLARI, N., CHOUKRI, K., MAEGAARD, B., MARIANI J., ODIJK, J., PIPERIDIS, S. & TAPIAS, D. (eds.), *Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC 10),* pp. 2981-2985. Paris: European Language Resources Association (ELRA).

TORREIRA, F. & ERNESTUS, M. 2011a. Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology* 2.2. 331-353.

TORREIRA, F. & ERNESTUS, M. 2011b. Vowel elision in casual French: The case of vowel /e/ in the word c'était. *Journal of Phonetics* 39:1. 50-58.

TORREIRA, F. & ERNESTUS, M. 2012. Weakening of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish. *Phonetica* 69. 124-148.

TORREIRA, F., ADDA-DECKER, M. & ERNESTUS, M. 2010. The Nijmegen corpus of casual French. *Speech Communication* 52. 201-212.

TSUNODA, T., 2008. Sonority Hierarchy in Warrongo (Australia). *Gengo Kenkyu* 133. 147-161.

TUNLEY, A. 1999. *Coarticulatory Influences of Liquids on Vowels in English.* PhD dissertation, Cambridge University.

ULLRICH, J. 2008. *New Lakota Dictionary & Incorporating the Dakota Dialects of Santee-Sisseton and Yankton-Yanktonai.* Lakota Language Consortium.

UZELAC, S. 1971. *Phonological Constraints in Serbo-Croatian Syllable Margins and Markedness in Generative Phonology.* MA thesis, Simon Frazer University, Burnaby, Canada.

VAUX, B. 1997. The Phonology of Voiced Aspirates in the New Julfa Dialect of Armenian. In AWDE, N. (ed.), *Armenian Perspectives*. Richmond: Curzon Press.

VAUX, B. 2003. Syllabification in Armenian, universal grammar, and the lexicon. *Linguistic Inquiry* 34. 91-125.

VENNEMANN, T. 1988. *Preference Laws for Syllable Structure and the Explanation of Sound Change*. Berlin: Mouton de Gruyter.

VIHMAN, M. 1996. *Phonological Development: The Origins of Language in the Child*. Cambridge & Oxford: Blackwell.

VOGT, H. 1971. *Grammaire de la Langue Georgienne*. Oslo: Universitetsforlaget.

WANG, M. D. & BILGER R. C. 1973. Consonant confusions in noise: A study of perceptual features. *Journal of the Acoustical Society of America* 54. 1248-1266.

WARNER, N. 2011. Reduction. In OOSTENDORP, M. VAN, EWEN, C. J., HUME, E. V. & RICE, K. (eds.), *The Blackwell Companion to Phonology*, vol. III, pp. 1866-1890. Oxford: Blackwell.

WARREN, P. & MARSLEN-WILSON, W. 1987. Continuous uptake of acoustic cues in spoken word recognition. *Perception and Psychophysics* 41. 262-275.

WEBER, D. J. 1989. *A Grammar of Huallaga (Huánuco) Quecha*. University of California Publications in Linguistics 112. Berkeley, CA: University of California Press.

WEST, P. 1999. Perception of distributed coarticulatory properties of English /l/ and /r/. *Journal of Phonetics* 27. 405-426.

WINTERS, S. 2000. Turning phonology inside out, or testing the relative salience of audio and visual cues for place of articulation. *Ohio State University Working Papers in Linguistics* 53. 168-199.

WOLFF, H. 1948. Yuchi Phonemes and Morphemes, with Special Reference to Person Markers. *International Journal of American Linguistics* 14. 240-243.

WRIGHT, R. 1996. *Consonant Clusters and Cue Preservation in Tsou*. PhD dissertation, UCLA.

WRIGHT, R., & LADEFOGED, P. 1994. A phonetic study of Tsou. *UCLA Working Papers in Phonetics* 87. 67-89.

YIP, M. 1982. Reduplication and CV skeleta in Chinese secret languages. *Linguistic Inquiry* 13:4. 637-662.

ZSIGA, E. 2013. *The Sounds of Language. An Introduction to Phonetics and Phonology*. Chichester, UK: Wiley-Blackwell.

ZURAIQ, W. & ZHANG, J. 2006. Phonological Assimilation in Urban Jordanian Arabic. *Kansas Working Papers in Linguistics* 28. 36-64.

ZWICKY, A. 1972a. On casual speech. In PERANTEAU, P., LEVI, J. N. & PHARES, C. G. (eds.), *Papers from the 8th regional meeting Chicago Linguistic Society*, pp. 607-615. Chicago, IL: Chicago Linguistic Society.

ZWICKY A. 1972b. Note on a phonological hierarchy in English. In STOCKWELL R. & MACAULAY R. (eds.), *Linguistic Change and Generative Theory*, pp. 275-301. Bloomington: Indiana University Press.

APPENDIX A

Nonce words presented in the experiment.

Fillers: tafun, famur, sapul, parun, kafum, lavus, mafun, narun, rakum, laruʃ, ʃamuk, fatup, fulat, vuras, bamuʃ, laduf, kunap, rufat, lutak, kufam, sulan, túʃal, vupar.

Target words (1)	*Target words (2)*
ʃtanur	ʃtukar
tʃafur	tʃuram
staful	stapun
tsulan	tsalun
ftamur	ftunal
tfalun	tfulan
ʃpalur	ʃpular
pʃamun	pʃagun
spalun	spatur
psalur	psamul
fpanur	fpunar
pfuran	pfarun
fkarul	fkural
kfalur	kfutar
ʃkatun	ʃkutan
kʃanum	kʃunam
skarun	skapur
ksafum	ksufam

sfarup	sfurap
fsamul	fsumal
ʃfapul	ʃfupal
fʃatul	fʃutal
ptafun	ptufan
ktaful	ktufal
tparum	tpalur
tkalum	tkumal
pkanur	pkunar
kpamun	kpunam
mnavul	mnuval
nmarul	nmapur
rlabul	rlubal
lrafun	lrufan
vulatʃ	nupatʃ
valuʃt	napuʃt
ruvats	ravuts
varust	vurast
savutf	suvatf
sulaft	kaluft
punaʃp	panuʃp
nalupʃ	nulapʃ
ruvasp	ravusp
karups	rukaps
kutapf	tukapf
tukafp	kutafp

nurapt	narupt
vumatp	vamutp
duvakf	davukf
vudafk	vatufk
tapukʃ	patukʃ
tamuʃk	tumaʃk
rafuks	rufaks
rufask	rafusk
simafs	samifs
mitasf	matísf
rikaʃf	kariʃf
dimafʃ	damifʃ
furakt	rafíkt
vumatk	vimatk
tulakp	tilakp
lutapk	litapk
rutamn	tarimn
turanm	ratinm
vudarl	vidarl
tugalr	tikalr

APPENDIX B

Words extracted from CLIPS (given in alphabetical order).

Gender and number are always given for adjectives, except the ones whose masculine and feminine forms have the same ending (in that case only number is given). As for nouns, number is given only if plural. For each verb person, tense and mood are noted. Given the geographical differences in the pronunciation of intervocalic *s*, it is always noted as voiceless.

Word	*Transcription (SF)*	*Gloss*
Abbondante	/ab.bon.ˈdan.te/	'abundant-SG'
Accanto	/ak.ˈkan.to/	'next to'
Alta	/ˈal.ta/	'tall, high-FEM.SG'
Altezza	/al.ˈtet.tsa/	'height'
Alto	/ˈal.to/	'tall, high-MASC.SG'
Altra	/ˈal.tra/	'other-FEM.SG.'
Altro	/ˈal.tro/	'other-MASC.SG'
Ampia	/ˈam.pja/	'wide-FEM.SG.'
Anche	/ˈaŋ.ke/	'also, too'
Ancora	/aŋ.ˈkoː.ra/	'again, still'
Anteriore	/an.te.ˈrjoː.re/	'front-SG'
Apertura	/a.per.ˈtuː.ra/	'opening'
Apposta	/ap.ˈpɔs.ta/	'on purpose'
Appunto	/ap.ˈpun.to/	'precisely, exactly'
Ascolta	/as.ˈkol.ta/	'listen-2SG.IMPER '

Aspetta	/as.ˈpɛt.ta/	'wait-2SG.IMPER'
Attorno	/at.ˈtor.no/	'around'
Attraverso	/at.tra.ˈvɛr.so/	'through, across'
Avanti	/a.ˈvan.ti/	'forward'
Avvallamento	/av.ˌval.la.ˈmen.to/	'subsidence'
Basso	/ˈbas.so/	'short, low-SG'
Basta	/ˈbas.ta/	'suffice-3SG.PRES.IND'
Bicicletta	/ˌbi.tʃi.ˈklet.ta/	'bicycle'
Càlcolati	/ˈkal.ko.la.ti/	'count-2SG.IMPER'
Camion	/ˈkaː.mjon/	'truck'
Canale	/ka.ˈnaː.le/	'canal'
Cancello	/kan.ˈtʃɛl.lo/	'erase-1SG.PRES.IND'
Capito	/ka.ˈpiː.to/	'understood'
Carla	/ˈkar.la/	given name
Caro	/ˈkaː.ro/	'dear-MASC.SG'
Cartina	/kar.ˈtiː.na/	'map'
Caso	/ˈkaː.so/	'case, chance'
Cavallo	/ka.ˈval.lo/	'horse'
Celeste	/tʃe.ˈlɛs.te/	'turquoise blue-SG'
Centimetri	/tʃen.ˈti.me.tri/	'centimeter-PL'
Centro	/ˈtʃɛn.tro/	'center'
Cerchio	/ˈtʃer.kjo/	'circle'
Certo	/ˈtʃɛr.to/	'certain-MASC.SG'
Chiuderla	/ˈkju.der.la/	'to close it-FEM.SG'
Cinque	/ˈtʃiŋ.kwe/	'five'
Circondato	/tʃir.kon.ˈdaː.to/	'surrounded-MASC.SG'

Circondi	/tʃir.ˈkon.di/	'surround-2SG.PRES.IND'
Circonferenza	/tʃir.ˌkoɱ.fe.ˈrɛn.tsa/	'circumference'
Circoscritta	/ˌtʃir.kos.ˈkrit.ta/	'circumscribed-FEM.SG'
Circoscrivi	/ˌtʃir.kos.ˈkriː.vi/	'circumscribe-2SG.PRES.IND'
Circoscrivo	/ˌtʃir.kos.ˈkriː.vo/	'circumscribe-1SG.PRES.IND'
Comprendere	/kom.ˈprɛn.de.re/	'to comprise'
Compresa	/kom.ˈpreː.sa/	'comprised-FEM.SG'
Comunque	/ko.ˈmuŋ.kwe/	'anyway'
Confine	/koɱ.ˈfiː.ne/	'boundary'
Contengono	/kon.ˈtɛŋ.ɡo.no/	'contain-3PL.PRES.IND'
Continua	/kon.ˈtiː.nwa/	'continue-2SG.IMPER′
Continue	/kon.ˈtiː.nwe/	'continuous-FEM.PL'
Contrario	/kon.ˈtraː.rjo/	'contrary-MASC.SG'
Corrispondenza	/ˌkor.ris.pon.ˈdɛn.tsa/	'correspondence'
Costeggia	/kos.ˈted.dʒa/	'skirt-3SG.PRES.IND'
Costellazione	/kos.ˌtel.lat.ˈtsjoː.ne/	'constellation'
Costituisce	/kos.ti.tu.ˈiʃ.ʃe/	'constitute-3SG.PRES.IND'
Costituiscono	/kos.ti.tu.ˈis.ko.no/	'constitute-3PL.PRES.IND'
Costituita	/kos.ti.tu.ˈiː.ta/	'constituted-FEM.SG'
Creare	/kre.ˈaː.re/	'to create'
Credo	/ˈkrɛː.do/	'believe-1SG.PRES.IND'
Crei	/krɛi/	'create-2SG.PRES.IND'
Crescente	/kreʃ.ˈʃɛn.te/	'growing-SG'
Croce	/ˈkroː.tʃe/	'cross'
Del	/del/	'of the-masc.SG'
Dentro	/ˈden.tro/	'inside'

Destra	/ˈdɛs.tra/	'right (direction)'
Dietro	/ˈdjɛː.tro/	'behind'
Dirlo	/ˈdir.lo/	'to say it-MASC.SG'
Distante	/dis.ˈtan.te/	'far, distant-SG'
Distanza	/dis.ˈtan.tsa/	'distance'
Diversi	/di.ˈvɛr.si/	'different-MASC.PL'
Dovresti	/do.ˈvres.ti/	'should-2SG.PRES.COND.'
Drasticamente	/ˌdras.ti.ka.ˈmen.te/	'drastically'
Dunque	/ˈduŋ.kwe/	'therefore'
Esattamente	/e.ˌsat.ta.ˈmen.te/	'exactly'
Est	/ɛst/	'East'
Esterna	/es.ˈtɛr.na/	'exterior-FEM.SG'
Esterno	/es.ˈtɛr.no/	'exterior-MASC.SG'
Estremi	/es.ˈtrɛː.mi/	'extreme-MASC.PL'
Estremità	/es.tre.mi.ˈta/	'end'
Faccia	/ˈfat.tʃa/	'face'
Facciate	/fat.ˈtʃaː.te/	'facade-PL'
Facendo	/fa.ˈtʃɛn.do/	'to do, to make-GERUND'
Fai	/fai̯/	'to do, to make-2SG.PRES.IND'
Fantastico	/fan.ˈtas.ti.ko/	'amazing-MASC.SG'
Fare	/ˈfaː.re/	'to do, to make'
Farfalla	/far.ˈfal.la/	'butterfly'
Fari	/ˈfaː.ri/	'headlight-PL.'
Farlo	/ˈfar.lo/	'to do it-MASC.SG'
Farne	/ˈfar.ne/	'to do (something) of it/them'
Fatti	/ˈfat.ti/	'made, done-MASC.PL'

Fatto	/ˈfat.to/	'made, done-MASC.SG'
Fermi	/ˈfer.mi/	'stop-2SG.PRES.IND'
Fermo	/ˈfer.mo/	'stop-1SG.PRES.IND'
Fianco	/'fjaŋ.ko/	'side'
Formano	/ˈfor.ma.no/	'form-3PL.PRES.IND'
Forse	/ˈfor.se/	'maybe, perhaps'
Fra	/fra/	'between, among'
Fregato	/fre.ˈɡaː.to/	'tricked-MASC.SG'
Fronte	/ˈfron.te/	'(in) front'
Giusta	/ˈdʒus.ta/	'correct-FEM.SG'
Giusto	/ˈdʒus.to/	'correct-MASC.SG'
Guarnizioni	/ɡwar.nit.ˈtsjoː.ni/	'topping-PL'
Ics	/iks/	'letter X'
Importa	/im.ˈpɔr.ta/	'matter-3SG.PRES.IND'
Inclini	/iŋ.ˈkliː.ni/	'incline-2SG.PRES.IND'
Indietro	/in.ˈdjɛː.tro/	'behind'
Infatti	in.ˈfat.ti/	'as a matter of fact'
Insieme	/in.ˈsjɛː.me/	'together'
Insomma	/in.ˈsom.ma/	'in short'
Interessa	/in.te.ˈrɛs.sa/	'interest-3SG.PRES.IND'
Interno	/in.ˈtɛr.no/	'internal'
Interrompi	/in.ter.ˈrom.pi/	'interrupt-2SG.PRES.IND'
Lasci	/ˈlaʃ.ʃi/	'to leave-2SG.PRES.IND'
Lasciato	/laʃ.ˈʃaː.to/	'left-MASC.SG'
Lateralmente	/la.te.ral.ˈmen.te/	'laterally'
Leggermente	/led.dʒer.ˈmen.te/	'slightly'

Lentamente	/ˌlen.ta.ˈmen.te/	'slowly'
Lenti	/ˈlɛn.ti/	'lens-pl'
Linea	/ˈli.ne.a/	'line'
Lupo	/ˈluː.po/	‘wolf’
Macchina	/ˈmak.ki.na/	'car'
Macchinetta	/mak.ki.ˈnet.ta/	'small car'
Mantieni	/man.ˈtjɛː.ni/	'maintain-2SG.PRES.IND'
Mappa	/ˈmap.pa/	'map'
Margine	/'mar.dʒi.ne/	'margin'
Media	/ˈmɛː.dja/	'medium-FEM.SG'
Meno	/ˈmeː.no/	'minus'
Messe	/ˈmes.se/	'put-FEM.PL'
Messi	/ˈmes.si/	'put-MASC.PL'
Metà	/me.ˈta/	'half'
Metti	/ˈmet.ti/	'put-2SG.PRES.IND'
Metto	/ˈmet.to/	'put-1SG.PRES.IND'
Mezzo	/ˈmɛd.dzo/	'middle'
Modo	/ˈmɔː.do/	'way, method'
Molto	/ˈmol.to/	'much, very'
Movimento	/mo.vi.ˈmen.to/	'movement'
Neanche	/ne.ˈaŋ.ke/	'not even'
Nella	/ˈnel.la/	'in the-FEM.SG'
Nero	/ˈneː.ro/	'black-MASC.SG'
Nessuna	/nes.ˈsuː.na/	'none-FEM.SG'
Niente	/ˈnjɛn.te/	'nothing'
Nordest	/nor.ˈdɛst/	'North-East'

Nordovest	/nor.ˈdɔ.vest/	'North-West'
Nuovo	/ˈnwɔː.vo/	'new'
Opposto	/op.ˈpɔs.to/	'opposite-MASC.SG'
Orizzontale	/o.rid.dzon.ˈta.le/	'horizontal-MASC.SG'
Ovest	/ˈɔ.vest/	'West'
Paio	/ˈpaː.jo/	'pair'
Pallido	/ˈpal.li.do/	'pale-MASC.SG'
Pallina	/pal.ˈliː.na/	'little ball'
Palline	/pal.ˈliː.ne/	'little ball-PL'
Palloncino	/pal.lon.ˈtʃiː.no/	'balloon'
Parabrezza	/ˌpa.ra.ˈbret.tsa/	'windshield'
Parafango	/ˌpa.ra.ˈfaŋ.ɡo/	'fender'
Parallela	/ˌpa.ral.ˈlɛː.la/	'parallel-FEM.SG'
Parallelo	/ˌpa.ral.ˈlɛː.lo/	'parallel-MASC.SG'
Paraurti	/ˌpa.ra.ˈur.ti/	'bumper'
Pare	/ˈpaː.re/	‘seem-3SG.PRES.IND'
Parlato	/par.ˈlaː.to/	'spoken-MASC.SG'
Parolaccia	/pa.ro.ˈlat.tʃa/	'dirty word'
Parte	/ˈpar.te/	'part-sg or leave-3SG.PRES.IND'
Partenza	/par.ˈtɛn.tsa/	'departure'
Parti	/ˈpar.ti/	'part.pl or leave-2SG.PRES.IND'
Partiamo	/par.ˈtja.mo/	'leave-1PL.PRES.IND'
Partono	/ˈpar.to.no/	'leave-3PL.PRES.IND'
Passa	/ˈpas.sa/	'pass-3SG.PRES.IND'
Passando	/pas.ˈsan.do/	'pass-GERUND'
Passare	/pas.ˈsa.re/	'to pass'

Passava	/pas.ˈsa.va/	'pass-3SG.IMPERF.IND.'
Passi	/ˈpas.si/	'pass-2SG.PRES.IND'
Pazienza	/pat.ˈtsjɛn.tsa/	'patience'
Pensavo	/pen.ˈsaː.vo/	'think-1SG.IMPERF.IND.'
Perché	/per.ˈke/	'why, because'
Percorso	/per.ˈkor.so/	'path'
Perfetto	/per.ˈfet.to/	'perfect'
Per	/per/	'for'
Perpendicolare	/per.pen.di.ko.ˈlaː.re/	'perpendicular-SG'
Perpendicolarmente	/per.pen.di.ko.lar.ˈmen.te/	'perpendicularly'
Polso	/ˈpol.so/	'wrist'
Porta	/ˈpɔr.ta/	'door'
Posta	/ˈpos.ta/	'put-FEM.SG'
Posteriore	/pos.te.ˈrjoː.re/	'rear-SG'
Posteriori	/pos.te.ˈrjoː.ri/	'rear-PL'
Posto	/ˈpos.to/	'put-MASC.SG or place'
Pratica	/ˈpra.ti.ka/	'practice'
Praticamente	/ˌpra.ti.ka.ˈmen.te/	'practically'
Precisamente	/pre.ˌtʃi.sa.ˈmen.te/	'precisely'
Preciso	/pre.ˈtʃi.so/	'precise'
Prendi	/ˈprɛn.di/	'take-2SG.PRES.IND'
Presa	/ˈpreː.sa/	'taken-FEM.SG'
Presenta	/pre.ˈsɛn.ta/	'display-3SG.PRES.IND'
Presente	/pre.ˈsɛn.te/	'present-SG'
Pressapoco	/pres.sa.ˈpɔː.ko/	'roughly'
Prima	/ˈpriː.ma/	'before, earlier, first'

Prime	/ˈpriː.me/	'first-FEM.PL'
Problema	/pro.ˈblɛː.ma/	'problem'
Procedendo	/pro.tʃe.ˈdɛn.do/	'proceed-GERUND.'
Procedi	/pro.ˈtʃɛː.di/	'proceed-2SG.PRES.IND'
Proietta	/pro.ˈjɛt.ta/	'project-3SG.PRES.IND'
Proiezione	/pro.jet.ˈtsjoː.ne/	'projection'
Proprio	/ˈprɔː.prjo/	'precisely'
Prosegui	/pro.ˈsɛː.gwi/	'continue-2SG.PRES.IND'
Prospettiva	/pros.pet.ˈti.va/	'perspective'
Punta	/ˈpun.ta/	'tip'
Punti	/ˈpun.ti/	'dot-PL'
Punto	/ˈpun.to/	'dot'
Qualcosa	/kwal.ˈkɔː.sa/	'something'
Quanto	/ˈkwan.to/	'how much'
Quattro	/ˈkwat.tro/	'four'
Questa	/ˈkwɛs.ta/	'this-FEM.SG'
Questo	/ˈkwɛs.to/	'this-MASC.SG'
Ridiscendi	/ri.diʃ.ˈʃɛn.di/	'come down again-2SG.PRES.IND'
Riga	/ˈriː.ga/	'line'
Riprendi	/ri.ˈprɛn.di/	'take again-2SG.PRES.IND'
Riscendi	/riʃ.ˈʃɛn.di/	'come down again-2SG.PRES.IND'
Rispetto	/ris.ˈpɛt.to/	'with respect to'
Rispiega	/ris.ˈpjɛː.ga/	'explain again-2SG.IMPER'
Rispieghi	/ris.ˈpjɛː.gi/	'explain-2SG.PRES.IND'
Risulta	/ri.ˈsul.ta/	'result-3SG.PRES.IND'
Ritroverai	/ri.tro.ve.ˈrai/	'find again-2SG.FUT.IND'

Ritrovi	/ri.ˈtrɔː.vi/	'find again-2SG.PRES.IND'
Ritrovo	/ri.ˈtrɔː.vo/	'find again-1SG.PRES.IND'
Rivolto	/ri.ˈvɔl.to/	'addressed-MASC.SG'
Sagoma	/ˈsa.go.ma/	'silhouette'
Sale	/ˈsaː.le/	'rise-3SG.PRES.IND'
Sali	/ˈsaː.li/	'rise-2SG.PRES.IND'
Salire	/sa.ˈliː.re/	'to rise'
Salita	/sa.ˈliː.ta/	'climb'
Sarà	/sa.ˈra/	'be-3SG.FUT.IND'
Saranno	/sa.ˈran.no/	'be-3PL.FUT.IND'
Sarebbe	/sa.ˈreb.be/	'be-3SG.PRES.COND'
Scendendo	/ʃen.ˈdɛn.do/	'go down-GERUND'
Scendere	/ˈʃen.de.re/	'to go down'
Scendi	/ˈʃen.di/	'go down-2SG.PRES.IND'
Scesa	/ˈʃeː.sa/	'gone down-FEM.SG'
Scesi	/ˈʃeː.si/	'gone down-MASC.PL'
Sceso	/ˈʃeː.so/	'gone down-MASC.SG'
Schermo	/ˈskɛr.mo/	'screen'
Sci	/ʃi/	'ski'
Sconto	/ˈskon.to/	'discount'
Scusa	/ˈskuː.sa/	'sorry'
Scusami	/ˈsku.sa.mi/	'forgive-2SG.IMPER me'
Secante	/se.ˈkan.te/	'secant'
Semicerchio	/ˌse.mi.ˈtʃer.kjo/	'semicircle'
Semicirconferenza	/ˌse.mi.tʃir.koɱ.fe.ˈrɛn.tsa/	'semicircumference'
Semplice	/ˈsem.pli.tʃe/	'simple'

Sempre	/ˈsɛm.pre/	'always'
Senso	/'sɛn.so/	'sense, direction'
Sentiero	/sen.tjɛː.ro/	'path'
Serpente	/ser.ˈpɛn.te/	'snake'
Sfera	/ˈsfɛː.ra/	'sphere'
Sferetta	/sfe.ˈret.ta/	'small sphere-SG'
Sferette	/sfe.ˈret.te/	'small sphere-PL'
Sfiorare	/sfjo.ˈraː.re/	'to touch lightly'
Sinistra	/si.ˈnis.tra/	'left (direction)'
Sopra	/ˈsoː.pra/	'above, on'
Sorta	/ˈsɔr.ta/	'sort'
Specchio	/ˈspɛk.kjo/	'mirror'
Specie	/ˈspɛ.tʃe/	'sort'
Sperando	/spe.ˈran.do/	'hope-GERUND'
Spero	/ˈspɛː.ro/	'hope-1SG.PRES.IND'
Spiegare	/spje.ˈɡaː.re/	'to explain'
Spiegati	/ˈspjɛ.ga.ti/	'explain-2SG.IMPER' yourself
Spieghi	/ˈspjɛː.ɡi/	'explain-2SG.PRES.IND'
Spiego	/ˈspjɛː.ɡo/	'explain-1SG.PRES.IND'
Spina	/ˈspiː.na/	'plug'
Spostata	/spos.ˈta.ta/	'moved-FEM.SG'
Spostato	/spos.ˈta.to/	'moved-MASC.SG'
Sposti	/ˈspɔs.ti/	'move-2SG.PRES.IND'
Sta	/sta/	'be, stay-3SG.PRES.IND'
Stai	/stai/	'be, stay-2SG.PRES.IND'
Stanno	/ˈstan.no/	'be, stay-3PL.PRES.IND'

Stare	/ˈstaː.re/	'to be, to stay'
Ste	/ste/	elided form of *queste* 'this-FEM.PL'
Stecche	/ˈstek.ke/	'temple-PL'
Stecchette	/stek.ˈket.te/	'small temple-PL'
Stella	/ˈstel.la/	'star'
Stelle	/ˈstel.le/	'star-PL'
Stellina	/stel.ˈliː.na/	'little star'
Stessa	/ˈstes.sa/	'same-FEM.SG'
Stessi	/ˈstes.si/	'same-MASC.PL'
Stesso	/ˈstes.so/	'same-MASC.SG'
Sto	/stɔ/	'be, stay-1SG.PRES.IND'
Stop	/stɔp/	'stop'
Storti	/ˈstɔr.ti/	'crooked-MASC.PL'
Storto	/ˈstɔr.to/	'crooked-MASC.SG'
Strada	/ˈstraː.da/	'road'
Stretto	/ˈstret.to/	'tight'
Stringere	/strin.ˈdʒe.re/	'to tighten'
Tangente	/tan.ˈdʒɛn.te/	'tangent'
Tangere	/ˈtan.dʒe.re/	'to touch'
Tantissimo	/tan.ˈtis.si.mo/	'very much'
Tanto	/ˈtan.to/	'much'
Termina	/ˈtɛr.mi.na/	'end-3SG.PRES.IND'
Terminato	/ter.mi.ˈnaː.to/	'ended-MASC.SG'
Tornante	/tor.ˈnan.te/	'hairpin bend'
Tornanti	/tor.ˈnan.ti/	'hairpin bend-PL'
Torni	/tor.ni/	'return-2SG.PRES.IND'

Torta	/ˈtor.ta/	'cake'
Tra	/tra/	'between, among'
Tracci	/ˈtrat.tʃi/	'draw-2SG.PRES.IND'
Tracciato	/trat.ˈtʃa.to/	'drawn'
Tragitto	/tra.ˈdʒit.to/	‘route’
Traiettoria	/tra.jet.ˈtɔː.rja/	'trajectory'
Tratto	/ˈtrat.to/	'line'
Tre	/tre/	'three'
Triangolo	/ˈtrjaŋ.ɡo.lo/	'triangle'
Trova	/ˈtrɔː.va/	'find-2SG.IMPER'
Trovato	/tro.ˈvaː.to/	'found'
Trovi	/ˈtrɔː.vi/	'find-2SG.PRES.IND'
Ultima	/ˈul.ti.ma/	'last-FEM.SG'
Ultimo	/ˈul.ti.mo/	'last-MASC.SG'
Verso	/ˈvɛr.so/	'towards'
Verticale	/ver.ti.ˈkaː.le/	'vertical-SG'
Vertice	/ˈvɛr.ti.tʃe/	'summit'
Visivamente	/vi.ˌsi.va.ˈmen.te/	'visually'
Visto	/ˈvis.to/	'seen-MASC.SG'
Volta	/ˈvɔl.ta/	'time'
Zampa	/ˈtsam.pa/	'paw'

LANGUAGE INDEX