PHONETICS AND PHONOLOGY
AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE

General Editor
E.F.K. KOERNER
Zentrum für Allgemeine Sprachwissenschaft, Typologie und Universalienforschung, Berlin
efk.koerner@rz.hu-berlin.de

Series IV – CURRENT ISSUES IN LINGUISTIC THEORY

Advisory Editorial Board

Lyle Campbell (Salt Lake City)
Sheila Embleton (Toronto)
Elly van Gelderen (Tempe, Ariz.)
Brian D. Joseph (Columbus, Ohio)
John E. Joseph (Edinburgh)
Manfred Krifka (Berlin)
Martin Maiden (Oxford)
E. Wyn Roberts (Vancouver, B.C.)
Joseph C. Salmons (Madison, Wis.)

Volume 306

Marina Vigário, Sónia Frota and M. João Freitas (eds.)

Phonetics and Phonology.
Interactions and interrelations
PHONETICS AND PHONOLOGY
INTERACTIONS AND INTERRELATIONS

Edited by
MARINA VIGÁRIO
SÓNIA FROTA
M. JOÃO FREITAS
University of Lisbon

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA
Library of Congress Cataloging-in-Publication Data
  p. cm. -- (Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory, ISSN 0304-0763 ; v. 306)
Includes bibliographical references and index.
1. Phonetics. I. Vigário, Marina Cláudia. II. Frota, Sónia. III. Freitas, Maria João.
P221.P4745    2009
414'.8--dc22    2009025637
ISBN 978 90 272 4822 0 (HB; alk. paper)
ISBN 978 90 272 8900 1 (EB)
© 2009 – John Benjamins B.V.
No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.
John Benjamins Publishing Co. • P.O.Box 36224 • 1020 ME Amsterdam • The Netherlands
John Benjamins North America • P.O.Box 27519 • Philadelphia PA 19118-0519 • USA
# Table of contents

Introduction

## Part I. Between phonetics and phonology

Schwa in American English V+/r/ sequences

*María Riera, Joaquín Romero & Ben Parrell*

Perception of word stress in Castilian Spanish: The effects of sentence intonation and vowel type

*Marta Ortega-Llebaria & Pilar Prieto*

Do complex pitch gestures induce syllable lengthening in Catalan and Spanish?

*Pilar Prieto & Marta Ortega-Llebaria*

Cues to contrastive focus in Romanian

*Alis Manolescu, Daniel Olson & Marta Ortega-Llebaria*

The phonetics of sentence-initial topic and focus in adult and child Dutch

*Aoju Chen*

## Part II. Segmental and prosodic interactions

Prosodic structure and consonant development across languages

*Timothy Arbisi-Kelm & Mary E. Beckman*

Rhythmic and prosodic contrast in Venetan and Sicilian Italian

*Laurence White, Elinor Payne & Sven L. Mattys*

Stem boundary and stress effects on syllabification in Spanish

*Teresa Cabré & Maria Ohannesian*

Prosodic and segmental effects on vowel intrusion duration in Spanish /r C/ clusters

*Benjamin Schmeiser*
PART III. Interactions between segments and features

Acoustic and aerodynamic factors in the interaction of features:
The case of nasality and voicing

*Maria Josep Solé*

205

Fixed and variable properties of the palatalization of dental stops in Brazilian Portuguese. In an Italian immigrant community

*Elisa Battisti & Ben Hermans*

235

Post-tonic vowel harmony in some dialects of Central Italy: The role of prosodic structure, contrast and consonants

*Stefano Canalis*

247

Vowel reduction and vowel harmony in Eastern Catalan loanword phonology

*Teresa Cabré*

267

Index of Subjects and Languages

287
Introduction

The thirteen papers in this book were selected from submissions by participants to the third Phonetics and Phonology in Iberia (PaPI) conference, hosted by the Universidade do Minho, Portugal, in June 2007. The main goal of the PaPI conferences is to bring together scholars interested in phonetics and phonology, with a special focus on the relationship and interactions between these two areas of research. The third edition of the conference was especially fruitful in addressing questions bearing on the issue of the status of phonetics-phonology connections. All the papers included in this volume are concerned with some of the multiple possible forms of interactions and interrelations in phonetics and phonology: the phonetic and/or phonological nature of speech patterns, segmental and prosodic interactions, and interactions between segments and features, both in child and in adult language, combining perception and production data, and doing so from theoretically as well as experimentally oriented perspectives.

It is by now well established that interactions within grammar are far from unlimited. Knowing what can interact in grammar and in what ways constitutes an important research program that has engaged a great number of linguists for the past few decades. The present volume Phonetics and Phonology. Interactions and interrelations is a novel contribution to this program. Since Keating’s (1988) description of the complex relation between phonological and phonetic descriptions of speech and related urge for cooperative research, and Beckman and Kingston’s (1990) foundations of a hybrid approach to research in phonetics and phonology known as Laboratory Phonology, and after two decades of the very successful Papers in Laboratory Phonology, the relation between the two fields is again the order of the day. Phonetics, commonly seen as the study of speech, and Phonology, traditionally concerned with sound patterning in language, largely share the same domain of inquiry. Therefore, interface explorations and interactions across the two fields are a natural ground for research. There is no consensus in the literature about what the division of labor between phonetics and phonology is, and the debate whether the two fields could be reduced to one (and which one) is open (see Ohala 1990 and Steriade 2000, a.o.). Further work on the phonetics-phonology connections is certainly required before any agreement or broader view can be reached. In this volume, the Phonetics and Phonology dualism is assumed: we take the linguistic prism, whereby phonetics, like phonology, is seen as a part of grammar that plays a role in the account of phonological patterns, and
experimental phonetics may deepen our understanding of such patterns. Thus, following a clear tendency in the field, this volume includes both contributions dealing with phonetic correlates of phonological categories and contributions focusing on the interplay of different aspects of sound patterning in language.

In their introduction to *Papers in Laboratory Phonology V*, Broe and Pierrehumbert (2000) judiciously highlight the gains of a research strategy that integrates phonetics and phonology, grounded in a multidisciplinary and multiframework approach to speech, which is open to all sorts of empirical data. In the same vein, but from outside the Laboratory Phonology view, Durand and Laks (2002) emphasize how phonological approaches divorced from phonetic substance may prove untenable, in much the same way as aspects of articulation are now known to be crucially related to aspects of perception, or different articulatory and acoustic dimensions may be critically dependent on one another.

A further fruitful domain for interactions, both between phonetics and phonology and within either of the fields, is how segments and subsegmental features are constrained by suprasegmental constituents or prosodies, and vice-versa (see Byrd, Kaun, Narayan & Saltzman 2000; Shattuck-Hufnagel 2006; inter alia). Not surprisingly, many of the themes and research questions of speech and language conferences over the last two years bear upon these issues, as it seems clear that general questions like the relevance of phonetic detail to grammar, the representation of phonetic variation within grammar, the interplay of gradual and/or categorical processes with the morphological and prosodic structures, or the extent of the dependency between the phonetics of features and processes and their form and function in language are yet to be fully understood (see Kingston 2007; and de Lacy 2008). The papers in the present collection bear upon these topics and address them in a wide array of languages, many of them underrepresented in the literature. Cantonese, Catalan, Greek, Brazilian Portuguese, Romanian, Spanish, several dialects from Italy (from the North, the Center and the South of Italy), Japanese, Dutch or English are among the languages covered. The present volume thus also contributes to enlarge the pool of languages for which new data now becomes available, a contribution that strengthens the empirical basis for cross-linguistic research as an avenue for understanding individual languages and language in general. The contributors come from the fields of phonology, experimental phonetics, acoustics and psycholinguistics, thus illustrating the commitment to interaction amongst researchers working on speech from various perspectives.

The volume has been organized into three parts: Part I – Between phonetics and phonology; Part II – Segmental and prosodic interactions; Part III – Interactions between segments and features. The papers in Part I deal with the relation between phonetics and phonology by either looking at the phonetic realization of phonological categories or evaluating the extent to which phonetic variation
and phonetic cues correlate with phonological contrast. Part II includes papers on
the interaction of segments with prosodic features, patterns or constituents. The
papers in Part III focus on the interplay between segments and features, namely
the dependency relations between features or between these and segmental processes
and their phonetic grounds.

Part I, *Between phonetics and phonology*, comprises five papers which address
the phonetic and phonological nature of segments, stress, pitch accents, and
pragmatic functions like topic and focus. All of them argue in favor of the laboratory
phonology approach, by exploring the use of production or perception experiments
to understand the issues investigated in both adult and child language.

Maria Riera, Joaquin Romero and Ben Parrell conduct an acoustic analysis of
V+/r/ sequences produced by native speakers of American English. The main goal
is to test the presence of a schwa-like element in the VC transitions and investigate
its phonetic and phonological status. The spectral and the durational properties of
this schwa-like element were registered and used as a means to assess the nature
of this element as a categorical entity or a product of the continuous nature of
speech production. The data gathered, namely the spectral values similar to those
of vowels preceding the schwa-like element and the increase in variability as an
effect of speaking rate, are seen to indicate that we are not in the presence of a
discrete segment (i.e. an output form of a phonological process of insertion or
epenthesis), but rather a product of a dynamic phonetic process of coarticulation.
Contra prior work that posits an epenthetic schwa and/or includes a schwa-like
element in the phonological representation of V+/r/ sequences, Riera and col-
leagues thus argue for an understanding of the phenomenon as an outcome of the
continuous nature of speech production that can be straightforwardly explained
from the point of view of articulation dynamics. By examining the nature and
properties of a segmental element in the speech string, this paper opens the dis-
cussion about the role played by phonetics and phonology in the characterization
of speech phenomena.

In the next three papers the discussion focus on prosodic phenomena and
information structure categories conveyed by prosodic means.

Marta Ortega-Llebaria and Pilar Prieto examine the way in which specific
phonetic cues, such as duration, spectral tilt, overall intensity, as well as vowel
quality affect the perception of stress contrast in Spanish in de-accented words.
They show that (i) the detection of stress is possible in the absence of pitch accents,
(ii) it is dependent on vowel type, and (iii) duration and overall intensity, but not
spectral tilt, are relevant phonetic cues to stress in Spanish. As the first study to
assess the cues to stress in Spanish without the confounding effect of pitch accent,
its results crucially bear on the issue of the phonetic nature of the stress contrast
across languages and across levels of prominence. Spanish has no phonological
vowel reduction, unlike English or Dutch, but as in English and Dutch, stressed vowels are produced with longer durations and louder intensities. More importantly, unlike in English and similarly to Dutch, Spanish listeners use these phonetic cues to perceive stress. Moreover, and contrary to Dutch, they weigh them according to vowel type. These results are seen as supporting the view that stress is not merely a structural device without phonetic content; stress has its own phonetic content, which varies across-languages according to aspects of the language-specific phonology. Among the languages that seem to use duration as a cue to stress in unaccented contexts, the authors list Dutch, German, Catalan, Spanish and also Romanian (according to data from Manolescu and colleagues, this volume). Ortega & Prieto's findings not only contribute to the understanding of the phonetics and phonology of stress across languages, but also across prominence levels by disentangling effects due to higher-level prominence, namely those expressed via sentence accentuation, from effects specific to the word-level stressed/unstressed contrast. A further development of the line of research pursued in their paper would be assessing whether stress is conveyed in similar ways both at the word-level and at lower-phrase levels (below the sentence level) in the absence of pitch accents. Ortega & Prieto have laid the groundwork for new research in the prominence hierarchy.

Focusing on the realization patterns of complex tonal categories, Pilar Prieto and Marta Ortega-Llebaria investigate the effects of complex pitch patterns on the durational properties of the syllable in Catalan and Spanish, in comparison with simple pitch movements. Results from a production experiment show that, in general, the presence of a complex pitch pattern tends to have a lengthening effect on the target syllable in both languages. However, complex contours may also be partially truncated, in which case no lengthening is triggered. It is thus concluded that syllable duration interplays with pitch accent realization, and that there is a trade-off between lengthening and truncation as two possible strategies for tone-bearing ability. These results have interesting implications for the cross-linguistic understanding of tonal compression and truncation and their status in intonation grammar. Contra to previous proposals that argued for compression and truncation as language-specific strategies, the Catalan and Spanish data clearly favor a view of truncation as a gradient phonetic effect that may be speaker-dependent and interacts with timing in a dynamic way. In other words, truncation is seen as a variable strategy of phonetic implementation, rather than a feature that characterizes the ways in which the intonation grammar of a given language or dialect is realized. These findings call for a detailed re-analysis of data from the so-called compression or truncation languages (or language-dialects) in search of patterns of variation across and within speakers similar to those found for Catalan and Spanish.
Alis Manolescu and colleagues inspect the phonetic correlates of phonological focus-marking in Romanian, thus fulfilling a gap in the intonation literature on focus in Romance. They find that pitch and duration are manipulated in well-defined ways in order to create a maximum contrast between the word under contrastive focus and the words in pre- and post-focal contexts. While F0 is expanded in the focused stretch and is flattened before and after it, segmental duration does not increase in the focused word. Nevertheless, a contrast is obtained in duration between focused and non-focused material via the shorter segmental durations of the latter. These findings add to the body of research on focus in Romance languages and make a positive contribution to the understanding of similarities and variation within Romance intonation systems. In particular, they speak to the dimension of pitch accent distribution, by showing that Romanian has a high density of pitch accents in broad focus utterances similar to most Romance languages, and that deaccenting is restricted and most probably a phonetic effect of pitch accent reduction in non-focal stretches, much like in other Romance languages and differently from English. They also speak to the issue of pitch accent categories and their pragmatic use/meaning, as Romanian, like many other Romance languages, also shows a contrast in pitch accent shape (more specifically in peak alignment) between broad and contrastive focus. Interestingly, in the case of Romanian this contrast may not be phonetically explained as an effect of a lengthening contrast, since contrastive focus does not yield increased segmental durations (unlike, for example, in Catalan and Spanish as reported by Prieto & Ortega in this volume, or in Dutch, as shown by Chen in this volume).

Aoju Chen’s paper also deals with the phonetic realization of information structure categories. Specifically, it investigates the acquisition of topic and focus marking by looking at the phonetic realization of Dutch sentence-initial topic and focus in adults and children. Prior work has shown that, in Dutch, the same phonological means is used to mark topic and focus, namely the same pitch accent type (H*L), unlike in Catalan, Spanish, or Romanian where focus is signaled by pitch accent choice (see, respectively, Prieto & Ortega, and Manolescu et al. this volume). The central question of Chen’s study is whether topic and focus are phonetically distinguishable by the use of gradient parameters, either F0 or duration-related, and whether children mark these categories in the same way as adults. It is found that adults clearly distinguish sentence-initial topic from sentence-initial focus by using specific gradient cues related to both F0 and duration. In contrast, 4- to 5-year-olds do not yet use any of the adult phonetic cues to distinguish these two information structural categories, but like adults they realize sentence-initial topic and focus most frequently with H*L. 7- to 8-year-olds only make use of F0 related, but not durational cues to focus and topic marking. Although children at the age of 7–8 are fully adult-like in their use of accent placement and accent type,
they are not fully adult-like in the phonetics of topic and focus marking. Chen thus concludes that the phonetic marking of topic and focus, being less salient than the use of discrete means like pitch accent type, is acquired later than the phonological marking of topic and focus. This paper makes two important claims about the phonetics and phonology of information structure categories and the way they are acquired: it shows that differences in phonetic realization are enough to distinguish information structure categories; however, it also shows that such differences are less salient than categorical differences and thus harder to acquire. These claims make the interesting prediction that the marking of information structure categories becomes adult-like earlier in languages where they are signaled by distinct phonological means, a prediction that waits empirical testing.

The four papers in Part II, *Segmental and prosodic interactions*, focus on prosodic factors and the ways they constrain segments. Using different frameworks, such as Optimality Theory or Articulatory Phonology, and different empirical bases, namely language development data, disfluent production data, standard adult speech, and a variety of acoustic measurements, they all argue for the relevance of prosody (prosodic structure, rhythm, stress) to the actual segmental outcome in speech.

Timothy Arbisi-Kelm and Mary Beckman elaborate on the idea that consonant development in first-language acquisition is related to the mastery of rhythmic structure. The argument is built on data from fluent speech, speech errors and disfluencies in the speech of stutterers. It is suggested that in actual speech production, metrical structures for larger prosodic constituents are put together before word-specific gestures and the latter are constrained by the former. In language development, the primacy of prosodic structure in constraining later word production is also argued for. A thorough review is provided of work suggesting how a hierarchy of language-specific metrical structures might emerge from a universal developmental progression of basic utterance rhythms in interaction with ambient language input. A cross-linguistic research project is presented, where recordings are analyzed from hundreds of children acquiring languages with different rhythmic properties, such as Cantonese, English, Greek, or Japanese. Several examples are offered of how patterns of disfluent consonant production differ across children acquiring the different languages, in ways that the authors suggest are related to the differences in metrical organization across languages. In this paper a new avenue in cross-linguistic research on how prosodic structure constrains segment production is proposed as a program to be pursued in the near future.

Intra-language rhythmic properties are examined by Laurence White, Elinor Payne and Sven Mattys, who explore the phonetic correlates of linguistic rhythm in two Italian varieties (Venetan, spoken in the North, and Sicilian, spoken in the South). In the last decade, several rhythm metrics were shown to capture
the traditional syllable-timed/stress-timed distinction quite successfully, as well as the more gradient rhythm contrasts within languages. The question remains, however, whether such metrics are able to capture all the temporal information that may be relevant to rhythm, and whether there may exist non-durational sources of information, such as segmental cues or intensity cues to stress that also contribute to the perception of rhythm (as is suggested, for example, in Ortega & Prieto, this volume). Varieties of Northern Italian are usually described as syllable-timed, whereas Southern Italian is usually reported to tend towards stress-timing. The goal of White and colleagues is to examine the relative contribution of durational and segmental factors to this perceived rhythmic difference between Northern and Southern Italian. They found no differences in the common rhythm metrics, with both varieties showing scores of the syllable-timed type. However, the results point to greater prosodic contrasts in Southern Italian, with greater vowel reduction (measured as the degree of centralization) according to stress and prosodic position, and greater vowel lengthening in nuclear utterance-final position. These findings suggest that multiple factors, both prosodic and segmental, contribute to perceived rhythm, and not all perceptually-salient rhythmic differences are captured by current rhythm metrics. The paper thus contributes to the ongoing debate on the acoustic measures that best account for perceived rhythmic differences among languages and dialects, and to the general discussion of whether differences in language rhythm are of a gradual or categorical nature.

The next two papers in this section inspect how prosodic constraints, together with other factors, may affect vowel production in different contexts in Spanish. Teresa Cabré and Maria Ohannesian examine the effect of morphological and prosodic factors on glide formation in vowel sequences of rising sonority in Spanish. The main goal is to compare first conjugation verbs whose stem ends in a high vowel with the lexically related nominal forms in the language, a goal that is pursued for the first time in the large literature on glide formation in Spanish. The formal tools of the Optimality Theory framework are used in order to account for the relation between prosodic and morphological factors blocking or promoting glide formation in the verbal and nominal paradigms. It is shown that the resulting hiatus/diphthong realization depends on the interplay of the role of stem boundaries, the theme vowel and several aspects of prosodic prominence, namely word stress, word-initial position and length of the word or stem. The analysis proposed not only accounts for the patterns of glide formation present in the language, but also explains why there is a gap in the system of possible patterns (specifically, why verbs with the diphthong realization and related nominals with the hiatus realization do not exist), and does so by relating the facts of glide formation with the processes of analogy in verbal environments in Spanish. It is well-known that the syllabification of vowel sequences with rising sonority differs among Romance
languages, and this paper makes a clear contribution to the general understanding of the phenomenon by highlighting the role played by prosodic prominence and prosodic position.

Within the Articulatory Phonology framework, Benjamin Schmeiser inspects the effect of prosodic and segmental variables on vowel intrusion duration in Spanish /rC/ sequences. Prior work has already shown that the intrusive vowel (a short, intervening vowel-like element) in Spanish is conditioned by both prosodic and segmental factors. However, previous studies have focused on /Cr/ clusters, and not on /rC/ sequences. The current study is intended to fill the existent gap in the research on vowel intrusion by measuring the effects on the durational variability of the intrusive vowel of various prosodic and segmental factors, such as word boundary and stress on the prosodic side and voicing, manner of articulation, order of constriction location and heterorganic/homorganic C2 on the segmental side. An acoustic analysis is conducted of data produced by Spanish speakers from countries as different as Spain, Mexico, Argentina, Guatemala, Colombia, and Ecuador. The results show that only one prosodic factor (across word boundary) and one segmental factor (order of constriction location) emerge as significantly affecting the duration of the intrusive vowel produced between the alveolar tap and the right adjacent C. These results contrast with those reported for /Cr/ clusters where only segmental factors play a role, thus suggesting that gestural overlap in heterosyllabic contexts differs from gestural overlap in tautosyllabic contexts. In addition, the fact that the intrusive vowel is sensitive to prosodic and segmental factors points to its phonetic nature as an articulatory transition, rather than an epenthetic vowel (in agreement with Riera and colleagues, this volume). Further, the results are taken to support the prosodic π-gestural model rooted in Articulatory Phonology (e.g. Byrd, Krivokapić & Lee 2006).

Part III, Interactions between segments and features, comprises four papers which address the relation between features and the role they play in segmental processes such as palatalization and vowel harmony. Phonetic detail and phonetic variation of different sources (individual, social, dialectal) are considered in the data, and the various authors take different stands on the best explanatory account of their findings: some argue in favor of a phonetically grounded account, others favor more phonologically grounded accounts along the lines of Feature Geometry or Optimality Theory approaches.

Maria-Josep Solé’s paper investigates the interaction between nasality and voicing, within and across segments. The goal is to account for the cross-linguistically common patterns of co-occurrence of these two features. The author elaborates on the effects of acoustic and aerodynamic factors in the interaction of features. She reviews typological, dialectal, language development and instrumental cross-linguistic data in order to evaluate dependency relations of the features referring to
nasality and voicing. Physical factors are shown to be responsible for the attested dependency; acoustic-auditory reasons are found to be at the origin of nasal segments being largely voiced. In contrast, abstract feature specifications devoid of detailed phonetic content and their current phonological representations fail to capture not only the likelihood of certain combinations of features in segments, but also how features may affect each other in contiguous segments, as in phonological processes or sound change. For example, the use of the nasal leakage to facilitate and preserve voicing in stops cannot be accounted for in phonological models where the nasal and the laryngeal features are in different branches. Solé argues that dependency relations between features due to aerodynamics, acoustics or perception are not explained in the formal accounts of Articulatory Phonology, Feature Geometry or the Optimality Theory, but are instead explained by phonetic theory. The central claim is that the sounds and sound sequence patterning in language can be largely accounted for by physical, physiological and auditory phonetic principles.

Unlike Sole’s approach, the last three papers assume a more phonological stand in their treatment of the interactions between segments and features. The chapter by Elisa Battisti and Ben Hermans examines fixed and variable properties of the palatalization of coronals in a dialect of Brazilian Portuguese spoken in Rio Grande do Sul. Although social factors may constrain the use of palatalization, the authors show that when this phonological process is activated, the process is linguistically conditioned and contextually predictable. The empirical data presented show that, in this dialect, high front vowels palatalize preceding dental stops. The analysis proposed is developed under the Optimality Theory framework and the Parallel Structures Model of feature geometry, by using a set of unranked constraints that generate the variation observed together with a hybrid representation of high vowels (which can be relatively consonantal or vocalic, depending on their distribution). The view that palatalization results from the requirement that the C(onsonantal) aperture feature of the high vowel be linked to a higher C node (the root node of the previous consonant) accounts for the empirical facts observed. The novelty of the proposal resides in its mixed approach that combines a phonological feature geometry-based representation of segments with a set of constraints that yield the variable palatalization outputs expressed by the quantitative data.

Stefano Canalis sets out to study the role of prosodic structure and of the feature architecture of consonants in the postonic vowel harmony found in some dialects of central Italy in proparoxytone words. The area where this harmony is attested includes Umbria, Marche, Northern Lazio and a part of Tuscany. In these dialects, postonic regressive vowel harmony copies all features of the word-final vowel to a preceding unstressed vowel, thus creating postonic sequences of identical vowels. The author offers phonetic and phonological arguments showing that
penultimate vowels of proparoxytone words are in a prosodically weak context, which makes them good targets for assimilation. In some dialects, vowel harmony is only active if the C node between the trigger and the target vowel is a liquid. Following Clements (2001), Canalis assumes that liquids are the only consonants without contrastive place features, thus explaining their transparency in the vowel harmony process in these dialects. The data under analysis is interesting and challenging, due to cross-linguistically uncommon features of the harmony process described, namely its direction and domain; and the account proposed assumes a relativized notion of universal feature geometry, thus adding to the debate on universal/language-specific feature geometry.

Finally, Teresa Cabré’s paper also deals with vowel harmony, but in a different language and in a specific part of the lexicon. The author explores the emergent vowel harmony processes affecting loanwords in Eastern Catalan (mostly imported from English and Spanish). The incorporation of non-native words in the language is generating a new phonology, by blocking the reduction of unstressed mid vowels to schwa and by allowing vowel harmony phenomena where mid vowels in stressed position are realized as close mid vowels as a result of the presence of close mid vowels in unstressed position. The paper specifically addresses the latter type of phenomena and investigates the spreading of the ATR feature across segments within the foot domain. Although dialectal and idiolectal factors affect the apparently random empirical facts reported, it is claimed that variation in the production of loanwords depends on the level of nativization (in the sense of Itô & Mester 1999). Furthermore, the insertion of a loanword in one of the specific levels identified depends on the quantity of phonological processes it has undergone. The account proposed is developed within the Optimality Theory framework and the theory of lexicon stratification. Crucially, the variation found is seen not as reflecting different parallel phonologies or as gradient phonetic variation, but as lexical strata organized into a core-periphery structure. As in Canalis’ chapter, the data under observation here are challenging, particularly due to their phonetic and phonological differences relative to the native vocabulary, but also due to the cross-linguistically uncommon direction of the harmony and trigger-target prominence relation; the analysis proposed by Cabré offers a straightforward account of the Catalan data. However, the question remains whether the interactions of features and segments in vowel harmony may be explained cross-linguistically in a principled way by means of similar phonological and/or phonetic principles, or by interactions among these.

We believe this volume is of interest to all researchers, teachers and students in the fields of phonetics and phonology, as well as to those interested in the interplay between production and perception, in the organization of grammar and language typology. In general, *Phonetics and Phonology. Interactions and interrelations* may
be a useful companion to all those wishing to widen and deepen their knowledge of the sound structure of language(s). Hopefully, it will contribute towards spreading further interest in this enticing field of inquiry between phonetics and phonology.

As a final word, we would like to thank all the contributors for their articles, and all the scholars who had agreed to review the papers submitted to this book, as well as the helpful advice received from one particular referee and the several anonymous reviewers’ comments forwarded by the CILT editor. Thanks are also due to the series editor, for his support and guidance. This work has been partially funded by grant POCTI-SFA-17-745, FCT, Portugal and by the Center of Linguistics of the University of Lisbon.

Lisbon, April 2009
Marina Vigário
Sónia Frota
Maria João Freitas

References


PART I

Between phonetics and phonology
Schwa in American English V+/r/ sequences*

María Riera, Joaquín Romero & Ben Parrell
Universitat Rovira i Virgili

This paper presents an acoustic study of word-final V+/r/ sequences in American English. The objectives were (i) to identify the presence of a schwa-like element in the VC transitions, (ii) to investigate how this presence is related to the phonetic/phonological nature of V, and (iii) to determine whether the spectral and durational characteristics of this element vary as a function of speaking rate. Two speakers participated in the experiment. Formant and duration measurements accounted for (i) differences between the schwa-like element and canonical schwa and (ii) variability in the schwa-like element and V. One-way ANOVAs tested for formant and duration differences, while two-way ANOVAs tested for the relationship between formant variability in the schwa-like element and V. The results suggest that the presence of a highly variable schwa-like element in the V+/r/ sequences is (i) a generalized process affecting all contexts and (ii) the result of coarticulation rather than epenthesis/insertion.

1. Introduction

In this study we investigate sequences of V+/r/ in American English, in particular the schwa-like element that is often perceived between the two segments in these sequences. We are interested in finding out whether the presence of this schwa-like element is due to a process of epenthesis/insertion or to one of coarticulation.

Trask (1996:132) defines epenthesis as a phonological process consisting of “the insertion of a segment into a word in a position in which no segment was previously present.” Coarticulation is defined by Hammarberg (1976:357) as “a process whereby the properties of a segment are altered due to the influences exerted on

*This research was funded by a doctoral grant from the Universitat Rovira i Virgili (Tarragona, Spain) and by projects 2005SGR00864 of the Generalitat de Catalunya and HUM2005-02746 of the Ministerio de Educación y Ciencia (Spain). The authors would like to thank Maria-Josep Solé and two anonymous reviewers for their very useful comments and suggestions on an earlier version of this paper.

1. Currently at the University of Southern California.
it by neighboring segments." Kühnert and Nolan (1999:7) state that coarticulation “refers to the fact that a phonological segment is not realized identically in all environments but often apparently varies to become more like an adjacent or nearby segment.” Finally, Ladefoged (2001:247) defines it as “the overlapping of adjacent articulatory gestures.” Coarticulation is considered a phonetic process, its phonological counterpart being known as assimilation (Hammarberg 1976, 1982; Roach et al. 2006), though the distinction and the relationship between the two remain an issue of debate among scholars. In fact, authors like Hammarberg (1976, 1982) and Kühnert and Nolan (1999) consider it necessary to take both the phonetic and the phonological levels into account when studying coarticulation. According to these authors, in any instance of coarticulation, behind the phonetic representation of the allophone (a physical tangible unit) there is always the phonological representation of the phoneme (a mental abstract unit).

Along these lines, a less clear-cut distinction between the phonetic and phonological levels can be made, however, as illustrated by the Articulatory Phonology approach (Browman & Goldstein 1986, 1989, 1990a, 1990b, 1992a, 1992b). Within this framework, articulatory gestures are simultaneously events of physical (traditionally phonetic) activity as well as units of linguistic (traditionally phonological) organization. Thus, coarticulation and assimilation could be viewed as instances of the same general process of gestural overlap and reduction. Language-dependent spatial and phasing relationships between gestures, as well as prosodic and contextual factors, would determine the extent to which a specific process remains at the level of colloquial, casual speech and is, therefore, under the control of the speaker, or rather, it becomes part of the lexical representation and is, supposedly, no longer affected by surface variability in aspects like speech rate. It is unclear, however, where the division between these two levels lies, as evidenced by much recent work on the phonetics/phonology interface. Thus, though for the sake of simplicity and convenience we will use the coarticulation vs. epenthesis/insertion dichotomy in this study, it is not our intention to imply a categorical distinction between the two or to present the study as a simple correlation between the phonetic and the phonological levels. Instead, we believe the two phenomena to be instances of the same generalized process of gestural conflict resolution, as an indication of the need to further our knowledge of the interaction between the two traditional levels.

In American English, a process of vowel neutralization takes place in the context of vowel sounds followed by /r/ (Avery & Ehrlich 1992; Giegerich 1992; Ladefoged 2001). The vowel contrast that is present in many other contexts between five pairs of vowel sounds (/i/ vs. /ɪ/, /ɛ/ vs. /ɛ/, /ɑ/ vs. /ɑ/, /o/ vs. /ɔ/ and /u/ vs. /ʊ/) disappears in this context, with the resulting combinations of vowel+ /r/ being /ɪr/, /ɛr/, /ɑr/, /ɔr/ and /ʊr/. Many authors (Avery & Ehrlich 1992; Baker & Goldstein 1990; Calvert 1986; Clark & Hillenbrand 2003; Dauer 1993; Giegerich 1992; Ladefoged 2001) claim that, in the existing V+/r/ combinations,
the exact realization of the vowel varies rather substantially between the two members of each contrasting pair. Thus, in the case of /ir/, the exact realization of the vowel could vary between /i/ and /ɜ/, based on speaker and/or dialectal differences. The same applies to the other pairs.

Vowels before /r/ are often called rhotacized, retroflexed or r-colored as a result of the influence exerted on them by the following /r/ (Avery & Ehrlich 1992; Clark & Yallop 1995; Davenport & Hannahs 1998; Ladefoged 2001; Olive et al. 1993; Rogers 2000). In these rhotacized vowels, the basic tongue configuration is retained, but the tongue tip is curled back in anticipation of the /r/ (Clark & Yallop 1995; Ladefoged 2001; Olive et al. 1993; Rogers 2000). The rhotacized effect may also be produced by keeping the tongue tip down and bunching the tongue body upwards towards the roof of the mouth (Ladefoged 2001; Olive et al. 1993). Ladefoged (2001:212) claims that “there may be in-between positions” to produce the rhotacized effect on the vowel and that in all cases “there is a slight narrowing of the pharyngeal cavity.” Rhotacized vowels before /r/ tend to show a considerable gradual and slow lowering of the third formant as a result of the influence of the low-frequency third formant of /r/ (Ladefoged 2001; Olive et al. 1993; Rogers 2000).

Sequences of V+/r/ do not occur syllable-finally in non-rhotic varieties of English. Instead, these varieties have the centering diphthongs /əʊ/, /eə/ and /ʊə/, as well as the centering triphthongs /əʊə/, /eɪə/, /ʊɪə/, /auə/ and /auə/. The development of the /ə/ in all these cases is understood as the result of the historic loss of the /r/ (Giegerich 1992; Ladefoged 2001; Laver 1994; Rogers 2000). However, the equivalence between the diphthongs and triphthongs of the non-rhotic varieties and the V+/r/ sequences of the rhotic ones is not straightforward. In a rhotic variety such as General American, even though /əʊ/, /eə/, /ʊə/, /auə/ and /auə/ have their equivalents in /ɪr/, /ɜr/, /ʊr/ and /ær/, the three triphthongs /eɪə/, /ʊɪə/ and /auə/ have no V+/r/ counterpart. According to standard descriptions (Kenyon & Knott 1953; Edwards 1997), diphthongs in General American are limited to /æɪ/, /oɪ/ and /au/, while /e/ and /o/ are considered diphthongized monophthongs; in addition, General American is described as having no triphthongs.

In his English pronunciation dictionary, Wells (2000) uses the superscript symbol /ə/ to show instances of schwa epenthesis/insertion (/rətr/, /eər/, /ʊər/, /aiər/ and /auər/, as in here, hair, poor, fire and power). He refers to these cases as examples of what he terms pre-r breaking. According to him, vowels and diphthongs undergo pre-r breaking as the result of the development of a glide sound /ə/ before /r/ and, consequently, become diphthongs and triphthongs, respectively. The diacritic /ə/ that is present in the transcriptions /aiər/ and /auər/ reflects the possibility of considering these words as composed of either one syllable or two. Similarly, Lavoie

---

2. Except in the case of linking /r/ (here is fine, far and away, poor or rich).
and Cohn (1999) state that monosyllables consisting of non-low tense pure vowels or diphthongs followed by a liquid, which they call sesquisyllables, can be pronounced with either one or two syllables. Finally, Roach et al. (2006:144) state that “opinions differ as to whether diphthongs should be treated as phonemes in their own right, or as combinations of two phonemes.”

In all the above-mentioned cases, Wells’s transcription system accounts for two possible pronunciations: one with schwa epenthesis/insertion and the other without it. According to him, the choice of one or the other is speaker-dependent as well as situation-dependent, with schwa epenthesis/insertion being more common in slow speech rates than in fast ones. As Wells (2000) points out, in his dictionary the superscript symbol /ə/ is also intended to help second and/or foreign language learners of English in their pronunciation. Other authors (Baker & Goldstein 1990; Calvert 1986; Celce-Murcia et al. 1996; Dauer 1993; Prator & Robinett 1985) also make a point of the inclusion of epenthetic schwa in the transcription of the V+/t/ sequences as a way for the second and/or foreign learner of English to associate the phonological representation of the sequences with how they are both produced and perceived.

Some authors within the American tradition make use of the term centering diphthong to refer to either the vowel in the V+/t/ sequences or the whole sequence. In their study, Clark and Hillenbrand (2003:1) define the combinations of V+/ə/ as diphthongs and choose “to use [ə] as the second part of the diphthong in question, based on the assumption that their first part is prominent and the second is an offglide to a relatively weaker centralized endpoint.” Dauer (1993) calls the V+/t/ sequences centering diphthongs and uses a transcription system that is consistent with the term. Her phonological transcriptions include epenthetic schwa in all cases. Van Riper and Smith (1992) also refer to these sequences as centering diphthongs, but they present the transcription with /ə/ only as a secondary option. Kenyon (1989) also uses the term centering diphthongs to name the V+/t/ sequences, but on no occasion does he suggest transcribing the sequences with epenthetic schwa, transcribing them instead with vowel+/t/. Roach et al. (2006) transcribe the centering diphthongs in here, hair and poor without epenthetic schwa and with vowel+/t/ as well, but they transcribe fire and power with final /ə/, and they do not consider the words far and pour to contain any centering diphthong. Prator and Robinett (1985) advocate for the term centering diphthongs as well, but their phonological representation of the sequences is with final /t/ in all cases and with epenthetic schwa included in /əə/ (here), /εə/ (hair), /aəə/ (fire) and

---

Schwa in American English V+/r/ sequences

In /auər/ (power), but not in /ɑr/ (far), /ɔr/ (pour) and /ʊr/ (poor). For them, the vocalic elements are the ones that make up the centering diphthong and the /r/ is not part of it.

Other authors (Baker & Goldstein 1990; Celce-Murcia et al. 1996) do not talk of centering diphthongs, but they suggest epenthetic schwa as an option in at least some cases. Baker and Goldstein (1990), for instance, do so with /ɪər/ (here), /ɛər/ (hair), /uər/ (poor), /ʌər/ (fire) and /uoər/ (power), not considering it possible in /ɑr/ (far) or /ɔr/ (pour). Celce-Murcia et al. (1996:104) associate epenthetic schwa directly only with the diphthongs /aɪər/ (fire) and /auər/ (power), but indirectly with at least one other case when they state that the conventional transcription of the V+/r/ sequence in beard as /ɪr/ “may not completely represent or capture the precise articulation of the /r/-colored vowel.”

The schwa-like element that is often perceived between the two elements of some VC and CC sequences has received different names by different authors. Warner et al. (2001), in their analysis of Dutch schwa in /l/+C clusters, refer to it as epenthetic schwa. Authors such as Baker and Goldstein (1990), Celce-Murcia (1996), Olive et al. (1993), Prator and Robinett (1985) and Wells (2000), despite not using the term epenthes, refer to the process as one of insertion, a term that is considered synonymous with epenthes. Hall (2003, 2006) makes a clear distinction between schwa intrusion and schwa epenthes/insertion, claiming that the schwa-like element in the CC clusters of her studies belongs to the former type. According to this author, intrusive vowels are phonologically invisible, are inserted late in the phonological derivation, cannot act like syllable nuclei, do not add a syllable to the word, and do not involve the addition of a vowel segment. Moreover, they are not likely to occur in the most marked types of CC clusters, tend to occur between heterorganic consonants, copy only over sonorants or gutturals and are either copy vowels or neutral and schwa-like in quality. Finally, they come in a restricted range of qualities, are often variable in duration and have a tendency to disappear in fast and/or casual speech. Schmeiser (2009, this volume) also provides justification for distinguishing between intrusive vowels and epenthetic vowels. Using data from Spanish, he claims that the result of diachronic phonologization of intrusive vowels is predictable, given the similarity in formant structure between the intrusive vowel and the nuclear vowel. Also, intrusive vowels are not specified lexically, as evidenced by syllable separation. In addition, the restrictions imposed by the three-syllable stress window of Spanish suggest the non-epenthetic nature of these vowels. Browman and Goldstein (1992b), in their analysis of CVCCVCV sequences, refer to this element as a targetless schwa, that is, a vocalic element with spectral and duration values similar to those that are often attributed to canonical schwa but, at the same time, somehow differ from them, mainly due to the influence exerted by neighboring vocalic segments on the vocalic element. Finally, Gick and Wilson (2001, 2006) call the schwa-like element in V+liquid
combinations excrescent schwa and consider it the result of the tongue movement required to pass through a schwa-like configuration on its way from the vowel to the /r/. Explanations along the same lines to account for the presence of a schwa-like element in American English V+/r/ sequences are provided by other authors like Calvert (1986), Prator and Robinett (1985) and Olive et al. (1993).

In a preliminary descriptive acoustic study, Riera and Romero (2006) investigated the presence of a schwa-like element in word-final V+/l/ and V+/r/ sequences in American English stressed monosyllables. The results showed that the VC transitions in these sequences are in some cases not easy to identify by means of visual spectrographic observation. The results also suggested a relationship between the phonological parameters that are used to classify vowels (tongue height, tongue advancement and the tense vs. lax distinction), and the clear presence in sequences containing high front tense vowels of a quite variable schwa-like element in terms of its duration and spectral characteristics. This study, however, did not provide acoustic measurements and was not able to account for speaking rate differences.

In a follow-up study, Riera and Romero (2007) looked at the phonetic and phonological nature of word-final V+/l/ sequences in American English stressed monosyllables as produced by only one speaker. The analysis involved acoustic (F1, F2 and duration) measurements. The results suggested the presence of a highly variable schwa-like element which differed from canonical schwa as a function of both the preceding vowel and speaking rate. This study concluded that the phenomenon under analysis was a generalized one affecting all V+/l/ contexts and that the presence of the schwa-like element was due to a process of coarticulation rather than to one of epenthesis/insertion.

Taking the results of previous studies into consideration, we designed an experimental acoustic study which expanded on work undertaken by Riera and Romero (2006, 2007) and partially replicated Riera and Romero’s (2007) study in an attempt to provide further evidence for the phonetic and phonological nature of word-final V+/r/ sequences in American English stressed monosyllables.

1.1 Objectives and hypotheses

The first of our main goals in designing the present study was three-fold, since our intentions were (i) to discover, by means of acoustic measurements, which of the V+/r/ sequences under study contain a distinguishable schwa-like element in their VC transitions, (ii) to see whether our findings are consistent with those of previous work, and (iii) to determine whether the presence of this schwa-like element can be extended to VC transitions containing vowels other than high, front and tense. We hypothesize that acoustic measurements will make it possible to account for the presence of a schwa-like element in sequences in which mere spectrographic observation renders it undetectable as well as in sequences not dealt with in previous
studies. If this is the case, we will find ourselves confronting a generalized process affecting all contexts, in line with the results obtained in Riera and Romero (2007), rather than a process affecting just some or a few contexts, as seems to be suggested by the focus and/or the results of some previous studies.

Our second main aim was to investigate the extent to which the presence of this schwa-like element is related to the phonetic/phonological nature of the preceding vowel and whether the spectral and durational characteristics of the schwa-like element vary as a function of speaking rate. Two hypotheses stem from this objective. First, we hypothesize that F1, F2, F3 and duration values for the schwa-like element in the VC transitions will differ from those of canonical schwa. F1, F2 and F3 will show greater variability in the schwa-like element than in canonical schwa. Also, duration will be shorter in the former than in the latter. This will be taken as evidence that this element cannot be considered an epenthetic schwa. Second, we hypothesize that F1, F2 and F3 values for the schwa-like element will vary considerably across the different contexts and also as a function of the preceding vowel, resembling more the values of this preceding vowel the faster the speaking rate. If one or both of these hypotheses are confirmed, we will be able to claim that any presence of a schwa-like element between the vowel and the consonant in the V+/r/ sequences under study, far from being the result of a discrete process of epenthesis/insertion, can be attributed to a dynamic process of coarticulation.

2. Method

2.1 Subjects

The subjects taking part in the experiment were two native speakers of the Midwestern variety of American English. Speaker 1 was a 23-year-old male who had been born and had lived most of his life in the Chicago, Illinois, area. He was a graduate student in phonetics and had some specialized phonetic training. Speaker 2 was a 50-year-old female who had lived in different parts of the United States but who spoke with a clearly defined Midwest accent. She had no specialized phonetic training. Both speakers were unaware of the purposes of the experiment at the time of the recording.

2.2 Stimuli and data collection

The stimuli that were selected for the experiment consisted of seven meaningful English monosyllables containing the sequences C₁VC₂. The VC₂ sequences consisted of /r/ preceded by each of the seven vowel sounds of American English that can appear before this consonant, namely, /ɪr/, /ɛr/, /ɜr/, /ɔr/, /ʊr/, /ʌr/, and /ær/. C₁ was one of /p/, /f/ and /h/. The choice of non-lingual (unlike /r/) and oral (like /r/)
consonants was considered most appropriate and was made for the purpose of minimizing \( C_1 \) coarticulatory influence on \( V \), and even on coda \( C_2 \). Seven additional monosyllables were chosen as distracters. In these, \( C_1 \) was /f/, /v/ or /h/, again non-lingual and oral consonants; \( V \) was once more one of the seven vowel sounds that can appear before /r/ in American English; and \( C_2 \) was either /t/ or /d/. Finally, the words \textit{arrive} and \textit{ahead}, both containing canonical schwa in their initial syllable, were also included, the former to serve as control and the latter as distractor. Table 1 shows the 16 stimuli (target words, control word and distractors) used in the experiment.

<table>
<thead>
<tr>
<th>Target words</th>
<th>Distractors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word</strong></td>
<td><strong>C₁VC₂</strong></td>
</tr>
<tr>
<td>here</td>
<td>/hir/</td>
</tr>
<tr>
<td>hair</td>
<td>/her/</td>
</tr>
<tr>
<td>far</td>
<td>/fər/</td>
</tr>
<tr>
<td>pour</td>
<td>/pɔːr/</td>
</tr>
<tr>
<td>poor</td>
<td>/pɔːr/</td>
</tr>
<tr>
<td>fire</td>
<td>/faɪr/</td>
</tr>
<tr>
<td>power</td>
<td>/ˈpəʊər/</td>
</tr>
<tr>
<td>arrive</td>
<td>/əˈrɑːv/</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Word</strong></th>
<th><strong>V₁C₁V₂C₂</strong></th>
<th><strong>V₁C₁</strong></th>
<th><strong>V₁</strong></th>
<th><strong>Word</strong></th>
<th><strong>V₁C₁V₂C₂</strong></th>
<th><strong>V₁C₁</strong></th>
<th><strong>V₁</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Control word</td>
<td>Distractor</td>
<td>Control word</td>
<td>Distractor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A total of 40 stimuli – the stimuli that were selected for this experiment plus the stimuli that were selected for the experiment reported in Riera and Romero (2007), which looked at \( V+1/ \) sequences – were inserted in the carrier sentence ‘Tell me ____ four times.’ The tokens were presented in a timed computer Power Point slide presentation for the speakers to read. Each speaker performed two readings of 12 randomized tokens of each sentence: first, at a slow rate and, second, at a fast rate. The speaking rate variable was controlled by presenting the tokens in the slow reading at three-second intervals, with a three-second break every 20 tokens, and the tokens in the fast reading at one-second intervals, with a three-second break every five tokens. Each speaker provided a total of 960 (40x12x2) tokens, 384 (16x12x2) of which were selected and used in the experiment reported here. The total involvement time for each speaker was about one hour and thirty minutes, including the instructions period, the two readings, a 15–20 minute break between the two readings, and two trial periods before each reading, of 20 tokens each, which were not used for the analysis.
The data were recorded at a 22,050 Hz sampling rate directly into an HP Pavilion dv4000 laptop computer using an M-Audio Nova condenser microphone, an M-Audio Firewire Solo mobile audio interface and the Praat (version 4.5) speech analysis software (Boersma & Weenink 2006). The latter was also used for the subsequent data processing and analysis.

2.3 Data processing and analysis

The procedure used for the data processing and analysis of the sequences under study replicates the one reported in Riera and Romero (2007). First, in order to account for general duration differences between the slow and fast productions, duration measurements were taken for the entire CVC sequence and the VC portion in the V+/r/ words, as well as for the entire word and the /ər/ portion in the control word arrive. These measurements, however, were not used in the final analysis, since this focused on the VC sequences under study only. Second, F1, F2 and F3 values were automatically calculated every five milliseconds for both the entire CVC sequence of the V+/r/ words and the /ər/ portion of the control word arrive. Third, the beginning and end of any VC transitions in the V+/r/ sequences and of the /ə/ in the /ər/ sequence in arrive were established on the basis of (i) F1, F2 and F3 values, (ii) spectrographic observation, and (iii) auditory perception. Fourth, mean F1, F2, F3 and duration values of the VC transitions were extracted for each V+/r/ sequence independently and were then compared with the mean values of the canonical schwa in arrive. Fifth, F1, F2 and F3 values for V were obtained from a single middle point in the most stable part of the pure vowels and in the offglide of the diphthongs. Finally, the differences between the values for V and those of the corresponding schwa-like element were calculated for each context in order to check for variability of the schwa-like element as a function of the preceding vowel. Figure 1 shows the measurement criteria followed for the analysis of the VC transitions and canonical schwa.

A first set of statistical analyses, consisting of one-way factorial ANOVAs, was aimed at obtaining overall significant differences in F1, F2, F3 and duration between the schwa-like element and canonical schwa. Fisher’s post-hoc tests were then carried out for individual comparisons between the schwa-like element in each of the contexts and the canonical schwa in the control word arrive, with context (each of the V+/r/ sequences) as the independent variable and F1, F2, F3 and duration as the dependent variables.

A second set of statistical analyses, using two-way factorial ANOVAs, was performed in order to test for F1, F2 and F3 variability between the schwa-like element in each of the contexts and the preceding vowel. In this case, the independent variables were rate (slow vs. fast) and context (each of the V+/r/ sequences), and the dependent variables were the F1, F2 and F3 values obtained from calculating the differences in F1, F2 and F3 between the schwa-like element and the preceding vowel.
Figure 1. Example tokens illustrating the measurement criteria for VC transitions (top) and canonical schwa (bottom) corresponding to one of Speaker 1’s slow readings of the words *fire* and *arrive*
3. Results

3.1 Schwa-like element and canonical schwa

The first set of statistical analyses, which consisted of one-way factorial ANOVAs performed to test for general differences in overall comparisons between the schwa-like element in each of the contexts and the canonical schwa in the control word *arrive*, revealed highly significant differences for all the dependent variables (F1, F2, F3 and duration) for both speakers. Significance level was set at p<.01. These results are provided in Table 2.

Table 2. Results of one-way factorial ANOVAs for overall significant differences in F1, F2, F3 and duration

<table>
<thead>
<tr>
<th></th>
<th>Speaker 1</th>
<th>Speaker 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>F(7,184) = 51.366, p&lt;.0001</td>
<td>F(7,184) = 44.792, p&lt;.0001</td>
</tr>
<tr>
<td>F2</td>
<td>F(7,184) = 430.990, p&lt;.0001</td>
<td>F(7,184) = 300.945, p&lt;.0001</td>
</tr>
<tr>
<td>F3</td>
<td>F(7,184) = 17.507, p&lt;.0001</td>
<td>F(7,184) = 9.123, p&lt;.0001</td>
</tr>
<tr>
<td>Duration</td>
<td>F(7,184) = 10.726, p&lt;.0001</td>
<td>F(7,184) = 3.204, p=.0032</td>
</tr>
</tbody>
</table>

Tables 3 and 4 show the results for Speaker 1 and Speaker 2, respectively, of the Fisher’s post-hoc tests carried out to test for particular differences in individual comparisons between the schwa-like element in each of the contexts and the canonical schwa in the control word *arrive*. Context (each of the V+/r/ sequences) was the independent variable, and F1, F2, F3 and duration were the dependent variables. Significance level was set at p<.01. Significant differences were found for duration and, more strikingly so, for F2 in all contexts for both speakers, thus showing that the schwa-like element in the sequences under study differs from the canonical schwa in *arrive* as regards F2 and duration. The results for both speakers concerning F1 and F3 differences between the schwa-like element and canonical schwa were significant only in some cases. In addition, no clear relationship can be established between the results for both speakers. For Speaker 1, F1 showed significant differences in the case of *here*, *far* and *fire*, but not in the case of *hair*, *pour*, *poor* and *power*. For this speaker, F3 showed significant differences for *hair*, *pour*, *poor* and *power*, but not for *here*, *far* and *fire*. For Speaker 2, F1 yielded significant

---

4. We thank Maria-Josep Solé and one anonymous reviewer for pointing out that, because the V+/r/ sequences are tautosyllabic while the canonical schwa+/r/ combination in *arrive* is heterosyllabic, duration differences between the two schwa types may be influenced by syllable affiliation. We are currently carrying out an experiment that takes this issue into consideration.
differences for *hair, far, fire and power*, but not for *here, pour and poor*. F3 for this speaker yielded non-significant differences in all cases except for *pour*.

Table 3. Results of individual comparisons (Fisher’s post-hoc tests) of F1, F2, F3 and duration values between the canonical schwa in *arrive* and the schwa-like element in each of the different contexts for Speaker 1, with context as the independent variable and F1, F2, F3 and duration as the dependent variables

<table>
<thead>
<tr>
<th>Context</th>
<th>F1 Mean differ.</th>
<th>p-value</th>
<th>F2 Mean differ.</th>
<th>p-value</th>
<th>F3 Mean differ.</th>
<th>p-value</th>
<th>Duration Mean differ.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>arrive-her</td>
<td>69.466</td>
<td>&lt;.0001</td>
<td>-475.838</td>
<td>&lt;.0001</td>
<td>55.043</td>
<td>.1935</td>
<td>37.521</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>arrive-hair</td>
<td>-24.015</td>
<td>.1655</td>
<td>-372.459</td>
<td>&lt;.0001</td>
<td>255.343</td>
<td>&lt;.0001</td>
<td>40.046</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>arrive-far</td>
<td>-215.499</td>
<td>&lt;.0001</td>
<td>83.645</td>
<td>.0002</td>
<td>43.138</td>
<td>.3077</td>
<td>36.217</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>arrive-pour</td>
<td>9.247</td>
<td>.5925</td>
<td>264.931</td>
<td>&lt;.0001</td>
<td>273.258</td>
<td>&lt;.0001</td>
<td>38.829</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>arrive-poor</td>
<td>7.935</td>
<td>.6460</td>
<td>271.124</td>
<td>&lt;.0001</td>
<td>271.723</td>
<td>&lt;.0001</td>
<td>40.483</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>arrive-fire</td>
<td>-104.422</td>
<td>&lt;.0001</td>
<td>-404.995</td>
<td>&lt;.0001</td>
<td>11.830</td>
<td>.7794</td>
<td>40.467</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>arrive-power</td>
<td>-43.457</td>
<td>.0126</td>
<td>215.422</td>
<td>&lt;.0001</td>
<td>180.697</td>
<td>&lt;.0001</td>
<td>39.658</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Table 4. Results of individual comparisons (Fisher’s post-hoc tests) of F1, F2, F3 and duration values between the canonical schwa in *arrive* and the schwa-like element in each of the different contexts for Speaker 2, with context as the independent variable and F1, F2, F3 and duration as the dependent variables

<table>
<thead>
<tr>
<th>Context</th>
<th>F1 Mean differ.</th>
<th>p-value</th>
<th>F2 Mean differ.</th>
<th>p-value</th>
<th>F3 Mean differ.</th>
<th>p-value</th>
<th>Duration Mean differ.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>arrive-hair</td>
<td>-67.239</td>
<td>&lt;.0001</td>
<td>-572.218</td>
<td>&lt;.0001</td>
<td>-10.821</td>
<td>.7470</td>
<td>15.908</td>
<td>.0096</td>
</tr>
<tr>
<td>arrive-far</td>
<td>-152.075</td>
<td>&lt;.0001</td>
<td>-90.189</td>
<td>.0007</td>
<td>21.504</td>
<td>.5217</td>
<td>17.288</td>
<td>.0049</td>
</tr>
<tr>
<td>arrive-pour</td>
<td>-23.378</td>
<td>.0526</td>
<td>87.557</td>
<td>.0010</td>
<td>169.414</td>
<td>&lt;.0001</td>
<td>17.580</td>
<td>.0043</td>
</tr>
<tr>
<td>arrive-poor</td>
<td>30.837</td>
<td>.0108</td>
<td>82.421</td>
<td>.0020</td>
<td>83.487</td>
<td>.0136</td>
<td>22.521</td>
<td>.0003</td>
</tr>
<tr>
<td>arrive-fire</td>
<td>-118.946</td>
<td>&lt;.0001</td>
<td>-587.411</td>
<td>&lt;.0001</td>
<td>-72.760</td>
<td>.0311</td>
<td>23.063</td>
<td>.0002</td>
</tr>
<tr>
<td>arrive-power</td>
<td>-65.738</td>
<td>&lt;.0001</td>
<td>86.577</td>
<td>.0012</td>
<td>23.755</td>
<td>.4791</td>
<td>23.486</td>
<td>.0002</td>
</tr>
</tbody>
</table>

3.2 Schwa-like element and preceding vowel

Figures 2, 3 and 4 show distribution graphs for F1, F2 and F3, respectively, of the variability between the schwa-like element in each of the contexts (*V+/t/ sequences*) and the preceding vowels. In the three figures, the top graphs display the results for
the slow tokens while the bottom graphs display the results for the fast ones. Similarly, the left graphs reveal the results for Speaker 1 while the right graphs reveal the results for Speaker 2. In addition, the squares stand for the mean formant values of the schwa-like element in the VC transitions, the triangles represent the mean formant values of the vowel in the CVC sequences, and the solid horizontal line depicts the mean formant values of the canonical schwa in the control word *arrive*.

Overall, the graphs in Figures 2, 3 and 4 show similar results for both speakers. First, F1, F2 and F3 values for the schwa-like element in the VC transitions are highly variable. Second, these values are also systematically different from those of the canonical schwa in *arrive*. Third, the values tend to resemble those in the preceding vowel. Finally, there is a higher degree of variability the faster the rate, that is, there is a smaller distance between F1, F2 and F3 values of the schwa-like element and those of the corresponding preceding value in the fast tokens than in the slow ones. Only a few tokens depart from this tendency. For Speaker 1, the distance between the values of the schwa-like element and of the preceding vowel is smaller in the slow tokens than in the fast ones in the case of *hair, far, pour, fire* and *power* (F1) and *here* and *far* (F3), and exactly the same in the case of *here* (F2). As regards Speaker 2, the distance is shorter for *here* (F1) and *far* (F2), and exactly the same for *fire* (F1). Once more, no clear relationship can be established between the results for both speakers.

The results of the second set of statistical analyses show significant effects in almost all cases and seem to corroborate the results shown by the data in the graphs of Figures 2, 3 and 4. In this case, two-way factorial ANOVAs testing for F1, F2 and F3 variability between the schwa-like element in each of the contexts and the preceding vowel were performed, with rate and context as the independent variables and with the F1, F2 and F3 values obtained from calculating the differences in F1, F2 and F3 between the schwa-like element and the preceding vowel as the dependent variable. Significance level was set at p<.01. The results are provided in Table 5.

4. **Discussion and conclusions**

The purpose of this study was to investigate the phonetic/phonological nature of word-final V+/*r*/ sequences in American English stressed monosyllables. We designed an acoustic experiment in an attempt to answer two main questions: (i) is there a schwa-like element in these V+/*r*/ sequences? and (ii) if so, is this schwa-like element the result of a process of epenthesis/insertion or of a process of coarticulation? So as to prove that the latter is the case, we formulated two more questions: (i) does this schwa-like element resemble canonical schwa? and (ii) does this schwa-like element vary systematically as a function of the vocalic context? The results of our study have provided answers to all these questions. Our findings corroborate those reported in a previous study that focused on V+/*l*/ sequences...
(Riera & Romero, 2007). Moreover, they expand on earlier studies that were preliminary in nature or that did not consider the full range of stressed vowels and diphthongs of American English in the treatment of V+/t/ sequences. Finally, our work is innovative in that it introduces the speaking rate variable to help explain the phenomenon under analysis.

Acoustic measurements have made it possible to determine that a vocalic schwa-like element can be distinguished, to a greater or lesser extent, in the VC transitions of all of the V+/t/ sequences under study. The presence of this element in all of the sequences leads us to conclude that we are dealing with a generalized process affecting all V+/t/ contexts, rather than a specific one concerning only a few, such as those involving high and/or front and/or tense vowels, which had been the object of previous research.

![Graphs of F1 values](image)

Figure 2. Distribution graphs for F1 values, in slow (top) and fast (bottom) tokens, for Speaker 1 (left) and Speaker 2 (right), with frequency shown in Hertz on the vertical axis and each of the contexts on the horizontal axis.
The results obtained in the one-way ANOVAs, which looked at overall comparisons between the schwa-like element and canonical schwa, yielded highly significant differences for both speakers. The results of the subsequent Fisher’s post-hoc tests, which focused on individual comparisons, despite not being significant for all contexts, corroborated the existence of these differences in most cases and for both speakers. The fact that this schwa-like element differs from canonical schwa in its F2 and duration values in all contexts and for both speakers provides evidence to conclude that this element is sufficiently different from canonical schwa for it not to be considered an epenthetic schwa. In a process of schwa epenthesis/insertion, the formant and duration values of this element were expected to be more similar to those of canonical schwa than our results proved it to be. Thus, we attribute the presence of this element to a process of coarticulation.

Figure 3. Distribution graphs for F2 values, in slow (top) and fast (bottom) tokens, for Speaker 1 (left) and Speaker 2 (right), with frequency shown in Hertz on the vertical axis and each of the contexts on the horizontal axis.
The analysis of the data that appear in the graphs of Figures 2, 3 and 4, as well as the subsequent two-way ANOVAs, were performed in order to explore the degree of F1, F2 and F3 variability between the schwa-like element in each of the contexts and the preceding vowel. The results obtained provide further evidence to confirm our hypothesis favoring coarticulation versus epenthesis/insertion. First, it is possible to detect the existence of a highly variable schwa-like element in terms of its spectral and duration values on the one hand, and as a function of the preceding vowel on the other. Second, this schwa-like element once more appears to be systematically different from canonical schwa. And third, the values of this schwa-like element tend to resemble those of its preceding vowel, with an increase in variability, the faster the speaking rate. Despite exceptions and the lack of significant results in some cases, these observations apply to the great majority of contexts in our study.

Figure 4. Distribution graphs for F3 values, in slow (top) and fast (bottom) tokens, for Speaker 1 (left) and Speaker 2 (right), with frequency shown in Hertz on the vertical axis and each of the contexts on the horizontal axis.
Table 5. Results of two-way factorial ANOVAs for F1, F2 and F3 variability between the schwa-like element in each of the contexts (V+/r/ sequences) and the preceding vowel, with rate and context as the independent variables and with F1, F2 and F3 values obtained from the differences between the schwa-like element and the preceding vowel as the dependent variables

<table>
<thead>
<tr>
<th></th>
<th>Speaker 1</th>
<th>Speaker 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>Rate: ( F(1,154) = 0.438, p=.5091 )</td>
<td>Rate: ( F(1,154) = 14.689, p=.0002 )</td>
</tr>
<tr>
<td></td>
<td>Context: ( F(6,154) = 9.297, p&lt;.0001 )</td>
<td>Context: ( F(6,154) = 10.108, p&lt;.0001 )</td>
</tr>
<tr>
<td></td>
<td>Rate*Context: ( F(6,154) = 0.850, p=.5333 )</td>
<td>Rate*Context: ( F(6,154) = 3.509, p=.0028 )</td>
</tr>
<tr>
<td>F2</td>
<td>Rate: ( F(1,154) = 52.928, p&lt;.0001 )</td>
<td>Rate: ( F(1,154) = 92.267, p&lt;.0001 )</td>
</tr>
<tr>
<td></td>
<td>Context: ( F(6,154) = 130.146, p&lt;.0001 )</td>
<td>Context: ( F(6,154) = 90.806, p&lt;.0001 )</td>
</tr>
<tr>
<td></td>
<td>Rate*Context: ( F(6,154) = 6.762, p&lt;.0001 )</td>
<td>Rate*Context: ( F(6,154) = 11.569, p&lt;.0001 )</td>
</tr>
<tr>
<td>F3</td>
<td>Rate: ( F(1,154) = 6.887, p=.0096 )</td>
<td>Rate: ( F(1,154) = 202.374, p&lt;.0001 )</td>
</tr>
<tr>
<td></td>
<td>Context: ( F(6,154) = 48.183, p&lt;.0001 )</td>
<td>Context: ( F(6,154) = 6.692, p&lt;.0001 )</td>
</tr>
<tr>
<td></td>
<td>Rate*Context: ( F(6,154) = 6.373, p&lt;.0001 )</td>
<td>Rate*Context: ( F(6,154) = 2.993, p=.0379 )</td>
</tr>
</tbody>
</table>

All in all, our results provide convincing evidence for considering the schwa-like element in V+/r/ sequences as the result of a dynamic process of coarticulation rather than a discrete process of epenthesis/insertion. A process of coarticulation requires two or more sounds to come in contact with each other and influence one another either in a progressive manner (from left to right) or in a regressive manner (from right to left). Consequently, a process of coarticulation also implies that, far from being produced independently from each other and one after another, neighboring sounds overlap, blend and even disappear in extreme cases. Our understanding of the processes under analysis is more in accordance with this continuous nature of speech production than with any other involving a simple categorical process, such as that of vowel epenthesis/insertion. Furthermore, considering this issue from the point of view of articulation dynamics allows for a convincing explanation of the behavior of the schwa-like element due to speaking rate differences. An increase in speech rate entails a decrease in time for articulatory gestures to attain their targets and, thus, the faster the rate is, the more overlap and blending there is in the transitions and the more difficult it is to predict exactly what these will look like.

Our conclusions lead us to claim that the term *epenthetic schwa* is not appropriate when defining this schwa-like element. Despite resembling schwa, this element also has spectral values similar to those of its preceding vowel, to the point of becoming a different vocalic element in each of the different contexts, as a function of the vowel that precedes it. Other terms such as *targetless schwa* (Browman & Goldstein 1992b), *excrecent schwa* (Gick & Wilson 2001, 2006) or *intrusive schwa* (Hall 2003, 2006) seem more appropriate, since they imply the existence of an element with
differing spectral and duration values according to context. Moreover, a phonological/phonemic representation of this schwa-like element, i.e., /ə/ or /ɐ/, such as the one present in some pronunciation dictionaries and text books, aimed mainly at aiding foreign speakers’ English pronunciation, does not seem very adequate. In our view, a representation of this element would only be appropriate as an aid in second/foreign language pronunciation learning.

The present study has some limitations that should be overcome in further studies so as to lead to more consistent and reliable results as well as to a better understanding of the processes involved in the VC transitions of the V+/ɾ/ sequences studied here. First, the number of participants should be increased, and speakers of different varieties of American English should be included. Second, a perception study would complement the production studies so far undertaken. Third, an analysis of the spectral and duration values of /ɾ/ would reveal the extent to which this coda consonant influences the presence and magnitude of the schwa-like element, and at the same time show how this /ɾ/ is influenced by the variability of the preceding schwa-like element. Fourth, an attempt to classify the preceding vowels, in terms of phonological parameters such as tongue height, tongue advancement, lip rounding and the tense-lax distinction, and according to the behavior of the schwa-like element, would provide a more accurate picture of the phenomenon. Fifth, canonical schwa should be studied and contrasted with the schwa-like element in greater and more varied contexts. Finally, no work on this subject would be complete without taking into consideration articulatory data such as that gathered by means of electromagnetic midsagittal articulometry systems. The authors of this paper are currently working on a wider study which aims to take into account these suggestions.

References

Schwa in American English V+/r/ sequences


Schmeiser, Benjamin. 2009. “Prosodic and Segmental Effects on Vowel Intrusion Duration in Spanish /r C/ Clusters”. This volume.


Perception of word stress in Castilian Spanish

The effects of sentence intonation and vowel type*

Marta Ortega-Llebaria¹ & Pilar Prieto²
¹University of Texas at Austin/²ICREA & Universitat Autònoma de Barcelona

We provide evidence for the perception of the stress contrast in unaccented contexts in Spanish. Twenty participants were asked to identify oxytone words which varied orthogonally in two bi-dimensional paroxytone-oxytone continua: one of duration and spectral tilt, and the other of duration and overall intensity. Results indicate that duration and overall intensity were cues to stress, while spectral tilt was not. Moreover, stress detection depended on vowel type: the stress contrast was perceived more consistently in [a] than in [i]. Thus, in spite of lacking vowel reduction, stress in Spanish has its own phonetic material in the absence of pitch accents. However, we cannot speak of cues to stress in general since they depend on the characteristics of the vowel.

1. Introduction

The goal of this article is to investigate which acoustic cues and cue-interactions Spanish speakers use to perceive primary stress in unaccented contexts. Previous studies in Spanish have examined cues to stress in accented contexts, where there is co-variation between stress and accent (Enríquez et al. 1989; Llisterri et al. 2004 and Solé 1985, among others). For example, Llisterri et al. (2004) examined the perception of stress in one word declarative utterances, which are consistently produced with a pitch accent on the stressed syllable and lengthening on the last syllable. Consequently, they found that pitch was the main cue to stress, followed first by duration, and then by intensity. More specifically, duration and intensity cues were not sufficient to perceive stress unless they were processed together with a pitch accent. However, as Beckman and Edwards (1994:13) explained, this is a source of misunderstanding:

in short utterances, pitch excursions are likely to be interpreted in terms of the sequence at a nuclear accent, as in Fry’s 1955 experiment showing the salience

*Many thanks to Eva Estebas and Ana Estrella for their help with data collection. This research was funded by a grant from the Ministerio de Educación y Ciencia, Spain.
of the F0 contour in cueing stress in pairs such as *pérmit* versus *permit*. This is probably the major source of the common misunderstanding in the experimental literature that F0 excursion is a direct acoustic correlate of the feature ‘stress’, a misunderstanding that has been incorporated into several standard textbooks.

Nevertheless, several studies have investigated how speakers of different languages use duration, intensity and vowel reduction in the production and perception of stress while controlling for the potential effects of accent. The earliest extensive instrumental study of this sort in English was that of Huss (1978). He embedded word minimal pairs that differed in primary stress, i.e. *import* noun – *import* verb, within post-focal contexts like those below. Note that vowels in *import* noun – *import* verb do not become reduced when they do not receive lexical stress, and that consequently, this word minimal pair contrasts vowels with primary and secondary stresses, not vowels with primary stress and unstressed vowels.

The GERMANS’ import sinks. (target noun)
The GERMANS import sinks. (target verb)

Huss’ results showed that although English speakers did produce small duration and intensity differences between vowels with primary and secondary stress, the very same speakers were not able to perceive these differences. Based on these results, Huss concluded that in the absence of pitch-accents and vowel reduction patterns, English speakers cannot perceive primary stress based solely on duration and intensity differences.

Beckman and colleagues (1994, 1997) obtained production results for English that complement those of Huss’ (1978). Campbell and Beckman (1997) compared fully stressed vowels, i.e. vowels with primary stress, with unreduced unstressed vowels, i.e. vowels with secondary stress, in different pitch-accent contexts. Their results revealed that the spectral balance did not differentiate vowels with primary stress from vowels with secondary stress in the absence of a pitch accent. They also observed that their 4 subjects varied in how reliably they used other cues, mainly duration, to mark stress. Based on Huss’ results and their own, they concluded that “there are no direct acoustic correlates for stress, and that, instead, the phonetic properties associated with stress at any level are parasitic on the phonetics of the relevant prominence marker”. More specifically, they agreed with Huss (1978) that at the lower levels of the prominence hierarchy, vowel reduction is a stronger correlate of word stress in English than duration and intensity:

categorical contrasts in ‘primary stress’ will be maintained in post-nuclear position [i.e. in a context where the word does not bear a pitch-accent] only if the words differ also in stress-foot structure. For example *digést* should be categorically perceived as different from *digest* only if the verb has [ɪ] rather than the full vowel [aɪ] in its first syllable (Beckman & Edwards 1994:15).
However, studies in other languages have examined lexical stressed syllables in deaccented contexts (Sluijter & van Heuven 1996; Sluijter, van Heuven & Pacilly 1997 for Dutch; Manolescu et al. 2009, this volume for Romanian; Dogil & Williams 1999 for German; Kastrikani 2003 for Greek; Ortega-Llebaria 2006; Ortega-Llebaria & Prieto 2009 for Spanish) and found that stressed syllables were consistently longer than their unstressed counterparts in the absence of pitch accents, revealing that duration was a cross-linguistic correlate of stress at the lower levels of the prominence hierarchy. For example, Manolescu and colleagues (2009, this volume) found that Romanian speakers consistently produced stressed syllables with longer durations than unstressed syllables in both declarative sentences with broad focus, where there is co-variation between stress and accent, and in unaccented post-focal contexts, where both stressed and unstressed syllables have a flat F0 melody.

Duration was also a strong correlate of stress in Dutch (Sluijter & van Heuven 1996; Sluijter, van Heuven & Pacilly 1997), even though Dutch and English are two closely related languages that have similar stress placement and vowel reduction patterns. Sluijter and colleagues found that in unaccented contexts, Dutch speakers, like English speakers, produced vowels with primary stress with longer durations, flatter spectral tilts and fuller vowel qualities than their unstressed counterparts. However, unlike English speakers, Dutch speakers mainly used duration and spectral tilt cues to perceive primary stress. In fact, when the perception task was performed with reverberation noise in the background, Dutch listeners continued to rely on duration cues but increased their reliance on intensity cues, demonstrating the ecological validity of their results. Thus, Sluijter and colleagues showed that at the lower levels of the prominence hierarchy, Dutch differed from English in that duration and intensity were stronger cues to perceiving primary stress than vowel reduction patterns. They conclude that

stress is not just a weaker degree of accent. One would expect to observe lower values along all measured correlates in stressed syllables of unaccented words. However, what we do observe is weakening along only those correlates that are related to the omission of the accent-lending pitch movement. (Sluijter & van Heuven 1996:2483).

Campbell and Beckman (1997) explained these cross-linguistic differences by appealing to the different strategies used by Dutch and American English in making stress perceptible in the absence of a pitch-accent: “Dutch differs from English in having relatively fewer words in which unstressed syllables are reduced” (Campbell & Beckman 1997:70). Consequently, vowel reduction constitutes a sufficient cue to stress only in English, while in Dutch, speakers need to phonologize other cues, such as duration and intensity, as correlates of stress.

```
Stressed vowels       Unstressed vowels
| i                        | i                        |
| e                        | ə                        |
| ε                        |                         |
| a                        |                         |
| ə                        |                         |
| o                        |                         |
| u                        | u                        |
```

Figure 1. Inventory of stressed and unstressed vowels in Central Catalan. Lines indicate reduction patterns

Thus, since Catalan has a consistent morphophonological alternation between full long /e, ε, a, ə, o/ in stressed syllables versus schwa or [u] in unstressed syllables, our working hypothesis was that Catalan should behave more like English in relying mainly on vowel quality and not needing to phonologize duration and loudness as correlates of primary stress in the absence of accent. Results from our production experiment showed that, as expected, Catalan speakers reduced vowel [a] into a schwa when it became unstressed, and that they produced [i] with similar vowel qualities in stressed and unstressed contexts. Crucially, in spite of their different reduction patterns, both vowels were produced with longer durations and louder intensities in stressed contexts and this effect was intensified in vowels that underwent vowel reduction. Thus, having vowel reduction as a cue to stress not only did not prevent speakers from phonologizing duration and intensity cues, but vowel reduction also extended the duration and deepened intensity differences to the stress contrast. Results from our perception experiment further
disproved our working hypothesis (Ortega-Llebaria, Prieto & Vanrell, submitted). They showed that like Dutch speakers, Catalan speakers did rely on duration and intensity cues to perceive stress in vowel [i] but at the same time, like English speakers, they relied on vowel quality to perceive stress on vowel [a]. However, in contrast to Huss’ English speakers who were unable to process duration and intensity cues in relation to stress, Catalan speakers still perceived the stress contrast by relying on duration and intensity cues when vowel reduction patterns for vowel [a] were neutralized in the speech signal. Moreover, once vowel reduction patterns were made available in the stimuli, Catalan speakers relied heavily on them to perceive stress.

Our results for Catalan indicate that the perception of primary stress is based on a cluster of cues whose weights change according to vowel type. Our subjects used vowel quality, duration and intensity cues to perceive primary stress in [a], and to duration and intensity cues to perceive primary stress in [i]. However, not a single cue, including vowel reduction, was absolutely necessary for the perception of stress. Even when we eliminated the vowel reduction patterns in [a], thereby creating an unusual context in Catalan, Catalan listeners, unlike Huss’ English speakers, still perceived primary stress by relying on duration cues and if they were absent, on intensity cues.

Results from production experiments in Castilian Spanish (Ortega-Llebaria 2006; Ortega-Llebaria & Prieto, submitted) also support that notion that word stress is expressed by a cluster of acoustic correlates, and because in Castilian Spanish there is no phonological vowel reduction (Hualde 2005), these correlates work independently of vowel reduction patterns. While Spanish speakers produced the five vowels, i.e. [a,e,i,o,u], with the identical qualities in stressed and unstressed contexts, they produced stressed vowels with longer durations and louder intensities than unstressed vowels in both declarative sentences, where there is co-variation between stress and accent, and reporting clauses, where all syllables are unaccented. However, the lengthening effect of stress was larger in vowel [o] than it was in vowel [i], and intensity correlates to stress were larger in reporting clauses than in declarative sentences. We concluded that, although duration and intensity were consistent cues in the production of primary stress in Spanish, these cues had different weights according to sentence intonation and vowel type.

In order to examine how the duration and intensity cues found in the previous production experiment are perceived in relation to primary stress, we prepared the present perception experiment. Twenty native speakers of Castilian Spanish were asked to listen to and identify oxytone words which contained different vowels, in unaccented sentences. The target words varied orthogonally in two bi-dimensional
paroxytone-oxytone continua: one of duration and spectral tilt, and the other of duration and overall intensity. The research questions were:

1. In the absence of pitch-accents and vowel reduction patterns, can Spanish speakers still perceive a stress contrast based on duration and intensity cues?
2. How do duration and intensity cues interact?
3. How do Spanish speakers use duration cues to perceive stress across vowels? More specifically, when identical duration modifications are applied to vowels [a] and [i], do speakers rely less on duration cues to perceive stress in vowel [i], reflecting production patterns?
4. Since results in production indicate that overall intensity is a stronger correlate of stress than spectral tilt, do listeners also rely more on overall intensity cues than on spectral tilt to perceive stress?

2. Methodology

2.1 Recordings

A 41 year-old female monolingual speaker of Spanish from Barcelona was recorded saying the sentence *Hola – saluda mama contenta* “Hello, greets mama happily”, where the reporting clause *saluda mama contenta* was consistently pronounced with a flat pitch melody. The paroxytone target word *mama* was replaced with oxytone *mamá*, oxytone *mimí* and paroxytone *mimi*, yielding a total of 60 sentences (4 target words * 15 repetitions). Measurements of duration, intensity and spectral tilt were made on all target words and the sentences containing the paroxytone *mama* and *mimi* items with values closest to the average were selected for further manipulation. Vowels from these tokens had the same vowel quality.

2.2 Materials

Three *mama-mamá* (and *mimi-mimí*) continua were created by manipulating the cues of duration, overall intensity and spectral tilt separately. For each continuum, stimulus 1 had the syllable ratio typical of paroxytone words. This ratio decreased in stimulus 2, was close to 0 in stimulus 3, and increased again but in the opposite direction for stimuli 4 and 5, with stimulus 5 replicating the ratio of oxytone words. For example, for the overall intensity continuum, syllable 1 was 3 dBs louder than syllable 2 in stimulus 1. This difference decreased to 1.5 dBs in stimulus 2, and came close to 0 in stimulus 3. In stimulus 4, the second syllable was 1.5 dBs louder than the first one, and this difference increased to 3 dBs in stimulus 5 (see Table 1).
In order to change the duration ratio between the two syllables of *mama* (or *mimi*) while maintaining the same word duration, we cut one cycle from vowel 1 while adding another cycle to vowel 2. Because we modeled our stimuli after our recorded speaker and she consistently made larger inter-syllabic duration differences in oxytone than in paroxytone words to the extent that both syllables of some paroxytone *mama* or *mimi* tokens had roughly the same duration while in oxytone tokens the second syllable was on average 28 ms longer than the first, we kept these ratio differences in our test materials. Therefore, starting from stimulus 1, the number of cycles in each vowel of *mama* was 13–11, 12–12, 11–13, 10–14, 9–15, and in *mimi* 10–8, 9–9, 8–10, 7–11, 6–12. Since each cycle lasted 7 ms, the first vowel in stimulus 1 was 14 ms longer than the second vowel. In stimulus 2, this difference dropped to 0 ms, while in stimulus 3, the second vowel was 14 ms longer than the first. In stimulus 4, the second vowel was 28 ms longer than the first, and in stimulus 5, the difference was 36 ms.

In the spectral tilt continuum, the increments were 4 dBs in both the *mama* and *mimi* tokens. The main difference between the spectral tilt and overall intensity manipulations was that in the former, the increments of loudness were only applied in frequencies ranging from 500 Hz to 3000 Hz, while in the latter, the amplitude increments were applied uniformly across all the frequencies of the spectrum. All increments had values very similar to those employed by Sluijter et al. (1996, 1997).

| Table 1. Increments for the overall intensity, spectral tilt and duration continua |
|-------------------------------------------------|-------------------------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Stimulus 1                                      | Stimulus 2                                      | Stimulus 3                       | Stimulus 4                       | Stimulus 5                       |
| Clear                                          | paroxytone                                     | neutral                         | oxytone                         | Clear                           |
| Syllable ratios                                 | Syll 1> syll 2                                 | Syll 1= syll 2                  | Syll 1< syll 2                  | Syll 1< syll 2                  |
| Over. Intensity                                 | +3 dBs                                        | +1.5 dBs                        | +1.5 dBs                        | +3 dBs.                         |
| Spectral tilt                                  | +8 dBs                                        | +4 dBs                          | +4 dBs                          | +8 dBs.                         |
| Syllable ratios                                 | Syll 1> syll 2                                 | Syll 1= syll 2                  | Syll 1< syll 2                  | Syll 1< syll 2                  |
| Duration                                       | +2 cycles                                      | + 2 cycles                      | + 4 cycles                      | + 6 cycles                      |

Finally, the five levels of the duration continuum were crossed with those of the overall intensity continuum, creating a 5*5 grid for each vowel. For example, grid 1 contained the 25 independent stimuli for the *mama-mamá* contrast and grid 2 included the 25 independent stimuli for *mimi-mimí*. Similarly, the five levels of duration continuum were also crossed with the five levels of spectral tilt, thus yielding grids 3 and 4. A summary of the crossed continua and resulting grids is depicted in Table 2.
Table 2. Summary of the five identification tasks

<table>
<thead>
<tr>
<th>Crossed continua</th>
<th>Tokens</th>
<th>Grid</th>
<th>Repetitions</th>
<th>Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration * overall intensity</td>
<td>25 mama</td>
<td>1</td>
<td>7</td>
<td>Group A = 10 subjects</td>
</tr>
<tr>
<td></td>
<td>25 mimi</td>
<td>2</td>
<td>7</td>
<td>Group B = 10 subjects</td>
</tr>
<tr>
<td>Duration * spectral tilt</td>
<td>25 mama</td>
<td>3</td>
<td>7</td>
<td>Group A = 10 subjects</td>
</tr>
<tr>
<td></td>
<td>25 mimi</td>
<td>4</td>
<td>7</td>
<td>Group B = 10 subjects</td>
</tr>
</tbody>
</table>

2.3 Subjects and listening tasks

Twenty native speakers of Spanish participated in the study. Their ages ranged from 21 to 60, and they had either been born in Madrid, Spain, or lived there for most of their lives. None of them reported having any speech or hearing problems.

A group of ten subjects (Group A in Table 2) was told that they would hear the words *mama* (with paroxytone stress) or *mamá* (with oxytone stress) inserted in the sentence *Hola – saluda ___ contenta*. They were asked to press the space bar on a keyboard as soon as they heard the oxytone word *mamá*. They listened to the 175 randomized sentences of the ‘duration * overall intensity’ condition (25 stimuli of grid 1* 7 repetitions) in 7 blocks of 25 stimuli with an ISI of 1500 ms and a 10-second break between blocks. After a longer rest, they then listened to the 175 sentences of the ‘duration * spectral tilt’ condition (25 stimuli of grid 3 * 7 repetitions). Orders of presentation between the two tasks were counterbalanced across subjects. The remaining 10 speakers (Group B in Table 2) performed analogous tasks with the *mimi-mimí* stimuli.

2.4 Statistics

The effect of duration and overall intensity cues in the perception of primary stress was assessed by performing a Repeated Measures ANOVA with the within-subjects factors of duration and overall intensity and the between-subjects factor of vowel ([a] and [i]) on the probability of identification scores obtained from grids 1 and 2 (see 3.1 below). An analogous analysis was performed on the probability identification scores obtained from the duration and spectral tilt modifications of grids 3 and 4 (3.2).

3. Results

3.1 Duration and overall intensity

The graph in Figure 2 illustrates the probabilities of *mamá* answers for the 25 stimuli. The slopes of the functions indicate that duration has a strong effect on
the perception of stress: for each intensity level, the oxytone responses increase along the duration continuum regardless of whether they are in a competing or enhancing relationship with intensity. As an example, the probabilities of identification of mamá in intensity level 1 (thick grey line) start close to 0 in duration 0.40 increasing progressively to 0.42, 0.55 and 0.63 until they reach 0.80 in duration 5. In addition to the strong effect of duration on stress perception, the spread of the intensity curves indicates that intensity also has an effect because the base rate of mamá responses increases with higher intensity levels. Thus, speakers rely heavily on duration, but the addition of intensity as an enhancing cue results in a shift of the response profile toward the alternative mamá.

Figure 2. Probabilities of mamá responses for the 25 stimuli resulting from crossing the 5-levels of duration with the 5-levels of overall intensity in a máma-mamá continuum.

It is easier to observe this effect of intensity in Figure 3, which depicts the same probabilities of mamá answers as in Figure 1, but with intensity along the x-axis. In general, there are fewer oxytone responses for intensity 1 than there are for intensity 5 which yields a slight ascendant trajectory along the intensity continuum for each duration level indicating that the oxytone responses increase along the intensity continuum. Although this ascendant trajectory becomes most visible at duration 3 showing that intensity has a stronger effect when duration cues are ambiguous, the slopes are by no means as strong as those depicted along the duration continuum in Figure 2. Therefore, both duration and intensity contribute to the perception of stress in [a] in an additive manner. Nonetheless, duration seems to be a stronger cue than intensity.
The two graphs in Figure 4 show the probabilities of mimí judgements. In contrast with the results from mamá, subjects do not rely on duration and intensity to predict stress. Rather, they rely exclusively on intensity since oxytone responses increase only along the intensity continuum (as shown in the right-hand graph), but not along the duration continuum (left-hand graph). Even in stimuli with conflicting cues, such as paroxytone duration 1 with oxytone intensity 5, speakers favor intensity over duration and tend to hear oxytone mimí. Interestingly, when duration cues become ambiguous (level 3 in the x-axes of the left graph, and the dotted red line on the right graph), all intensity levels score around chance (.5) showing that speakers stop using intensity cues in the absence of clear duration cues to stress. Thus, in contrast with [a], duration and intensity do not seem to be in an additive relationship because listeners rely only on intensity cues to perceive stress in vowel [i]. However, duration shapes intensity judgments in that listeners use intensity solely when duration is not an ambiguous cue.

Results from a Repeated Measures ANOVA with the factors of duration and intensity on the probability scores confirmed the above patterns. Duration was significant only in mamá ([a]: F(4,32)=15.753, p<.0001; [i]: F(4,32)=1.041, p=.401) while intensity was strongly significant in mimí and only marginally significant in mamá ([a]: F(4,32)=3.343, p=.042; [i]: F(4,32)=7.933, p<.0001) showing that cues to stress depend on vowel type. While listeners use both duration and intensity cues
to perceive stress in [a], in [i] they only use intensity. Moreover, since interactions were not significant, they show that duration and intensity are additive in [a]. Partial Eta-Square estimates further confirm these results showing that duration in [a] explains 63% and intensity 23% of the variance, while in vowel [i] intensity alone accounts for 50% of the variance.

Multiple comparisons between intensity levels at neutral duration (stimulus 3) yielded significant results only for vowel [a]. Stimuli 1 and 2 were significantly different from stimuli 4 and 5. Therefore, in the absence of clear duration cues, intensity has an effect on the perception of stress only for vowel [a].

### 3.2 Duration and spectral tilt

The two graphs in Figure 5 illustrate the probabilities of mamá answers (graph on the left) and mimí answers (graph on the right) for the 25 mama and 25 mimí stimuli that resulted from crossing the 5-step duration continuum with the 5-step spectral tilt continuum. As in Figure 1, oxytone mamá or mimí answers increase along the duration continuum showing that duration has a strong effect on the perception of stress. However, the spread amongst the curves of the 5 spectral tilt levels is not as wide as that in Figure 2. Thus, speakers consistently relied on duration over intensity to perceive stress either when intensity was computed as changes in overall intensity or as variations in spectral tilt. Yet variations in spectral tilt seem to have a weaker effect than those in overall intensity.
Results from a RM ANOVA with duration and spectral tilt factors reveal that duration is the only significant cue in the perception of stress (duration: $F(4,68)=7.034$, $p=.003$, $dur*vowel$: $F(4,68)=301$, $p=.872$, intensity: $F(4,68)=1.679$, $p=.211$, $int*vowel$: $F(4,68)=1.906$, $p=.165$), thus confirming that duration is a strong cue to stress while spectral tilt is not. Partial Eta-Square Estimates further corroborate this result by showing that duration explains most of the variance in the data, i.e. 60% in [a] and 55% in [i].

A visual comparison of all six graphs shows that probability scores are consistently higher for [a] than they are for [i]. While 72% of the data for [a] score above .5, fewer than 50% of the data for [i] score above .5. Mean d-prime scores confirm that it was easier for speakers to detect oxytone words in vowel [a] than in vowel [i], i.e. d-prime for [a] scores 1.8, while for [i] it only reached 0.72.

In summary, results from the perception tasks show that Spanish speakers do perceive stress in reporting sentences, thereby confirming that stress can be perceived in the absence of pitch accents, and since Spanish has no vowel reduction, also in the absence of vowel reduction patterns. Perception of stress was based on a cluster of cues, namely duration and overall intensity, whose weights changed according to vowel type. In [a], speakers relied more on duration than on overall intensity cues. Yet, when duration cues to stress became ambiguous, speakers relied heavily on overall intensity. In contrast with vowel [a], overall intensity, not duration, was the main cue to stress in vowel [i]. Moreover, when duration cues became ambiguous for stress perception, speakers did not rely on overall intensity instead. Finally, spectral tilt did not have any effect on the perception of stress in Spanish.
4. Discussion

Our results demonstrate that even in the absence of pitch-accent and vowel reduction patterns in the speech signal, Spanish speakers still detect the stress contrast on the basis of duration and overall intensity differences between adjacent syllables, while ignoring differences in spectral tilt. Thus, the answer to our first research question is affirmative: speakers of Castilian Spanish perceive stress in unaccented contexts confirming that at the lower levels of the prosodic hierarchy, stress in Spanish has its own phonetic material which works independently of vowel reduction patterns.

However, the use that our Spanish speakers make of duration and intensity cues to perceive word stress differs across vowels. When listening to the *mama* stimuli, our Spanish speakers obtained higher identification scores in the stimuli where both cues were in an enhancing relationship, rather than in the stimuli where these cues were in a competing relationship. This shows that they use duration and overall intensity cues in an additive manner to perceive the stress contrast in vowel [a]. However, for vowel [i], our listeners use duration – and not very successfully – only if overall intensity is not an available cue in the signal. As mentioned in the introduction, the production experiments (Ortega-Llebaria & Prieto 2009) also shows a different use of duration in relation to stress across vowels. The lengthening effect of stress was smaller in vowel [i] than it was in other vowels, possibly because [i] is the vowel used in processes of speaker normalization (Johnson 1990). In contrast, overall intensity differences between stressed and unstressed vowels were similar across vowels. Thus, mirroring production patterns, results from this perception experiment showed that listeners relied more on duration cues when listening to stress in vowel [a] than in [i] in spite of performing the same manipulations of duration and intensity on the *mama* and *mimi* target words. Perception patterns therefore reflect the knowledge that speakers have of production. Since duration cues to stress in production are less salient in [i] than in other vowels, listeners rely less on duration when perceiving stress in vowel [i] than in vowel [a].

Moreover, listeners obtained higher d-prime scores for [a] than for [i] showing that they were more sensitive to the stress prominence in [a]. This asymmetry is related to the finding that manipulations of intensity tend to have a weaker effect on stress perception than manipulations of duration (Llisterri et al. 2004 for Spanish, Turk & Sawusch 1996 for English). For example, Turk and Sawusch (1996) found that English listeners’ perception of duration and intensity in relation to stress was non-orthogonal. More specifically, they showed that irrelevant variations in duration when listeners were attending to intensity had a greater effect on the perception of stress than irrelevant variations in intensity when listeners were attending to duration. Thus, since our Spanish listeners attended more to duration cues when perceiving
stress in *mama*, and to intensity cues when perceiving stress in *mimi*, and duration cues have a stronger effect than intensity cues in the perception of stress, listeners were more sensitive to the stress contrast in *mama* than they were in *mimi*.

In short, our results indicate that the perception of stress in Spanish is based on a cluster of cues whose weights and interactions vary according to the phonology of the language, requirements of sentence intonation and vowel type. The phonology of Spanish rules out vowel reduction as a possible cue to stress (Hualde 2005). Moreover, the flat F0 contour of reporting clauses prevents pitch accents from cueing stress. In spite of this, Spanish listeners still perceive the stress contrast based on variations of duration and overall intensity. Moreover, vowel type determines the weights of duration and overall intensity cues. Spanish speakers produced longer vowels in stressed than in unstressed syllables, but these lengthening effects of stress had a more restricted range in vowel [i] than in other vowels. Mirroring production patterns, listeners relied less on duration when perceiving stress in [i] than when perceiving stress in [a]. However, it is necessary to test the perception of stress in the remaining Spanish vowels in order to fully understand the effect of vowel type in the weighting and integration of cues to stress.

Cross-linguistic research provides cumulative evidence that stress is not a reduced type of accent in stress-accent languages like Catalan, Spanish, and Dutch. While in stress-accent languages pitch accents aid to the perception of stress in the sense that only stressed syllables can bear a pitch accent, in the absence of pitch accents, the stress contrast is still perceptible. Duration and overall intensity in Spanish, duration and spectral tilt, and vowel reduction in Dutch, and duration, overall intensity and vowel reduction in Catalan constitute the phonetic material that cue the stress prominence in unaccented contexts showing that stress has its own phonetic material. This material is language specific because it varies according to the phonology of each language, and as shown for Catalan and Spanish, this material also depends on vowel type. In view of these data, it seems equally difficult to define stress as a mere structural device empty of any phonetic content as it is to find cross-linguistic cues to stress. Following Lieberman (1960), it seems reasonable to propose that the perception of stress is based on a cluster of cues that conveys rhythm to an utterance and that the variation we observed in the specific cues and cue weights that make up this cluster could be explained in relation to the perception of the rhythmic patterns of a language.

5. Conclusions

Results from this experiment indicate that in the absence of vowel reduction and pitch accents, Spanish speakers perceive stress by extracting the remaining
information from the signal. This information, in turn, is conditioned by vowel type. Listeners in this study relied mostly on duration to perceive stress in vowel [a] and on overall intensity in vowel [i]. Thus, it seems that perceived prominence in an unaccented context is based on a cluster of parameters, whose weights and interactions are determined by the phonology of the language and the requirements of sentence intonation and vowel type.

References


Manolescu, Alis, Daniel Olson & Marta Ortega-Llebaria. 2009. Cues to Contrastive Focus in Romanian. This volume.


Do complex pitch gestures induce syllable lengthening in Catalan and Spanish?*

Pilar Prieto¹ & Marta Ortega-Llebaria²
¹ICREA & Universitat Pompeu Fabra/²University of Texas at Austin

In both Spanish and Catalan, narrow contrastive focus and presentational broad focus in nuclear position have different pitch accent choices, namely a rising or a falling pitch accent, respectively. In words with final stress, narrow contrastive focus displays a rise-fall complex pitch gesture in the last syllable of the utterance. This article investigates the effects of the complexity of such a pitch pattern on the durational properties of the syllables in both languages when compared to the simpler falling pitch movement. The results of the production experiment reveal that, in general, the presence of a complex pitch pattern tends to have a lengthening effect on the target syllable. Yet we also find that some instances of this complex contour can be partially truncated, in which case it does not trigger lengthening. In sum, even though truncation and compression have been claimed to be language- and dialect-specific strategies (Ladd 1996; Grabe 1998; Grabe et al. 2000), in our data, truncation can be considered a speaker phonetic realization strategy that interacts with timing in such a way that there is a trade-off relationship between the two factors.

1. Introduction

In languages with a contrast between contour and non-contour tones, there are often restrictions on where those categories can occur. These restrictions vary from language to language, but there are some such patterns that recur independently

---

*The article further develops materials presented at the 3rd PaPI Conference (Braga, June 2007). We are grateful to the audience at this conference and especially to John Kingston, G. Elordieta, J.I. Hualde, M. Simonet, and F. Torreira for very useful feedback. We are indebted to our Catalan and Spanish speakers (M. Nadeu, M. Magrans, M. Bosch, M. Albert, A. García, M.P. García). Finally, we would like to thank M. Nadeu for her help in recording and performing the segmentation of some of the speakers. This research was funded by grant 2005SGR-00753 from the Generalitat de Catalunya, and by grants HUM2006-01758/ FILO and CONSOLIDER-INGENIO CSD2007-00012 from the Spanish Ministry of Science and Education.
and in unrelated languages. Contour tones tend to manifest themselves in contexts that are longer, while shorter contexts tend to produce neutralization. For example, it has long been noted that contour tones in Chinese languages such as Mandarin, Cantonese, and Fuzhou appear with fewer restrictions in language varieties that have coda contrasts, diphthongs, and longer rhyme durations, that is, in bimoraic syllables (Duanmu 1990, 1994a, b). Similarly, in Tokyo Japanese (unlike Kansai Japanese), a contour tone cannot be realized on a single mora (Maeda & Venditti 1998).\(^1\) Nonetheless, the strictly moraic approach has been challenged by Zhang (2001, 2004), who points out that contour tones appear cross-linguistically in syllables which are long for independent reasons and which do not need to be bimoraic, e.g., syllables at the end of prosodic domains, syllables in shorter words, etc., and argues that the tone-bearing ability is rooted in phonetic behavior. In the realm of perception, experiments by Diehl and Kluender (1989) have revealed that high tones in Chinese may be misperceived as rising tones when syllables are longer.

In pitch-accent languages, it has also been observed that when complex pitch gestures appear as a result of two or three associated tones with a single syllable, either the contour is fully realized and ‘compressed’, or certain repair strategies appear, such as contour truncation (for examples of these different strategies, see Ladd 1996:132–136; Grabe 1998; and Grabe et al. 2000). Truncation and compression are two distinct phonetic realization strategies: while compression generally involves a speeding up of the pitch realization in order to produce a complete accent shape, truncation involves no pitch velocity change in the contour, which is only partially realized (see Grabe et al. 2000:162). In compressing languages, a few studies have observed that syllables bearing a complex pitch accent are longer than syllables bearing simpler pitch accents. For example, Gili-Fivela (2006) reports that syllables in contrastive focus (e.g., rise-fall gestures) are 7% to 10% longer than syllables in broad focus (e.g., rise gestures), regardless of their position within the word. Similarly, Ortega-Llebaria and Prieto (2006) recently investigated the durational properties of words with broad vs. narrow contrastive focus in Catalan and Spanish. They found that duration was amplified in narrow-focused words although only in final position. Since it is only in final position that we have a complex tonal gesture, the working hypothesis is that complex tonal gestures trigger lengthening. Crucially, durational differences only appeared between narrow vs. broad focus in words with final stress, a context where we find the pressure of realizing a complex rising-falling tonal gesture only in narrow focus. In contrast,

\(^1\) On the other hand, in Cantonese, contour tones have a lengthening effect: specifically, there is a statistically significant difference between the duration of a level tone and that of either a morphologically-derived or a sandhi-derived rising tone (Yu 2003).
the lack of duration effect in words with penultimate stress is attributed to the lack of realization of the complex pitch gesture.

The goal of this paper is to systematically investigate the effects of the presence of complex pitch movement on the durational properties of syllables in Catalan and Spanish and to test whether the abovementioned restrictions are rooted in phonetic behavior. Crucially, in both languages, narrow contrastive focus and presentational broad focus have different pitch accent choices (e.g., *She broke her neck, right? – No, she broke her LEG*) (see de la Mota 1995; Face 2002 for Spanish; Prieto 2002; Estebas-Vilaplana 2000; Astruc-Aguilera 2006 for Catalan). As Figure 1 shows, in Catalan narrow focus is realized as a rising accent and broad focus as a falling accent. Interestingly, when the accented syllable is in phrase-final position (*mamá*), the rising accent is realized as a complex rise-fall gesture, as follows.  

![Figure 1. Schematic representation of the realization of the nuclear pitch accent in broad focus and in narrow contrastive focus utterances, in words with penultimate stress and final stress.](image)

This contrast will enable us to test the hypothesis that the presence of a complex pitch gesture in words with final stress in narrow focus will trigger an extra amount of lengthening in this syllable. The present article presents an analysis of 1280 utterances produced with the abovementioned contrasts. The analysis revealed the following: (a) in cases where the complex contour is fully realized, a clear difference is revealed between the duration properties of syllables carrying a narrow focus accent in final position vs. penultimate position; this is attributed to the pressure of realizing a complex tonal gesture only in final position (see Figure 1); (b) in cases of truncated (or partially truncated) contours, there is no clear lengthening effect. The potential implications of these results for cross-linguistic work on the constraints and strategies of contour compression and truncation will be discussed.

---

2. Both in Spanish and Catalan we have two possible words meaning 'Mum', namely, *mamá* and *mama* in Spanish, and *mamà* and *mama* in Catalan.
The article is organized as follows. Section 2 presents the methodology of the production experiment. Section 3 presents the main results of pitch range and duration of the production experiment. Finally, Section 4 discusses the main implications of this work for cross-linguistic studies on the interaction between pitch realization and duration.

2. Experimental investigation

2.1 Method

2.1.1 Materials. In both Catalan and Spanish, broad focus has been described as having a different prosodic realization from narrow contrastive focus (e.g., *She broke her neck, right? – No, she broke her LEG*) (see de la Mota 1995 and Face 2002 for Spanish; Prieto 2002; Estebas-Vilaplana 2000; Astruc-Aguilera 2006 for Catalan; for Romanian, see Manolescu, Olson & Ortega-Llebaria 2009, this volume). In Spanish and Catalan, nuclear broad focus is typically realized with a falling nuclear pitch accent. In contrast, the nuclear narrow contrastive focus is realized with a non-downstepped rising pitch accent, as Figure 2 shows. While the narrow contrastive pitch accent is generally transcribed as L+H*, there is no consensus as to what is the phonological analysis of the falling nuclear pitch accent in both languages. Some analyses such as Astruc-Aguilera (2006) for Catalan and Beckman et al. (2002) for Spanish have proposed the H+L* nuclear pitch accent, identifying this pitch accent with the broad focus nuclear pitch accent of European Portuguese (Frota 2002) and of Italian (D’Imperio 2002). On the other hand, Sosa (1999) transcribes it as having an L* nuclear accent – see the discussion about this topic in Beckman et al. (2002). Yet the main contrast between the two accent types is upheld under any of the analyses that seem tenable.

In order to test the hypothesis that the presence of a complex pitch gesture in words with final stress with narrow-corrective focus will trigger an extra amount of lengthening, we planned a controlled production experiment comparing near-minimal-pair words bearing the two types of pitch accents. A previous study with similar speech materials for Catalan and Spanish (Ortega-Llebaria & Prieto 2006) found that durational differences between stressed syllables in narrow vs. broad focus were only found in words with final stress, and not in words with penultimate stress. This contrast was attributed to pressure on the speaker to realize a complex tonal gesture only in narrow focus. Yet these materials did not control segmental content, which is done in the present experiment. The sentences under study were the following: (1) a broad-focus utterance (e.g., *Catalina me desanimó ‘Catalina discouraged me’*), and (2) a narrow-contrastive focus (e.g., *¿Catalina te animó?* “Catalina encouraged you?”).
'Did Catalina encourage you?' – No, Catalina me DESANIMÓ ‘No, Catalina DISCOURAGED me’).

To enhance the control for vowel and consonantal effects on duration (i.e., segmental effects), the target words selected are four two-syllable nouns that have the same segmental composition and that contrast only in the stress position in both languages: Spanish/Catalan mamá ‘Mum’ vs. mama ‘Mum’, and Mimí ‘proper noun’ vs. Mimi ‘proper noun’. Each target word was placed in: (1) a broad-focus utterance (e.g., Span. ¿Qué pasa? – Se lo dice a mama ‘(S)he is telling Mum’); and (2) a narrow-contrastive focus, which is indicated in capital letters (e.g., Span. ¿Se lo dice a mamá? – No, se lo dice a MAMA ‘No, (s)he is telling MUM’), as follows:

Catalan

/ə/  /i/  
L’hi mano a la mamà L’hi mano a la Mimi
‘I am giving it to Mum’ ‘I am giving it to Mimi’
L’hi dono a la MAMA L’hi dono a la MIMI
‘(No), I am giving it to MUM’ ‘(No), I am giving it to MIMI’
L’hi dono a la mamà L’hi dono a la Mimi
‘(No), I am giving it to Mum’ ‘I am giving it to Mimi’

As mentioned above, there are two possible words in Spanish and Catalan for ‘Mum’, each of which have two different stress patterns, namely, mamá and mama in Spanish, and mamà and mama in Catalan. Even though both forms are used, mamá is more common in Spanish, and mama in Catalan.
These nouns in two focus contexts were inserted in 20 frame sentences with different verb types (see Appendix) and placed in sentence-final position.

2.1.2 Experimental procedure. Subjects were asked to look at a Power Point presentation comprising 80 slides (20 utterances x 2 words x 2 focus types) which contained suitable contexts for triggering an utterance with either broad or narrow focus, together with the answers, as follows:

**Broad focus**

Context: — ¿Qué pasa?
‘What is going on?’
— L’hi dono a la mamà
‘I am giving it to Mum’

**Narrow-contrastive focus** (with emphasis and assertiveness)

Context: — ¿L’hi dones a la mamà?
‘Will you give it to “mamà”?’
— No, l’hi dono a la MAMA
‘No, I will give it to “mama”’

First, the experimenter read the context question aloud to the subject. The subject was supposed to read out the answer in the appropriate intonation. If the experimenter thought that the utterance had been mispronounced, the speaker was asked to repeat the sentence. This happened in only a handful of cases, and not for every speaker. The same process was repeated with each one of the slides.

Speakers were recorded individually in a quiet room, using a Sennheiser MKH20P48U3 omnidirectional condenser microphone and a Pioneer PDR609 digital CD-recorder. Speech samples were digitalized at 32000 Hz in 16-bit mono, double-checking that the target utterances were produced with the intended prosody.
2.1.3 Subjects. Four young female speakers of Central Catalan and four young speakers of Peninsular Spanish (between 23 and 40 years of age) participated in the experiment, giving a total of 160 utterances per speaker (20 verb types x 2 focus conditions (broad vs. narrow) * 4 target words stress conditions (mama, mamá, mimi, mimí) = 160 utterances per speaker. Thus, we obtained a total of 1280 utterances (160 utterances x 8 speakers = 1280 utterances).

2.2 Data analysis and measurements

The following measurements were made with Praat (Boersma & Weenink 2005; Wood 2005) on each of the 1280 target words. Figure 2 shows the labels that we used to segment the target words. In tier 1, we marked the beginning and end of the target segments, for example, ‘m’, ‘a’, ‘m’ and ‘a’. In tier 2, we marked the valley (L) and the peak (H) of the pitch accent. In the cases where the pitch line was flat or descending, marks were placed at the beginning and end of the syllable. A Praat script extracted the F0 value in Hz at the marked points and calculated the pitch excursion size by subtracting the F0 values at L from the F0 values at H for each of the 1280 tokens.

2.3 Statistical analysis

In order to ensure that our data did indeed include two types of accents with different intonation properties, a Repeated Measures ANOVA with the within-subject factor of intonation (broad focus vs. narrow focus) and the between-subject factor of language (Catalan and Spanish) was performed on the pitch-range measurements. After this, we performed a Repeated Measures ANOVA on the duration of stressed vowels with three within-subject factors: stress position ([+word-final]/[-word-final]) and accent-type (broad focus vs. narrow focus) and vowel (i, a), and the between-subject factor of language (Catalan and Spanish). Since the vowel was not significant, we collapsed our data across vowels.

3. Results

The main research question posed by the study is whether syllable duration is amplified in contrastive narrow-focused words with respect to broad-focused words. That is, we investigate whether the presence of a complex F0 gesture is accompanied by an increase in duration.
3.1 Pitch excursion size differences

In this section, we check that the two types of focus (broad focus and narrow focus) in nuclear position were indeed realized using two different pitch accents, namely, a falling pitch accent and a rising pitch accent. The boxplots in Figure 3 show the mean pitch distance (in Hz) between LH values (in the case of a rising pitch movement) and between HL values (in the case of a falling pitch movement) of stressed syllables in broad-focused and narrow-focused conditions (striped boxes vs. dotted boxes, respectively) for all four Catalan speakers and all four Spanish speakers. It should be remembered that in cases in which the pitch was descending and no peak and valley could be visually identified, the pitch measures were taken at the beginning and at the end of the target accented syllable. As is clear from the graph, both Catalan and Spanish subjects consistently used a substantial pitch increase in narrow-focused sentences (a mean of 69.20 Hz, s.d. 28.47 in Spanish and 56.82 Hz, s.d. 45.04 in Catalan) and a negative increase in the broad-focus case (i.e., a mean of -33.65 Hz, s.d. 17.16 in Spanish and -28.36 Hz, s.d. 14.01 in Catalan).

If we plot the data separately by speaker (see Figure 4 below), the same pattern emerges for each one of the subjects, namely, speakers produce pitch accents with larger pitch ranges in narrow focus than in broad focus. While this difference is maintained, speakers also show some variation in the amount of pitch range values, especially in the narrow-focus case. For example, Spanish speaker (MB) has a mean pitch range of 34.93 Hz, while Spanish speaker (MN) has a much larger mean pitch range of 81.60 Hz.
Complex pitch gestures induce syllable lengthening

A Repeated Measures ANOVA on the pitch range of stressed syllables revealed highly significant effects of the accent factor on F0 variation (measured as pitch excursion size), at $F(1,587) = 989.798; p<.0001$ and no interaction between the accent*language factor, at $F(1,587) = 1.676; p = .196$. Hence, as expected, narrow focus was consistently cued by a rising pitch accent in the two languages. In contrast, broad focus in both languages was cued by a falling nuclear pitch accent.
3.2 Durational differences

Figure 5 below displays the mean duration (in ms) of the target syllable in penultimate (upper graph) and word-final (lower graph) position in Catalan and Spanish across vowels. In all four plots, the duration of the stressed syllables is compared between the narrow-focused (dotted boxes) and broad-focused (striped boxes) words. The graphs reveal that, in general, stressed syllables in narrow-focused words are longer than in broad-focused words in both languages, in both penultimate and word-final position (mean differences in penultimate and final position for Catalan: 39 ms and 55 ms; for Spanish: 69 ms and 207 ms). In general, the data works as expected: we find greater duration values in narrow-focused than

![Graph showing duration values for Catalan and Spanish](image)

**Figure 5.** Mean duration values (in ms) of the target syllable in penultimate (upper graph) and word-final (lower graph) position in broad focus (dashed boxes) and narrow focus (dotted boxes), for both Spanish and Catalan
in broad-focused syllables both in final and penultimate positions and in both languages (that is, dotted boxes are always to the right of striped boxes). Yet the graphs also show that there is a contrast between Spanish and Catalan. While Spanish speakers produce an extra amount of lengthening in narrow-focused words in final position (see lower graph), this does not seem to be the case for Catalan speakers (mean and standard deviation for narrow focus in Spanish 437.5 ms (105.5) vs. 281.5 ms (33.7) in Catalan).

The boxplots in Figure 6 depict the same data for each one of the four Spanish subjects. In general, it is very clear that all Spanish speakers display substantial durational differences between narrow- and broad-focused syllables, both in penultimate

![Figure 6. Mean duration values (in ms) of the target stressed syllables in broad focus (dashed boxes) and narrow focus (dotted boxes) in Spanish, plotted separately for all four speakers. Word-final syllables are plotted in the upper graph and penultimate syllables in the lower graph.](image)
and word-final position, and, importantly, there is a greater difference between the two when in word-final position, that is, when a complex tonal pattern is realized on the target syllable. For example, for speaker AG, the mean difference in duration between segments in broad focus and segments in narrow focus is 386 ms in words with final stress, while this difference falls to 67 ms in words with penultimate stress, thus showing that the lengthening effect of narrow focus with respect to broad focus is 316 ms longer in words with final stress than in words with penultimate stress. For speaker MO these mean differences are 180 ms and 96 ms respectively, and for speaker MP they are 204 ms and 92 ms, showing that the lengthening effect of narrow focus in words with final stress is 84 ms longer than that of words with penultimate stress for speaker MO, and 112 ms for speaker MP. For speaker MA, the lengthening effect of narrow focus in words with final stress is only 56 ms longer than in words with final stress since mean differences between narrow and broad focus in both conditions are 76 ms and 20 ms respectively.

Yet for Catalan subjects, subject differences may be found (see boxplots in Figure 7): while subjects MN and PP have an extra lengthening effect in narrow-focused syllables only of words with final stress, speakers MB and MM do not display this difference. Thus, similarly to Spanish speakers, for Catalan speakers MN and PP, the lengthening effect of narrow focus in words with final stress is around 40 ms longer than that of narrow focus in words with penultimate stress (mean differences between narrow and broad focus for words with final and penultimate stress for MN: 87 ms and 44 ms, for PP: 40 ms and 2 ms). However, this lengthening effect decreases to less than 10 ms for speaker MM (40 ms in words with final stress and 31 ms in words with penultimate stress) and even shows the opposite direction for speaker MB (52 ms in words with final stress and 83 ms in words with final stress).

We performed a Repeated Measures ANOVA on the duration of stressed vowels with two main within-subject factors: Position ([+word-final]/[-word-final]) and Accent-type (narrow focus vs. broad focus) and vowel (i, a), and the between-subject factor of language (Catalan and Spanish). Both Accent type and Position have a significant effect on the duration of the target syllables (Accent type: F(1,318) = 910.65, p<.0001; Position: F(1,318) = 793.22, p<.0001) corroborating the hypothesis that narrow focus extends the duration of segments and that this lengthening effect is larger in words with final stress. There was no significant vowel effect, thus indicating that lengthening effects are similar in vowels [a] and [i]. Importantly, there is a significant interaction between Accent type*Language (F(1,318) = 910.65, p<.0001), meaning that the lengthening effect of narrow focus is larger in Spanish than in Catalan. In the next section, we investigate possible
Complex pitch gestures induce syllable lengthening

When searching for sources of variation in the data, we noticed that several of the complex pitch accents that appear in narrow-focused words in utterance-final

Figure 7. Mean duration values (in ms) of the target stressed syllables in broad focus (dashed boxes) and narrow focus (dotted boxes) in Catalan, plotted separately for all four speakers. Word-final syllables are plotted in the upper graph and penultimate syllables in the lower graph

sources of variation that can explain the distinctive and surprising behavior of these two Catalan speakers.

4. Discussion and conclusion: Sources of variation

When searching for sources of variation in the data, we noticed that several of the complex pitch accents that appear in narrow-focused words in utterance-final
position were truncated. Figure 8 shows the waveforms, spectrograms, and F0 contours of the Catalan utterance *No, l’hi mano a la MAMÀ* ‘I am asking this to MUM’ as uttered by two different speakers. In the left panel we see a fully compressed contour while the right panel shows a truncated contour.

![Figure 8](image)

**Figure 8.** Waveforms, spectrograms and F0 contours of two possible prosodic realizations of the Catalan utterance *No, l’hi mano a la MAMÀ* ‘No, I am asking this to MUM’. Left panel shows a fully compressed contour and right panel a truncated contour.

We hypothesize that, in our data, unexpected subject differences in durational patterns might be attributed to the presence of truncated contours. Figure 9 shows the mean relative truncation of the final F0 value (in Hz) for the four Catalan speakers (lower panel) and the four Spanish speakers (upper panel). This value was calculated as the distance in Hz from the final F0 value in each of the narrow-focused contours to the reference baseline of the speaker (that is, the bottom of the speaker’s pitch range, a value that is obtained at the end of broad focus statements) so that higher values correlate with truncation and lower values with fully realized pitch contours. The graphs show that speakers use different degrees of truncation and further support a view of truncation as a gradient acoustic effect. In general, Spanish speakers produce fewer truncated contours than Catalan speakers. Within the Spanish speakers, MA realizes full contours less often than the other speakers. And, among the Catalan speakers, MB and MM truncate contours more often. Interestingly, if we compare this data with the graphs in Figure 7, we can see that the Catalan subjects with more truncation (MB, MM, and MN) are exactly the ones who display less lengthening in narrow-focused words in final position. Similarly, the Spanish speaker who used full contours less often also showed the shortest lengthening effect of narrow focus in word-final position.

The scatter plot in Figure 10 shows the duration of the final syllable (in ms) in narrow focus contexts as a function of the degree of truncation (in Hz). For each speaker, the ‘degree of truncation’ was calculated relative to the speaker’s baseline.
Results show that there is a slight negative correlation between the two factors: in general, the more truncated the contour is, the shorter the accented syllable. Yet, we should be cautious in interpreting these results because this effect is not very strong and the data seems to be grouped in three different data sets. Linear

Figure 9. Mean truncation of the final F0 value (in Hz) in narrow-focus utterances relative to the baseline for all Spanish speakers (upper panel) and all Catalan speakers (lower panel)
regression analyses show that truncation alone explains 15% of the variation in the syllabic duration of oxytone words with complex tones.

![Figure 10. Duration of the final syllable (in ms) in narrow focus contexts as a function of the degree of truncation (in Hz), calculated relative to the baseline, for all speakers](image)

In sum, the results of our production experiment reveal that in both Catalan and Spanish, duration is amplified in narrow-focused words, whether in penultimate or final position. In general, duration is also amplified in syllables with a complex tonal gesture. Figure 11 shows the mean duration of word-final target syllables in different conditions in our data for the two languages: [–stress, –complex pitch gesture], [+stress, –complex pitch gesture], and [+stress, + complex pitch gesture]. It is clear that once we control for final lengthening factors, we have to consider two factors that affect duration and that combine in an additive fashion: (a) stress; and (b) tonal complexity.

In our data, the presence of a complex pitch contour is the strongest lengthening factor. Multiple Comparisons with the Bonferroni adjustment show that the duration differences between these three types of syllables ([–stress, –complex pitch gesture], [+stress, –complex pitch gesture], and [+stress, + complex pitch gesture]) are statistically significant. The mean duration value and standard
Complex pitch gestures induce syllable lengthening

deviation are the following: [–stress, –complex] 206.5 ms (31.1), [+stress, –complex] 228.6 ms (37.5), and [+stress, +complex] 359.5 ms (110.5).

In sum, truncation is used by some speakers as a phonetic realization strategy that interacts with timing. As we have just seen, some Catalan speakers truncate (or partially truncate) the complex contour: it is precisely these speakers who do not display the expected amount of extra lengthening in these syllables. Thus, in our data, the truncation of complex contours can be regarded as a phonetic realization strategy that interacts with timing in such a way that there is a trade-off relationship between the two factors. The observed phenomenon has consequences for cross-linguistic work on tonal realization strategies, namely, truncation and compression. Different studies have shown that there are cross-linguistic differences in the application of truncation and compression in standard varieties of English and German, and cross-dialectal differences within Swedish and Danish (see Ladd 1996; Grabe 2008; Grabe et al. 2000, among others). For example, while speakers of Cambridge English and Newcastle English compress rising and falling accents, speakers of Leeds English, in identical contexts, perform truncation. The data in this article challenge the view that truncation and compression are language and dialect-specific strategies (Ladd 1996, Grabe et al. 2000), and favor the view that they have to be regarded as phonetic realization strategies that interact in a dynamic way with timing.

Figure 11. Mean duration of word-final target syllables in different conditions: [–stress, –complex pitch gesture], [+stress, –complex pitch gesture], and [+stress, +complex pitch gesture]
References


Manolescu, Alis, Daniel Olson & Marta Ortega-Llebaria. 2009. “Cues to Contrastive Focus in Romanian”. This volume.


Appendix

TARGET VERBS (CATALAN)  TARGET VERBS (SPANISH)
L’hi mano a la mama       Se lo mando a mama
L’hi cuso a la mama       Se lo coso a mama
L’hi dono a la mama       Se lo vendo a mama
L’hi porto a la mama      Se lo debo a mama
L’hi torno a la mama      Se lo hago a mama
L’hi bullo a la mama      Se lo ruego a mama
L’hi bato a la mama       Se lo nombre a mama
L’hi poso a la mama       Se lo pongo a mama
L’hi dicto a la mama      Se lo dicto a mama
L’hi envio a la mama      Se lo envio a mama
L’hi pago a la mama       Se lo noto a mama
L’hi deixo a la mama      Se lo pago a mama
L’hi rento a la mama      Se lo pido a mama
L’hi cuino a la mama      Se lo leo a mama
L’hi busco a la mama      Se lo busco a mama
L’hi pinto a la mama      Se lo lavo a mama
L’hi brodo a la mama      Se lo llevo a mama
L’hi ballo a la mama      Se lo dejo a mama
L’hi tallo a la mama      Se lo presto a mama
L’hi baixo a la mama      Se lo guiso a mama
Cues to contrastive focus in Romanian

Alis Manolescu, Daniel Olson & Marta Ortega-Llebaria
University of Texas at Austin

In this study we measured patterns of pitch alignment, pitch range and duration in relation to broad and contrastive focus in Romanian. In declarative sentences with broad focus, speakers place a pitch accent on each lexically stressed syllable with peaks that become progressively lower towards the end of the sentence. In pre-nuclear accents, the peaks align with the post-tonic syllable. In declarative sentences with contrastive focus, speakers use strategies based on pitch and duration in order to build a maximum contrast between the word under focus and those in pre- and post-focal contexts: an expanded pitch range under focus and a reduced pitch range and shorter stressed syllables in pre- and post-focal contexts. Thus, the flat F0 and shorter segmental durations in pre- and post-focal contexts constitute a background that, in contrast, highlights the segmental durations and expanded pitch ranges found under contrastive focus.

1. Introduction

To our knowledge, there are only a few impressionistic studies on the intonation patterns of declarative sentences in broad (BF) and contrastive focus (CF) in Romanian (Dascalu-Jinga 1998; Winkler-Gobbel 2002; Swerts 2007). Dascalu-Jinga (1998) provides a descriptive overview of Romanian intonation contours using the INTSINT transcription method (Hirst & DiCristo 1998), which shows that the basic broad declarative pattern is a rising-falling one with a declination pattern apparent in longer declaratives. In the case of contrastive focus which may affect any item of an utterance, there is a positive prominence expressed by a high and/or rising pitch on the stressed syllable of the word under focus. Winkler-Gobbel (2002) uses the AM model of tonal transcription to claim that in BF utterances, syntactic arguments are associated with bitonal accents (L+H* and H+!H*), whereas verbs may either be de-accented or associated with the default H* accent. Winkler-Gobbel’s (2002) primarily syntactic analysis of p-movement shows that BF utterances may contain defocused material such as in English or German, namely that there is evidence for contextual de-accenting of an internal argument which does not give rise to a narrow focus interpretation.
However, Swerts’s (2007) empirical study refutes Winkler-Gobbel’s results by providing evidence that, as in Italian or Spanish, Romanian also resists de-accentuation inside syntactic constituents. Moreover, he observes some cases in which complex noun phrases consisting of an adjective and a noun are completely unaccented. These cases always occur on the first NP in the sentence whereas a final NP almost always receives a single accent on the second focalized word while the first word is de-accented. According to the author, these de-accentuation patterns serve a demarcative function, in that they mark the right edge of a speech unit and cannot be explained on the basis of contrast relations.

This study investigates the intonation patterns of BF and CF in Romanian in declarative sentences with a relatively simple syntactic structure, namely SVO sentences with a subject and object NP and a VP, all consisting of one single constituent. We expect to provide a detailed phonetic description of the pitch contours and segmental durations linked to pitch accents in BF and CF as well as shed further light on the controversial questions of de-accentuation in Romanian.

Although there is controversy about the definitions of broad and contrastive focus (Bolinger 1958; Gussenhoven 1984, among others), we may define broad focus as a carrier of new information where the whole constituent or sentence is previously unknown. Contrastive focus, on the other hand, highlights a subset of the information through a contrast, which implies the exclusion of contextually relevant alternatives. For example, when the sentence ‘Mary is coming’ is pronounced as an answer to the question ‘What’s happening?’ the entire sentence is new information with no specific element being emphasized. However, when the same sentence is an answer to the question ‘Is Peter coming?’, ‘Mary’ is highlighted by contrastive focus.

Focus has been shown to be marked by means of intonation and syntactic variation. Syntactically, focus can be indicated by word order variation, as in Italian or Spanish, whereas such scrambling is not possible in languages with a fixed word order, such as English. However, when word order is invariable between broad and contrastive focus utterances in languages with free word order, as in the present experiment, speakers use phonetic strategies to distinguish these two types of focus. Romanian allows this sort of distinction between broad and contrastive focus: while the sentence structure stays the same for the two conditions, a change in intonation indicates a difference in the pragmatic interpretation, as in (a) and (b):

a. broad focus declarative [What’s happening?]
   Maria vine. ‘Mary is coming’

b. contrastive focus declarative [Is Peter coming?]
   MARIA vine. ‘It is Mary who is coming.’
Based on patterns from other romance languages, we expect that the interpretations of broad and contrastive focus in Romanian are conveyed by manipulating pitch alignment, pitch range and segmental durations. In Spanish (Face 2002), Italian (D’Imperio 2002), Portuguese (Frota 2002), the peak of BF pitch accents in pre-nuclear position aligns with the post-tonic syllable, while in CF, the peak is on the stressed syllable. This contrast has been analyzed by some scholars (for example, Beckman et al. 2002 for Spanish) as a phonological contrast between two pitch accents, a late rise L*+H for broad focus and an early rise L+H* for contrastive focus. For European Portuguese, this distinction is marked by an H*+L accent for the focalized word, which contrasts with the H+L* counterpart in the broad declarative utterance, with a similar distinction for Neapolitan Italian (L+H* vs H+L*, in D’Imperio 1997) and Standard Italian (H* vs H+L*, in Avesani & Vayra 2000). Nevertheless, Face (2002) has shown that in Spanish either an L+H* accent or the L*+H pitch accent can be used in BF, the latter accompanied by boundary tones following the contrasted element (H-, L- in the AM model) and a higher F0 peak height.

The existence of an actual F0 pitch range increase for Spanish is highly controversial, with studies that suggest that it is not an acoustic correlate of contrastive focus (Face 2000, 2002), as against those which claim that it has a significant role in marking focus by an acoustically more salient accent (De la Mota 1995, 1997). In other languages such as Neapolitan Italian (D’Imperio 2002: 57), a broad focus utterance is characterized by “a relatively shallow F0 variation as opposed to the greater F0 excursion within the narrow focus.” Furthermore, it has been shown that, as a correlate of the tonal complexity associated with narrow focus, duration also serves as a cue to narrow focus (Prieto & Ortega-Llebaria 2009, this volume).

Duration has also been found to be a relevant phonetic cue to focus. De Jong (2004) discusses the effect of “localized hyperarticulation”, through which elements of the speech signal are emphasized in the duration contrast between stressed and unstressed syllables. Empirical studies show that stressed vowels are longer than their unstressed counterparts, as in Dutch (Sluijters & van Heuven 1996), in English (Beckman & Edwards 1994) and Italian (Marotta 1995, Kori & Farnetari 1983) among others. This “magnifying effect” extends to focalized contexts, in that contrastive focus elements expand their duration when stressed. Several studies support these claims: Face (2000) and De la Mota (1995, 1997) for Spanish among others. In her discussion of focus in Dutch, Chen found both an increase in duration as well as a change in the F0 contour, however acquisition of the durational cues occurred later than the F0 cues (Chen 2009, this volume).
It is apparent from the previously-mentioned studies that languages employ several strategies to convey the pragmatic opposition between broad and contrastive focus, although languages differ in their employment of these strategies. The Romance languages in particular show variations in their use of pitch alignment, pitch range increase, and increased vowel duration in distinguishing BF and CF. Our investigation of these acoustic cues for Romanian will add to the current body of research as well as to the understanding of the principles and variation of pan-Romance intonation. The remainder of this paper is organized as follows. Section 1 describes the methodology used for the production experiment. In Section 2, we present our results, the main effects of focus condition and stress on the measured variables, namely F0 pitch range, F0 peak alignment and stressed vowel duration. Finally, Section 3 presents and discusses our results, highlighting the acoustic correlates of contrastive focus in Romanian.

2. Methodology

2.1 Subjects

Ten female native speakers of Romanian (ages 20–30) recruited in Sibiu, Romania were recorded for this study. The subjects had been living in Sibiu for a period of at least 5 years. Their native tongue was Romanian and they did all their schooling in Romanian including their university education. They spoke Romanian with their parents, siblings and family, speaking and having been educated in the standard variety of Romanian. They had neither studied nor lived abroad for a period of time longer than a few weeks (considered travel/holiday time). They reported having normal speech and hearing.

2.2 Materials

The experiment used the same set of sentences spoken in two different intonations, broad declarative intonation (BF) and contrastive focus intonation (CF). Each utterance had a BF condition and 3 CF conditions, one CF for each of the 3 lexical constituents. The lexical constituents had 2 syllables each, controlling for paroxytonic and oxytonic stress. For example, O mama vinde mere has paroxytonic stress on each word, while in Dorel veđa maiori, words have oxytonic stress. In both stress patterns, the number of intervening unstressed syllables was constant. The distinction between the oxytonic and paroxytonic stress patterns was controlled in order to compensate for a possible word boundary crowding effect on the realization of the F0 peak. A total of 8 sentences used in the study may be seen in Table 1 below, giving a total of 320 utterances in the study (10 subjects x 8 sentences x 4 conditions).
Table 1. The eight sentences used in the study

| O mama vinde mere. | “A mother sells apples.” |
| Un mire vede marea. | “A groom sees the sea.” |
| O nora vede norul. | “A daughter-in-law sees the cloud.” |
| Un rege linge mierea. | “A king licks the honey.” |

| Dorel vedea maiori. | “Dorel was seeing mayors.” |
| Ionel dorea marar. | “Ionel wanted dill.” |
| Ninel vindea aluni. | “Ninel was selling hazelnut trees.” |
| Marian lingea magiun. | “Marian was licking preserves.” |

All sentences are declarative utterances. Each one was spoken with a broad or contrastive focus intonation. In order to elicit broad focus intonations, the subjects were instructed to read a statement presented to them as an answer to the question ‘What’s happening’ or ‘What happened’. To elicit contrastive focus, speakers were asked questions by the first author where one of the constituents in the sentence was replaced with another word. For example, after the subject had read the broad focus declarative ‘O mama vinde mere’, they were asked ‘O sora vinde mere?’ and they answered ‘Nu, o MAMA vinde mere’ with the contrastive focus on MAMA (as seen in Table 2). Since each sentence had three words, contextualizing questions were formulated so as to place each one of them under contrastive focus.

Table 2. Sample of a declarative sentence with broad focus and contrastive focus. The questions asked to elicit these declarative sentences are in italics while words under contrastive focus are in bold letters

<table>
<thead>
<tr>
<th>Utterance read:</th>
<th>Broad focus:</th>
</tr>
</thead>
<tbody>
<tr>
<td>O mamă vinde mere.</td>
<td>“A mother sells apples.”</td>
</tr>
</tbody>
</table>

Contrastive Focus:

| Question1: | O sora vinde mere? |
| Response—Focus word 1: | O mama vinde mere. |

| Question2: | O mama cumpara mere? |
| Response—Focus word 2: | O mama vinde mere. |

| Question3: | O mama vinde pere? |
| Response—Focus in word 3: | O mama vinde mere. |

2.3 Recordings

The declarative utterances were randomized and presented to the subjects. They were instructed to read the broad focus declaratives displayed on flashcards. Based on the information given in the broad focus utterance, the first
author posed 3 questions that triggered responses with contrastive focus on each of the lexical constituents of the broad focus declarative (see Table 1 for examples). In line with other previous intonation research, this methodology was designed to create a corpus that closely resembles natural exchange. No other specific instructions as to the nature and purpose of the experiment were given the informants.

The recording of the utterances was performed using a vacuum-tube microphone and the Praat software. Each utterance was isolated in the Praat software and partitioned in syllables as well as their respective vocalic and consonantal constituents.

2.4 Measurements

After marking all the syllables in each sentence by looking at spectrograms and F2 movements in Praat, we measured durations, pitch range and pitch alignment. Due to pitch track failure as a result of a creaky voice at the end of the sentence, the pitch changes that took place in the third constituent (word 3) of the BF sentences could not be measured.

2.4.1 Duration. Syllables were segmented in vowels and consonants. The segments were manually labeled, with particular attention being paid to the second formants. A sample of the data was cross-checked by a second researcher to ensure consistency. A Praat script extracted the durations of all the vowels in milliseconds, as well as the duration for the syllable rhymes.

2.4.2 Pitch range. F0 pitch range (Hz) was calculated by subtracting the minimum F0 value (valley) from the F0 subsequent maximum value (peak), as seen in Figure 1 below. The peak and the valley are associated with the stressed target syllable – potential recipient of the pitch accent for that specific lexical word. Using a Praat script, the values for the F0 peak and valley were extracted and marked automatically.

![Figure 1](image)

**Figure 1.** F0 peak and valley associated with the stressed syllable. In the broad declarative utterance ‘O mama vinde mere’ ‘A mother sells apples’, each stressed syllable has a pitch accent. For example the stressed ‘ma’ is associated with a valley and a peak
2.4.3 *Pitch alignment.* We defined pitch alignment as the distance from the peak of the pitch accent to the syllable boundary. In order to normalize duration, we divided the distance from the vowel onset to the F0 peak (d1 in Figure 2) to the total rhyme duration (d2 in Figure 2), calculated as a percent of the rhyme duration. Thus, peak alignment results appear as d1/d2 x100 percentage values. A value under 100 shows that the peak is aligned within the stressed syllable while a value higher than 100 is related to a post-tonic alignment.

![Figure 2](image)

**Figure 2.** Example of the F0 pitch track with d1 and d2 measurements. d1 is the distance from the vowel onset to F0 peak and d2 represents the total rhyme duration measured from the vowel onset to the end of the syllable

2.5 Statistics

We compared the measurements of duration, pitch range and pitch alignment in each of the sentence constituents in the BF declaratives with those in the CF sentences. For the alignment data, the last accented word of the BF and CF utterances (i.e. nuclear accent) was not considered for this study because speakers tended to have a creaky voice at the end of the sentence preventing the extraction of accurate pitch values. We compared the pitch range of the first constituent, i.e. the first word in the sentence when the sentences are produced with a broad focus intonation (Figure 3a), with cases when there was a contrastive focus on this constituent (Figure 3b), and when this same constituent was pre-focal in sentences where the contrastive focus was placed on the second or third constituent (Figure 3c). Similar comparisons were performed for vowel duration and pitch alignment, whereby the latter did not have comparisons for the nuclear accent, due to the reasons described above.

ANOVAS with factors of stress (oxytone and paroxytone words) and sentence intonation (broad focus declaratives and 3 contrastive focus conditions, one for each sentence constituent) were performed on each set of measurements on each word. Post-hoc tests were performed on each significant factor. For the alignment results, a paired t-test was performed, comparing the alignment for broad focus and contrastive focus in the two stress conditions (oxytone and paroxytone).
Figure 3a. Sentence ‘O mama vinde mere’ in broad focus. ‘Mama’ bears a pitch accent

Figure 3b. Sentence ‘O MAMA vinde mere’, with contrastive focus on MAMA

Figure 3c. Sentence ‘O mama VINDE mere’, contrastive focus on ‘vinde’ and de-accentuation on ‘mama’

3. Results

3.1 Pitch range

The graph in Figure 4 shows the mean pitch ranges of the pitch accents placed on the first (in dotted pattern), second (in black), and third constituents (in white) of the target sentences when they are spoken in a broad focus intonation, and with a contrastive focus on the first, second or third words. In the broad focus declarative sentences, pitch accents show a progressively smaller range in each word (means for word 1: 71 Hz, word 2: 33 Hz, word 3: 11 Hz). However, in sentences with contrastive focus, the largest pitch range is placed on the accents that express contrastive focus and their means are larger than those in broad focus. This shows that pitch range expands in CF. For example, when word 1 is in contrastive focus, its pitch range is 58 Hz larger than the pitch range of the same word in broad focus. The mean difference between contrastive and broad focus in word 2 is 58 Hz, and for word 3, it is 62 Hz.

In contrast with the expanded pitch accent range in CF, pitch accents on words adjacent to those that bear contrastive focus show a reduced pitch range,
in some cases reaching values close to 0 Hz, especially in post-focal contexts. For example, in sentences where contrastive focus is placed on word 1, the pitch range in post-focal contexts has a mean of 5 Hz in word 2 and 1 Hz in word 3, thus showing a strong tendency to full de-accentuation. These post-focal contexts, which for the purpose of this study are defined as those constituents occurring immediately after the contrastive focus elements, were compared with accents in similar sentence position in broad focus. These reduced pitch accents not only have a smaller pitch range than that of their adjacent contrastive focus, but also when compared with the pitch ranges of accents in broad focus, i.e. they were reduced by 28 Hz in word 2 and by 10 Hz in word 3.

In pre-focal contexts, again defined as those constituents immediately following the contrastive focus elements, we also observe a reduction in the pitch range. For example, when contrastive focus is on word 3, the pre-focal accents in word 1 and word 2 have mean pitch ranges of 35 Hz and 17 Hz respectively. Pitch accents on the same words have a pitch range of 71 Hz and 31 Hz respectively when realized in broad focus sentences.

Thus, while Romanian speakers placed a pitch accent in each word of the broad focus sentences, in sentences with contrastive focus, speakers produced a pitch accent with an expanded pitch range on the contrasted word, and reduced the pitch range of accents in pre- and post-focal position; this reduction is especially visible in post-focal positions.

![Figure 4](image_url)  
**Figure 4.** Mean pitch range values for each word in the sentence spoken with broad focus (BF) and contrastive focus (CF). Pitch range 1 refers to the pitch range in the first word of the sentence, pitch range 2 to the pitch range in the second word, and pitch range 3 to the third word. Contrastive focus can be placed on the first word (CF word 1), second (CF word 2) and third words (CF word 3).
The differences in the range of reduction between pitch accents in pre- and post-focal position are due to the patterns displayed below in Figures 5a, 5b, and 5c. In pre-focal position, speakers produce a flat F0 at the beginning of the sentence until it reaches the onset of the stressed syllable of the word with contrastive focus, where F0 increases abruptly (Figure 5a). This flat F0 contour was realized in 60% of the examined pre-focal tokens. Another possibility is for speakers to reduce only the second pitch accent of the utterance (5b), where a pitch accent is visible for the first pre-focal word whereas the pitch accent for the second word, immediately preceding the CF, is flattened. They may also produce an F0 that increases progressively from the beginning of the sentence until the peak of the pitch accent with CF (5c). Since this increment is progressive, it is difficult to distinguish pre-focal pitch accents. In spite of the above variability in the realization of the pre-focal intonation contours, there is a clear reduction of pitch accents in these contexts which may reach a completely flat F0.

In contrast with pre-focal contexts, speakers consistently produce the mirror image of (5a) in post-focal contexts. They drop the pitch abruptly after the CF pitch accent, producing a flat F0 until the end of the sentence, as it is evident in Graph 1 where the post-focal pitch range for CF 1 is considerably reduced.

![Figure 5a](image)

**Figure 5a.** Example of pre-focal reduction of 'pitch range 2'; 3rd word 'mere' in contrastive focus

![Figure 5b](image)

**Figure 5b.** Lack of pre-focal reduction for 'pitch range 1', with a pre-focal 'pitch range 2' reduction, with lack of pitch movement associated with the stressed syllable 'vin'; 3rd word 'mere' in contrastive focus
Cues to contrastive focus in Romanian

Pitch range 1
Pitch range 2
Pitch range 3

Figure 5c. Lack of pre-focal reduction for ‘pitch range 1’, with interpolation for ‘pitch range 2, with pitch movement visible for the stressed syllable ‘vin’; 3rd word ‘mere’ in contrastive focus.

An ANOVA with the factors of word (word 1, word 2, word 3) and sentence intonation (broad focus, CF in word 1, CF in word 2, CF in word 3) showed that the differences in pitch range between BF and CF were significant in each word position (pitch range in word 1: $F(3, 292)=135 \ p<0.001$, pitch range in word 2: $F(3, 292)=136 \ p<0.001$, pitch range in word 3: $F(3, 212)=119 \ p<0.001$). Multiple Comparisons with the Bonferroni adjustment confirm that on the one hand, pitch accents in CF have significantly higher pitch ranges than those in BF. On the other hand, they show that pitch ranges in pre- and post-focal accents in CF sentences are significantly reduced when compared with those in BF. Therefore, pitch range differentiates pitch accents in BF from those in CF by increasing the range of the pitch accents in CF and reducing the range of pitch accents in pre- and post-focal positions in CF.

3.2 F0 peak alignment

In Figure 6, the peak alignment of the BF pitch accents with respect to the syllable boundary is compared with that of CF pitch accents. It should be borne in mind that that 100% represents the end of the stressed syllable, so that values above 100 indicate that the peak is aligned with the post-tonic syllable while values below 100 show that the peak is within the stressed syllable. Peaks of CF pitch accents tend to be aligned earlier in the stressed syllable than those of BF pitch accents, which align closer to the syllable boundary or even in the post-tonic syllable. This variation in the peak alignment of BF accents seems to be related to stress and word boundaries.

The stimuli, designed to include both paroxytone and oxytone tokens, allows for an examination of this possible boundary effect. The post-tonic alignment is more frequent in paroxytone than in oxytone words, showing that post-tonic alignment is more likely to occur if it does not cross a word boundary. Nevertheless, the
earlier peak alignment of CF accents does not seem to be affected by this variation in the peak alignment of BF pitch accents. Thus, pitch alignment, like pitch range, may differentiate the two types of accents in Romanian.

Results from the paired t-test show that there is a statistically significant difference in the alignment of BF as compared with CF for word 1 and word 2 (2-tailed significance p<0.001 in all cases) confirming the earlier alignment of CF pitch accents. As discussed in Section 2.1 above, the BF alignment for the 3rd word was not considered due to pitch track errors and thus not included in our calculations.

![Figure 6](image_url)

**Figure 6.** F0 peak alignment for broad (BF) and contrastive focus (CF) with respect to the syllable rhyme. Since a value of 100% represents the rightmost boundary of the stressed syllable, values under 100% represent peaks aligned within the rhyme of the stressed syllable, while values over 100% refer to peaks that align with the post-tonic syllable.

### 3.3 Duration

In order to examine a possible effect of pitch range on duration, we compared the duration of the stressed vowels in BF, CF and de-accented contexts. For example, we compared the duration of the stressed vowel in the underlined vowel in Table 3 below) when it receives a BF pitch accent (sentence 1 in Table 3), when it receives a CF pitch accent (sentence 2), and when it is de-accented as in sentences 3 and 4. Similar comparisons were performed on the stressed vowels of words 2 and 3 for both paroxytone and oxytone words. Since comparisons were performed on the same word across different contexts, the variation in syllabic structure among target words did not bias our results. That is, duration differences were not computed between the stressed syllables of vin-de and ve-de, but between 'vin' in vin-de across different intonation contexts.
Table 3. Example of paroxytonic utterance ‘O mama vinde mere’ in broad focus (BF) and contrastive focus conditions on each of its constituents (CF word1, CF word2, CF word3); comparison of stressed vowel durations in BF vs. CF conditions

<table>
<thead>
<tr>
<th>Sentence type</th>
<th>Sentence examples</th>
<th>BF</th>
<th>CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Broad focus (BF)</td>
<td>O mama vinde mere.</td>
<td>121msec</td>
<td>125msec</td>
</tr>
<tr>
<td>2. CF in word1</td>
<td>O MAMA vinde mere.</td>
<td>85msec</td>
<td>89msec</td>
</tr>
<tr>
<td>3. CF in word2</td>
<td>O mama VINde mere.</td>
<td>158msec</td>
<td>159msec</td>
</tr>
<tr>
<td>4. CF in word3</td>
<td>O mama vinde MERE.</td>
<td>139msec</td>
<td>144msec</td>
</tr>
</tbody>
</table>

Table 4 below compares the mean durations of the stressed vowels that receive a BF pitch accent with those that bear a CF pitch accent, for each word position and stress pattern. Although vowels were longer in CF than in BF, these differences were so small, i.e. they ranged from 1 to 7 ms, that they became statistically non-significant. Thus, stressed vowels with CF pitch accents do not have larger durations than stressed vowels with BF pitch accents. Since CF pitch accents have larger pitch ranges than BF pitch accents, we can infer that larger pitch ranges did not increase vowel durations.

Table 4. Comparison of mean vowel durations (msec) in broad focus (BF) and contrastive focus (CF) condition

<table>
<thead>
<tr>
<th>Stress</th>
<th>Word in sentence</th>
<th>BF</th>
<th>CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paroxytone</td>
<td>Word 1</td>
<td>121msec</td>
<td>125msec</td>
</tr>
<tr>
<td></td>
<td>Word 2</td>
<td>85msec</td>
<td>89msec</td>
</tr>
<tr>
<td></td>
<td>Word 3</td>
<td>158msec</td>
<td>159msec</td>
</tr>
<tr>
<td>Oxytone</td>
<td>Word 1</td>
<td>133msec</td>
<td>136msec</td>
</tr>
<tr>
<td></td>
<td>Word 2</td>
<td>141msec</td>
<td>148msec</td>
</tr>
<tr>
<td></td>
<td>Word 3</td>
<td>139msec</td>
<td>144msec</td>
</tr>
</tbody>
</table>

In contrast, the durations of stressed syllables in pre- and post- focal positions in CF were much shorter than those in BF. As shown in Table 5, the pre- and post-focal vowels significantly compress their duration with respect to instances when the same vowels are in the BF context in both oxytone and paroxytone utterances. For example, when word 1, like mama in O mama vinde mere, has a BF pitch accent, the stressed vowel has a mean duration of 121 ms. When this vowel is in pre-focal position, its mean duration decreases to 96 ms. When the second word in a sentence, like vinde in O mama vinde mere, is in BF, the mean duration of the stressed vowel is 85 ms. However, in pre-focal position, its mean duration is 68 ms.
and in post-focal position is 35 ms. Multiple comparisons with the Bonferroni adjustment yield significant results for all the comparisons confirming that vowels in pre- and post-focal contexts are shorter than those in BF.

Table 5. Pre- and post-focal vowel reduction. Mean vowel durations (msec) in broad focus (BF) and in pre-/post-focal contexts; significance decrease independent of sentence stress (paroxytone and oxytone)

<table>
<thead>
<tr>
<th>Vowel compared &amp; stress type</th>
<th>Mean duration (msec)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BF</td>
</tr>
<tr>
<td>PAROXYTONE</td>
<td></td>
</tr>
<tr>
<td>Word 1</td>
<td>121msec</td>
</tr>
<tr>
<td>Word 2</td>
<td>85msec</td>
</tr>
<tr>
<td>Word 3</td>
<td>158msec</td>
</tr>
<tr>
<td>OXYTONE</td>
<td></td>
</tr>
<tr>
<td>Word 1</td>
<td>133msec</td>
</tr>
<tr>
<td>Word 2</td>
<td>141msec</td>
</tr>
<tr>
<td>Word 3</td>
<td>146msec</td>
</tr>
</tbody>
</table>

Finally, we compared the duration of stressed and unstressed vowels within each word. Since all words have two syllables and are either oxytone or paroxytone, the duration of the stressed vowel in a word was compared with that of its adjacent unstressed syllable. For example, in paroxytone words such as those in ‘o mama vinde mere’, the duration of the stressed vowels was compared with the duration of their adjacent post-tonic vowels. These comparisons were calculated for words with CF, BF and in de-accented contexts.

As shown in Figure 7, while stressed vowels are longer than their unstressed counterparts, this difference is larger in vowels with BF and CF than in vowels in pre- and post-focal contexts. For example, the mean difference between stressed and unstressed vowels for CF is 29 ms and for BF is 17 ms. Yet this difference scores only 8ms in pre- and post-focal contexts because the stressed vowel in this context is much shorter than in BF and CF contexts. Results from the ANOVA test confirm that in general, stressed vowels are significantly longer than unstressed vowels (F(3, 296)=132, p=0.05). Post-hoc tests confirm that this difference is significant only in BF and CF contexts. Thus, stressed vowels in Romanian are longer than their unstressed counterparts. However, this difference is larger in words with CF and BF pitch accents. In de-accented words, i.e. word in pre- and post-focal contexts, there is less difference due to a significant shortening of the stressed vowel.
3.4 Summary

The preceding sections illustrated that there are several strategies in Romanian to cue contrastive focus: (1) through the F0 pitch range increase under CF and the pre- and post-focal pitch range reduction, (2) the tonal alignment within the rhyme bounds for CF constituents, closer to the vowel mid-point as compared with the constituents in BF. It was also shown that (3) vowel duration cues are an essential element of CF expression, signaling it through a significant compression of duration in pre- and post-focal contexts which mirrors the behavior of the pre- and post-focal accents. The shape of a CF contour is thus dependent on phonetic factors at both the level of the melodic curve as well as the segmental level, which conspire to create meaning by foregrounding the contrasted elements with an increased F0 pitch range, and by surrounding it with the material in the background through de-accentuation and vowel duration compression.

4. Discussion and conclusions

This experiment describes the phonetic make up of broad focus and contrastive focus declarative sentences in Romanian. In broad focus sentences, speakers of Romanian consistently place a pitch accent on each syllable with lexical stress. The pitch range of these accents is progressively lower towards the end of the sentence and the peaks of the pre-nuclear accents align with the post-tonic syllable. Syllables with lexical stress are consistently longer than their unstressed counterparts.
In contrastive focus declarative sentences, speakers use several strategies based on pitch range, pitch alignment and duration to make the accented word more prominent than the rest. In CF the focalized word bears a pitch accent, which has an expanded pitch range whose peak aligns within the stressed vowel. However, this expanded pitch range does not have a lengthening effect on segmental duration since stressed syllables under CF pitch accents have similar durations to those under BF pitch accents, which have a smaller pitch range. In post-focal contexts, F0 has the lowest pitch values in the sentence and shows very little movement. In pre-focal positions, there is also a strong tendency to reduce F0 movement by either displaying a flat F0 or by reducing pitch accents. Although this reduced F0 movement is realized with a degree of variability, as seen in figures 5(a–c), all tokens show an important reduction or de-accentuation of the pre-focal pitch accent. Moreover, segmental durations of words immediately adjacent to the word in CF are compressed, particularly in the stressed syllables.

Thus, Romanian speakers highlight the accented word in CF by contrasting the large pitch excursion of the CF pitch accent against the flat F0 trajectories of pre- and post-focal contexts, especially in those words that are adjacent to the CF. They also make the word under CF sound longer, not by lengthening the segmental durations of the word under focus, but by compressing the durations of the words adjacent to the CF.

These results contribute to the discussion on the distribution and nature of pitch accents in Romance languages (Hualde 2002). Concerning the distribution of pitch accents in the sentence, Hualde shows that Romance languages either exhibit a high density of accents by placing one pitch accent on each syllable with lexical stress, as in Spanish and Italian. Or, like European Portuguese, only the first and last word of a declarative statement bear a pitch accent while syllables in between these two words show no prominence marking. Winkler-Gobbel’s (2000) analysis of Romanian showed that in declarative statements, some words, especially verbs, tend to be de-accented suggesting that Romanian, an SVO language, could pattern more like European Portuguese than like Spanish or Italian. However, our data has shown that such de-accentuation does not exist, but that each lexical stress of the utterance was linked to a pitch accent. Thus, our results indicate that Romanian patterns more like Spanish and Italian than like European Portuguese.

These contradictory results may, in part, be explained by the effect of lexical items on prominence. There seems to be reasonable evidence that the properties of the segmental string, such as lexical items, have an independent effect on prominence marking (Calhoun 2006). For example, certain types of words are more likely to be prominent than others. Function words like articles, determiners, prepositions and pronouns are often unstressed at sentence level (for Spanish, see Hualde 2006/2007). Moreover, some classes of content words, i.e. nouns, are more likely
to be accented than others, i.e. verbs. Face’s (2003) study of spontaneous speech in Spanish, a language that, like Romanian, exhibits high pitch density, seems to corroborate these patterns. He finds that verbs, especially high frequency verbs, tend to be de-accented. Thus, it is possible that in Winkler-Gobbel’s (2000) database, this effect of word class de-accentuation is more apparent than in ours since our data is based on a question-answer elicitation method and Face’s results are based on spontaneous speech.

With regard to the nature of pitch accents, Hualde (2002) explains that some Ibero- and Italo-Romance languages use pitch accent shapes in a pragmatically contrastive manner. Thus, speakers can choose from among different pitch contours on stressed syllables in order to express different pragmatic meanings. More specifically, CF pitch accents have a different shape from pitch accents that do not have this contrastive meaning. For example, in Spanish, the peaks of pre-nuclear accents align with the post-tonic syllable. However, when pitch accents have a contrastive meaning, the peak aligns within the stressed syllable. These alignment differences determine two distinct pitch shapes, which in AM notation are transcribed as L*H and LH*. In Italian and Portuguese, contrastive pitch accents on the last word of the sentence align its peak within the stressed syllable while neutral or non-contrastive accents in the same position show a falling pattern (D’Imperio 2003; Frota 2000).

Similarly, our results indicate that CF pitch accents in Romanian have a different shape from neutral pitch accents in pre-nuclear position: the peaks of CF accents align with the stressed vowel and peaks of neutral accents align closer with the syllable boundary or with the post-tonic syllable. These alignment differences within neutral pitch accents seem to be related to word boundaries since words with paroxytone stress exhibit the post-tonic alignment while oxytone words align closer at the end of the syllable showing that these alignment differences are related more to phonetic factors rather than to meaning itself. In contrast, the peak of CF pitch accents consistently aligns with the center of the stressed vowel. Therefore, the feature that differentiates pitch shapes in Romanian aligns the peak with the center of stressed vowel, as in CF pitch accents, or closer to the syllable boundary, as in neutral pitch accents.

However, some researchers question the definition of pitch accents types solely in terms of tonal alignment and propose that meaning differences attributed to different pitch accent types are in fact signaled by multiple phonetic cues, potentially at different levels of the prosodic structure (Calhoun 2006:67 and references therein). As Face points out for Spanish, pitch alignment is one amongst several strategies to mark CF. Similarly, our results indicate that Romanian speakers not only use pitch alignment, but also pitch range and segmental duration to express CF. They highlight the accented word in CF by contrasting the large pitch excursion of
the CF pitch accent against the flat F0 trajectories of pre- and post-focal contexts, and make the word under CF sound longer, not by lengthening the segmental durations of the word under focus, but by compressing the durations of the words adjacent to the CF. Thus, in addition to pitch shape, a word in CF is marked by contrasts in pitch range and segmental duration. It would be interesting to examine the perceptual salience of pitch alignment, pitch range and segmental duration in relation to CF in order to shed light upon what the most important phonetic factors are for conveying phonological meaning of CF in Romanian.

In conclusion, this study contributes towards making a description of intonation in the Romance languages by examining the phonetic characteristics of broad focus and contrastive focus declarative sentences in Romanian. Broad focus sentences in Romanian are very similar to those in Spanish and Italian in that they have a high density of pitch accents: each lexical stress in the sentence bears a pitch accent. CF pitch accents in Romanian, such as in Italian, European Portuguese and Spanish, exhibit a different shape from neutral pitch accents. However, since Romanian speakers also use strategies based on pitch range and segmental duration to highlight the word in CF, it is necessary to test the perceptual salience of these cues in order to understand fully how CF works in Romanian. Perceptual testing will both highlight the interaction of various phonetic cues, as well as support the phonological assessment of these cues.

References


The phonetics of sentence-initial topic and focus in adult and child Dutch*

Aoju Chen  
Max Planck Institute for Psycholinguistics

This study investigates whether adults and children use phonetic means to distinguish sentence-initial topic and focus marked with the same accent type (H*L) in Dutch declaratives. It was found that in adults’ speech, the falling accent starts to fall earlier and has a larger F0 excursion and lower F0 minimum in focus than in topic. Further, the low F0 is maintained longer in focus. Moreover, the accented syllable and word are longer in focus than in topic. In contrast, children do not yet use any of the phonetic cues to distinguish topic and focus at the age of 4 or 5. At the age of 7 or 8, they become adult-like only in the use of F0 lowering. Considering that children are fully adult-like in phonological marking of topic and focus at the age of 7 or 8, our findings suggest that phonetic marking is acquired later than phonological marking.

1. Introduction

The intonational realization of information structural categories in terms of gradient parameters such as F0 excursion and word duration (i.e. phonetic realization) has been widely studied across languages. The majority of previous studies are, however, concerned with the phonetic realization of different types of focus in adults. Furthermore, they do not explicitly deal with phonetic differences that can be inherent in phonological differences, e.g. differences in accent placement and accent type (cf. Manolescu, Olson & Ortega-Llebaria 2009, this volume). The present study contributes to this line of research by examining the use of phonetic means to distinguish topic and focus that are marked with the same phonological means (i.e. the same type of pitch accent) in children as well as adults in Dutch declarative sentences. Topic is generally defined as what the sentence is about; focus is defined as what is predicated about the topic (Lambrecht 1994). In this study, topic is operationalized as

*I thank Rebecca Defina and Marieke Hoetjes for their assistance in acoustic annotation, Bettina Braun for her help with R, and the anonymous reviewers for their constructive comments. I am also grateful to Rik van den Brule for writing the much needed Praat scripts.
the referent about which a WH-question is asked; focus is the information required by the WH-word about the referent, as illustrated in (1). Topic typically conveys information known to the receiver (person B in example (1)), whereas focus typically conveys information new to the receiver (person A in example (1)) (e.g., Lambrecht 1994; Gundel 1999). Topic is thus assumed to have a lower information value than focus. Both topic and focus can become contrastive if there are overt alternatives in the discourse that may fill the same position, as illustrated in (2). Focus can also have different scopes, i.e. a single lexical word (narrow focus) vs. larger than one lexical word (broad focus) (Ladd 1980). Contrastive focus usually has a narrow scope.

(1)  
Person A: A boy was in the room a while ago. The boy drew something on the door. What did the boy draw?
Person B: [The boy]topic drew [a dragon]focus.

(2)  
Person A: Two kids were in the room a while ago. The kids drew something on the door. The girl drew a butterfly. What did the boy draw?
Person B: [The boy]contrastive topic drew a [dragon]contrastive focus.

In the remaining part of the introduction, we consider the phonological marking of topic and focus in West Germanic languages in detail in Section 1.1 and specify the questions addressed in this study in Section 1.2. Following a review of prior work on phonetic realization of information structural categories in Section 1.3, we describe the phonetic parameters examined in this study and the predictions in adults and children in Section 1.4.

1.1 Similarity in phonological marking of topic and focus

It has been suggested in the literature that topic and focus are distinguished phonologically in West Germanic languages. For English, Jackendoff (1972:237) proposes that focus and topic are marked with different types of pitch accent. Focus receives a fall accent (A accent) and topic receives a fall-rise accent (B accent) regardless of whether they are contrastive or not. Many authors have made similar claims on the accent type associated with focus in declarative sentences (Ladd 1980; Gussenhoven 1983; Pierrehumbert & Hirschberg 1990; see also Manolescu et al. 2009, this volume, for discussion on accent type and focus marking in Romance languages). But they assign different labels to Jackendoff’s focus accent due to the use of different notations, namely, H* in ToBI-like notations (Beckman & Ayers-Elam 1997) but H*L in the ToDI-like notations (Gussenhoven 2005).1 Opinions about

---

1. Jackendoff’s B accent is also transcribed differently by different authors and sometimes even in different papers by the same author. See Hedberg and Sosa (2007:109) for an overview of various interpretations of B accent. The differences in labels seem to stem from whether
the intonational realization of topic in English are, however, more diverse. Different from Jackendoff (1972), Gundel (1978) and Steedman (2000) argue that only contrastive topics are spoken with B accent (transcribed as L+H* in ToBI). Vallduví and Engdahl (1996) claim that the assignment of the L+H* accent is optional on subject topics but obligatory on contrastive subject topics, topicalized non-subject topics and non-subject topics in situ. For German, Féry (1992) states that in declarative sentences focus is realized with H*L and topic is realized with either L*H or H* (in a ToDI-like notation). Vallduví and Engdahl (1996) claim that Dutch is similar to German in this respect.

Different from the traditional view, Lambrecht (1994) argues that topic, associated with givenness, is usually not accented and becomes accented only if its topical status is not yet ratified at the time of utterance. Further, topic accent is not necessarily different from focus accent in terms of form. These claims are borne out by Hedberg and Sosa’s (2007) corpus-based study on the intonational marking of topic and focus in spontaneous dialogues in American English. Hedberg and Sosa found that ratified topics are indeed mostly deaccented and that both focus and unratified topic are frequently realized with H* and L+H* [in ToBI] independent of contrast.

The use of the same set of accent types to mark topic and focus in American English also appears to hold for German. Braun (2006) reported that in read speech, sentence-initial topic is realized with either L+H* and L*+H (in ToBI for German) independent of contrast. The two accent types are followed by either a high plateau or a fall. In ToDI, the L+H* accent in Braun’s study would be transcribed as H* when followed by a high plateau, and H*L when followed by a fall; the L*+H accent would be transcribed as L*H when followed by a high plateau, and L*H L when followed by a fall. The reinterpretation of Braun’s results suggests that topic is associated with more than two accent types, including the so-called focus accent H*L. This in turn suggests that both topic and focus are sometimes realized with the same accent type or the same shape of pitch movement in sentence-initial position in German.

The use of the same accent types in topic and focus observed in American English and German has also been reported for Dutch. Chen (2007) examined the marking of (non-contrastive) topic and focus in Dutch declarative sentences produced as answers to WH-questions in a semi-spontaneous production task by adults and children. She found that although adults typically realize focus with H*L the transcription includes the tone of the syllable preceding the accented syllable (vs. H* L H% vs. L+H* L H%) and whether it includes the phrasal accents and boundary tones (e.g. L+H* vs. L+H* L H%).
and topic with no accent in sentence final position, they frequently accent both topic and focus with H*L in sentence initial position. In a subsequent perception experiment, she found that listeners ranked the melody of answer sentences to WH-questions about the object to be melodically more pleasant when both the subject noun and the object noun were accented with H*L than when only the object noun was accented with H*L. Chen thus suggested that Dutch speakers accent sentence-initial topic for rhythmic motivation, like English speakers (Horne 1991).

Taken together, the similarity in the phonological marking of topic and focus found in recent empirical work casts doubts on the traditional view that topic and focus are phonologically distinguishable and calls for a refined analysis on the intonational realization of topic and focus. Prior work on the phonetic realization of information structural categories has shown that categories differing in information value tend to differ along phonetic parameters like F0 excursion and word duration (more discussion on this in Section 1.3). As topic has a lower information value than focus, a question that arises is whether topic and focus are distinguishable along gradient phonetic parameters even though they are marked with the same accent type for rhythmic motivation. In the present study, we addressed this question by investigating the phonetic realization of sentence-initial topic and focus accented with H*L in Dutch declarative sentences.

Regarding children's marking of topic and focus, Chen (2009) has found that children are not completely adult-like in the phonological realization of topic and focus at the age of 4 or 5 but become fully adult-like at the age of 7 or 8. The differences between 4- to 5-year-olds and the others (7- to 8-year-olds and adults) are mainly in sentence-final position. In sentence-initial position, like adults and 7- to 8-year-olds, 4- to 5-year-olds realize both topic and focus frequently with H*L. The question that arises then, is whether 4- to 5-year-olds are similar to 7- and 8-year-olds and adults in the phonetic realization of sentence-initial topic and focus. This is the second question addressed in this study.

1.2 Previous work on phonetic realization of topic and focus

Previous work on the phonetic realization of information structural categories is mostly concerned with different types of focus in read speech. For example, Eady and Cooper (1986) studied the use of F0 and duration in narrow focus and broad focus in question-answer dialogues in American English and found that sentence-final nouns are spoken with both a longer duration and a higher F0 peak in narrow focus than in broad focus. Sityaev and House (2003) examined the use of F0 and duration in contrastive focus as well as narrow focus and broad focus in question-answer dialogues in British English. They found that narrow or
contrastive focus differed from broad focus in duration but not in F0-related cues. Recently, Ishihara and Féry (2006) investigated the use of F0 and duration in the marking of first occurrence focus (i.e. a constituent that is in the scope of a focus sensitive operator, e.g. ‘only’, and that appears for the first time) and second occurrence focus (i.e. a constituent that is in the scope of a focus sensitive operator, e.g. ‘only’, and a repetition of an earlier focused constituent) in declarative sentences in German. They found that second occurrence focus in prenuclear position is accented, like a first occurrence focus but it is realized with a significantly lower mean F0 than a first occurrence focus. In a more recent study, Baumann, Becker, Grice and Mücke (2007) examined both the intonational (F0 and duration) and articulatory (i.e. vowel quality) marking of contrastive focus, narrow focus and broad focus in question-answer dialogue in German. They found that in nuclear pitch accents, the peak is higher and aligned later in the accented word in contrastive focus than in narrow focus and broad focus. Further, the F0 excursion is higher in contrastive focus than in broad focus. Moreover, nuclear syllables and feet are the longest in contrastive focus and the shortest in broad focus. As regards vowel quality, vowels in nuclear syllables are more peripheral in narrow focus and contrastive focus than in broad focus. Hanssen, Peters and Gussenhoven (2008) investigated the intonational realization of these three types of focus in question-answer dialogue in Dutch. The accent-bearing word was spoken with a falling accent across focus conditions. However, narrow focus and contrastive focus were phonetically distinguishable from broad focus. In comparison with broad focus, the accented-word in narrow focus and contrastive focus was spoken with a longer duration of the onset and coda of the accented syllable, an earlier alignment of pitch peak, a steeper fall, a lower and earlier elbow after the pitch peak, and a lower onset pitch of the following stressed vowel.

We are aware of only one study on the phonetic marking of different types of topic. Braun (2006) examined the phonetic realization of non-contrastive topic and contrastive topic at sentence-initial position that were realized with the same accent types (L+H* and L*+H). She found that contrastive topics are realized with a higher peak, a later peak and a larger F0 excursion than non-contrastive topics. There has been no published work comparing the phonetic marking between topic and focus.

Taken together earlier findings have shown that the type of focus and topic with a higher information value tends to be realized with more acoustic prominence, assuming that contrast and narrow scope increase information value though not necessarily in tandem. More acoustic prominence can be implemented as a higher mean F0, a higher F0 peak, a wider F0 excursion and a longer duration. Languages can however differ in the choice of phonetic parameters to achieve acoustic prominence. For example, F0-related cues vary with focus
types in American English but not in British English (Hanssen et al. 2008; see also Manolescu et al. 2009, this volume, for discussion concerning Romance languages). Furthermore, variations in peak alignment can go different directions in different languages. For example, peak alignment in contrastive focus is later in German but earlier in Dutch than in broad focus.

1.3 Phonetic parameters and predictions in the present study

In the present study, we examined the phonetic properties of sentence-initial topic and focus marked with H*L along the following parameters: F0 maximum, F0 minimum, F0 excursion, alignment of F0 maximum (peak alignment), alignment of F0 minimum, and duration of accented vowel, syllable and word. On the assumption that focus has a higher information value than topic, we predicted that sentence-initial focus would be realized with a larger F0 excursion (resulting from a higher F0 maximum and possibly a lower F0 minimum), a later peak alignment - as a substitute for a higher peak (Gussenhoven 2002), an earlier alignment of F0 minimum (Hanssen et al. 2008), and a longer duration in the accented vowel, syllable and word.

As regards children's phonetic marking of topic and focus, two conflicting hypotheses can be put forward. According to Bolinger (1983), variations in young children's pitch can be a reflection of the universal physiological mechanism whereby an increase in excitation leads to an increase in pitch. Assuming that focus is produced with more excitation than topic, we hypothesized that children would at least use a larger F0 excursion in focus than in topic at the age of 4 or 5 and could use the other phonetic cues like adults at the age of 7 or 8. On the other hand, focus may not trigger more excitation in the speaker than topic; a child may get excited about part of a sentence for non-pragmatic reasons. As gradient variations in phonetic parameters seem to be less salient than changes in accent placement and accent type and consequently may be harder to learn, we hypothesized that children would not be able to use any of the phonetic cues to mark topic and focus at the age of 4 or 5 but would be able to use some of the cues like adults at the age of 7 or 8.

2. Method

Our data consisted of sentences with the topic-focus structure spoken by speakers from three age groups: 4- to 5-year-olds, 7- to 8-year-olds, and adults (mean age: 22 years). The sentences were selected from semi-spontaneous production reported in Chen (2007, 2009). We will briefly describe Chen's production task. We refer the reader to Chen (2007, 2009) for a more detailed description of the task. Following this, we will describe the criteria used to select the sentences for the present purpose and how the phonetic measurements were made.
2.1 Data elicitation

A picture-matching game was used in Chen (2007, 2009) to elicit topic-focus structures from children aged between 4 and 8 as well as adults. Prior to the game, the experimenter showed the participant two boxes full of pictures. The participant was told that a picture from one box went together with a picture from the other box and that the experimenter needed his/her help to sort the pictures out. The procedure of the game is as follows. First, the experimenter took a picture (e.g. a picture of a cleaning-lady) from one box. She then drew the participant’s attention to the picture and established what the picture was by saying ‘Kijk! Een poetsvrouw!’ “Look! A cleaning-lady!” with either H*L or L*H on the verb and H*L on the noun. In the picture, the cleaning-lady seemed to be picking up something. The experimenter then asked a question about the picture (e.g. ‘Wat pakt de poetsvrouw?’ “What does the cleaning-lady pick up?”), again in a prescribed intonation contour. The WH-word was spoken with H* the noun was spoken with either no accent or H*L with a reduced F0 excursion. Second, the participant turned to a robot for help by clicking on a picture of the robot displayed on his/her computer screen. The participant received the answer (in SVO word order) in abnormal intonation from the robot via a headphone set so that the experimenter could not hear it. 2 Third, the participant then used the same words as the robot to answer the experimenter’s question but in his/her own intonation (e.g. ‘De poetsvrouw pakt de vaas.’ “The cleaning-lady picks up the vase.”). Finally, the experimenter looked for the matching picture from the other box and handed both pictures over to the participant.

The experiment was conducted by means of a portable set-up equipped with the NESU experiment software. It began with four practice trials. Children were tested individually in a quiet room at their school during school time. Adults were tested individually in an experiment room at the Max Planck Institute for Psycholinguistics. They were told prior to the experiment that the game was originally developed for children and therefore was very simple. Each session was recorded with an external high-quality microphone connected to a portable DAT recorder at 48 kHz sampling rate with 16-bit resolution. The microphone was placed 10–15 cm away from the mouth of the participant.

Thirty-six answer sentences were elicited from each participant. One half of the sentences were answers to WHO-questions; the other half of the sentences were answers to WHAT-questions. Thus in one half of the sentences, the sentence-initial noun (subject) was the focus and the sentence-final noun (object) was the

2. The robot’s answer sentence was generated by splicing together the words (with a 200 ms pause in between) recorded in a wordlist reading, as in Müller et al. (2005). The original intonation was then erased and the pitch level was set at 200 Hz to obtain a flat intonation.
topic; in the other half of the answer sentences, the sentence-initial noun was the topic and the sentence-final noun was the focus. Each noun served both as focus and topic in different sentences, as shown in (3).

(3)  
Experimenter: Wie eet de biet?  
"Who eats the beet?"

Participant: [De poetsvrouw]$_{focus}$ eet [de biet]$_{topic}$.  
"The cleaning-lady eats the beet."

Experimenter: Wat pakt de poetsvrouw?  
"What does the cleaning-lady pick (up)?"

Participant: [De poetsvrouw]$_{topic}$ pakt [de vaas]$_{focus}$.  
"The cleaning-lady picks (up) the vase."

2.2 Data selection

For the present purpose, we needed pairs of answer sentences with identical subject nouns spoken with H*L. In each pair, the subject noun was the focus in one sentence and topic in the other sentence. To compose such a data set, we selected all answer sentences with the subject noun accented with H*L present in Chen’s (2007, 2009) data obtained from nine adults, twelve 4- to 5-year-olds and twelve 7- to 8-year-olds. We further trimmed this selection of sentences on the basis of the phrasal structure, the maintenance of the SVO word order, and the quality of the pronunciation. Specifically, as phrasing can be a strategy to distinguish topic and focus in addition to accent placement and accent type, a sentence was excluded if an intonational phrase boundary was perceived after the subject noun. Further, a sentence was excluded if it had a non-SVO word order. Moreover, a sentence was excluded if the subject noun was produced with stammering, a pause between the two morphemes (in the case of compound nouns), a mispronounced phoneme, a misplaced word stress, or laughter. As the last step in our data selection procedure, sentence pairs containing the two renditions of the same subject noun from the same speaker were selected from the trimmed selection. Although speakers across age groups frequently used H*L in both sentence-initial topic and focus, they did not necessarily use H*L in the sentences containing the same subject noun. The final selection gave us a set of 113 sentence pairs (48 from eight adults, 29 from nine 4- to 5-year-olds, and 36 from ten 7- to 8-year-olds).

2.3 Acoustic analysis

The subject nouns in the 113 sentence pairs were acoustically annotated by examining the waveform, wide-band spectrum and pitch track in Praat in combination with auditory impressions (Boersma 2001). Two F0-related landmarks were labeled in each subject noun, as illustrated in Figure 1:
- **H**: the point at which F0 maximum was reached (the begin of the fall)
- **L**: the point at which F0 minimum of the fall was reached

When labelling the F0-related landmarks, we discarded micro-prosodic effects by searching for the highest F0 after the first three to five periods of the accented vowel and the lowest F0 before the voice started to fade out towards the end of the word. Halving errors were observed occasionally in the region where the F0 minimum was expected because of the transition from one phoneme to another and creaky voice. They were manually corrected after the F0 values at the H and L landmarks were automatically extracted.

Further, a number of segmental landmarks were labeled in each subject noun:

- **C0**: the beginning of the accented syllable (also the beginning of the word)
- **V0**: the beginning of the accented vowel
- **C1**: the end of the accented vowel (or the end of the accented syllable if the coda consonant was absent)
- **C2**: the end of the accented syllable (also the end of the word in the case of monosyllabic words)

Three additional landmarks were labeled in subject nouns with more than one syllable:

- **V1**: the beginning of the post-accent vowel
- **C3**: the end of the post-accent vowel (not labeled if the word-final consonant was absent, as in Figure 1)
- **C4**: the end of word

---

**Figure 1.** Landmarks in a disyllabic word ‘poetsvrouw’ “cleaning-lady” produced in the topic condition: F0 maximum (H), F0 minimum (L), beginning of accented syllable and word (C0), beginning of accented vowel (V0), end of accented vowel (C1), end of accented syllable (C2), beginning of post-accent vowel (V1), and end of post-accent syllable and word (C4)
F0 values (in semitones with 1 Hz as the reference point) at the F0-related landmarks and time values (in seconds) at all segmental landmarks were automatically extracted by means of Praat scripts. Eight measurements were then obtained from each subject noun:

- F0 maximum: F0 at the H landmark (F0_H)
- F0 minimum: F0 at the L landmark (F0_L)
- F0 excursion: F0_H - F0_L
- Alignment of F0 maximum: Time_{C1} - Time_H
- Alignment of F0 minimum:
  - Time_L - Time_{C1/C2} in monosyllabic words
  - Time_L - Time_{C4} in words with more than one syllable
- Word duration:
  - In monosyllabic words: Time_{C1/C2} - Time_{C0}
  - In words with more than one syllable: Time_{C4} - Time_{C0}
- Syllable duration: Time_{C1/C2} - Time_{C0}
- Vowel duration: Time_{C1} - Time_{V0}

Alignment of F0 maximum was calculated relative to the end of the accented vowel, because the peak of prenuclear H*L (referred to as a prenuclear rise in ToBI terms) has been found to be aligned differently depending on the vowel length in Dutch (Ladd, Mennen & Schepman 2000). Alignment of F0 minimum was calculated as the distance from the time point of L to the end of the word, capturing the duration of the low plateau following F0 maximum.

3. Results and discussion

To assess the effect of topic and focus on the F0-related and duration-related measurements, mixed-effect models were built in R (an open source implementation of the S language for statistical analysis) for each measurement with information structure (i.e. topic, focus) as the fixed-effect factor and participant and target word as the random-effect factors (Baayen 2008). P-values were estimated as posterior probabilities based on Markov Chain Monte Carlo sampling with 10000 samples at the significance level of 0.05.

3.1 Adults

3.1.1 F0 measurements. Figure 2 shows the mean F0 maximum, F0 minimum and F0 excursion for the focal subject nouns and their topical counterparts in
the adults’ data. As can be seen, F0 excursion was larger in focus (5.59 st) than in topic (3.43 st). Further, F0 minimum was lower in focus (84.54 st) than in topic (86.46 st), whereas F0 maxima was similar in focus and topic. Mixed-effect modeling showed that the effect of the fixed-effect factor information structure reached significance regarding F0 excursion (p < 0.0005) and F0 minimum (p < 0.0005). This result indicated that the larger F0 excursion in focus was caused primarily by the lower F0 minimum in focus. In other words, the fall in H*L fell to a lower F0 point in focus than in topic.

![Figure 2. Mean F0 excursion, F0 minimum, and F0 maximum in topic and focus in adults](image)

3.1.2 *Duration measurements.* Figure 3 shows the mean durations of the accented word, syllable and vowel for focus and topic in adults’ data. As word duration was the same as the duration of the accented syllable in monosyllabic words, only the mean syllable duration of the words with more than one syllable was shown in Figure 3. As can be seen, each duration measurement was larger in focus than in topic, although the magnitude of the difference between focus and topic differed among the duration measurements (0.054 seconds in word duration averaged over all words and 0.057 seconds in word duration averaged over words with more than one syllable, 0.022 seconds in syllable duration, and 0.004 seconds in vowel duration). The models for word duration and vowel duration were built with measurements obtained from all subject nouns; the model for syllable duration was built with measurements obtained from the subject nouns with more than one syllable only. The effect of the fixed-effect factor information structure was significant regarding word duration (p < 0.005) and syllable duration (p < 0.01). The observable difference in the magnitude of lengthening between the accented syllable and the word in focus indicated that the post-accent syllable was substantially lengthened in addition to the accented syllable in this condition. This pointed to a general lengthening of the segments in the focal word.
3.1.3 Alignment of F0 maximum and F0 minimum. Figure 4 displays the mean temporal distances in seconds from F0 maximum to the end of the accented vowel and from F0 minimum to the end of the word in topic and focus in the adults’ data. As can be seen, F0 maximum was reached 0.008 seconds earlier in focus than in topic, meaning that H*L started to fall earlier in focus than in topic. Further, F0 minimum was reached 0.032 seconds earlier in focus than in topic, suggesting that there was a longer low plateau in focus than in topic. Mixed-effect modeling showed that the difference was significant in both alignment of F0 maximum (p < 0.05) and alignment of F0 minimum (p < 0.005).

The use of earlier alignment of F0 minimum (or a longer low plateau) in focus may be related to informational completeness. The lengthening of the low plateau is perceptually similar to the lengthening at phrase-final position. It follows that the longer the low plateau is, the more final the subject noun sounds. In the focus condition, the subject noun has provided the information required by the WH-word. In terms of informationality, it is appropriate to end the answer after the subject noun. The speakers in our study provided full-sentence answers. It was highly likely that they demarcated the informational completeness at the end of the subject noun by means of a longer low plateau.

The earlier alignment of F0 maximum in focus was contra our prediction. It was however possible that speakers started to decrease their F0 earlier in order to reach a lower F0 point in focus as it takes more time to complete a fall with a larger F0 excursion (Xu 2002). This was supported by the fact that the temporal distance between F0 maximum and F0 minimum was 0.016 seconds longer in focus than in topic in spite of the earlier alignment of F0 minimum in focus. Together with lengthening of segments, the earlier alignment of F0 maximum in focus could account for the difference in the duration of the fall.

To sum up, although adult speakers of Dutch frequently accent focus and topic with H*L in sentence-initial position, they clearly distinguish the two information
The phonetics of sentence-initial topic and focus

1. Structural categories by means of a larger F0 excursion, a lower F0 minimum, an earlier alignment of F0 maximum, a longer low plateau, and general lengthening of segments in the focal words.

2. Children

With respect to the data from 4- to 5-year-olds, the effect of information structure did not reach significance in any of the mixed-effect models. This indicated that although children from this age range, like adults, used H*L to realize sentence initial topic and focus, they were not yet able to use any of the phonetic cues to distinguish topic from focus.

With respect to the data from 7- to 8-year-olds, the effect of information structure reached significance only in F0 excursion and F0 minimum. As shown in Figure 5, F0 range was larger in focus (2.71 st) than in topic (2.03 st) and F0 minimum was lower in Focus (95.42 st) than in topic (96.07 st). This result showed

---

**Figure 4.** Mean alignment of F0 maximum (relative to end of accented vowel) and F0 minimum (relative to end of word) in topic and focus in adults

**Figure 5.** Mean F0 excursion, F0 minimum and F0 maximum in topic and focus in 7- to 8-year-olds
that 7- to 8-year-olds were adult-like in the use of F0 but did not yet exploit variations in duration, alignment of F0 maximum, and alignment of F0 minimum to distinguish between topic and focus.

4. Conclusions

We have examined the phonetic realization of sentence-initial topic and focus marked with H*L in adults and children in Dutch declarative sentences. Our results show that in sentence-initial position topic and focus are phonetically distinguished in adult Dutch even though topic is accented with H*L, like focus. In comparison with topic, focus is realized with a larger F0 excursion, a lower F0 minimum and lengthening in both the accented syllable and the post-accent syllable(s), as expected. Interestingly, focus is also distinguished from topic in alignment of F0 maximum and F0 minimum. F0 maximum is reached earlier in focus, allowing sufficient time for F0 to fall to a lower F0 point than in topic together with segmental lengthening. F0 minimum is reached earlier in focus than in topic relative to the end of the word, resulting in a longer low plateau in focus. It is argued that speakers use the longer low plateau to demarcate the informational completeness at the end of the subject noun in full-sentence answers to WHO-questions.

In contrast, children do not yet use any of the phonetic cues to distinguish topic and focus in sentence-initial position at the age of 4 or 5. At the age of 7 or 8, they become adult-like only in the use of F0 excursion and F0 minimum. These findings lend support to the hypothesis that the use of variations in gradient cues are harder to learn. Considering that children are fully adult-like in the phonological marking of topic and focus at the age of 7 or 8, it is concluded that phonetic marking is acquired later than phonological marking.

References

105


Manolescu, Alis, Daniel Olson & Marta Ortega-Llebaria. 2009. “Cues to Contrastive Focus in Romanian”. This volume.


PART II

Segmental and prosodic interactions
Prosodic structure and consonant development across languages*

Timothy Arbisi-Kelm¹ & Mary E. Beckman²
¹University of Wisconsin-Madison/²The Ohio State University

This paper relates consonant development in first-language acquisition to the mastery of rhythmic structure, starting with the emergence of the “core syllable” in babbling. We first review results on very early phonetic development that suggest how a rich hierarchy of language-specific metrical structures might emerge from a universal developmental progression of basic utterance rhythms in interaction with ambient language input. We then describe salient differences in prosodic structures across the languages being studied in a cross-language investigation of phonological development, in which we are eliciting and analyzing recordings from hundreds of children aged two years through five years who are acquiring Cantonese, English, Greek, or Japanese. Finally, we present examples of how patterns of disfluent consonant production differ across children acquiring the different languages in this set, in ways that seem to be related to the differences in metrical organization across the languages.

1. Introduction

An enormous body of work over the past half century and more highlights the role of metrical structure in aligning different types of information from different parts of the grammar. Work on languages such as English, for example, supports the existence of a hierarchy of structures such as the syllable, the stress foot, and the intonational phrase, which is parsed, in part, from the role that these structural elements play in aligning intonation patterns with morphosyntactic constituents in the planning of attentional flow during an unfolding discourse (e.g., Gussenhoven 1983; *This work was supported by an NIH traineeship to the first author and NIDCD grant 02932 to Jan Edwards, Principal Investigator of the παιδολογος project. We thank the adult and child participants who produced the utterances that were recorded on this project and the native speaker transcribers (Wai-Yi Peggy Wong, Sarah Schellinger, Asimina Syrika, and Junko Davis) who transcribed the utterances and noted the disfluencies. We thank these transcribers, Jan Edwards, Eunjong Kong, and Fangfang Li for discussion of the prosodic analyses.
Selkirk 1984; Welby 2003). Just within the phonological component, lower-level metrical structures such as the syllable are evident from their role in aligning consonant gestures with vowel gestures in order to realize features that otherwise might not be audible (e.g., Mattingly 1981; Browman & Goldstein 2000). Moreover, studies such as Levelt and Cutler (1983) and Arbisi-Kelm (2006), among many others, provide striking evidence of the integral role also of higher-level metrical structures in facilitating the fluent production of consonants. In particular, these studies show that the most notable consonant disfluencies occur around structural positions such as the onsets of pitch-accented words, where entropy is maximal. Much of this evidence on adult disfluencies involves studies of speakers of English. However, there are a few studies of phenomena such as speech errors in other languages (e.g., Kubozono 1989). These studies suggest that some metrical structures and/or the associated constraints on consonant alignment and production can differ from language to language. If metrical structures and/or consonant alignment and production constraints are to some extent language specific, how can they develop?

In this paper we will first clarify our assumptions about metrical structure and its role in organizing spoken discourse, and then review previous results on very early phonetic development that suggest how a rich hierarchy of language-specific metrical structures might emerge from a universal developmental progression of basic utterance rhythms in interaction with ambient language input. For example, while the basic motor rhythms of canonical babbling develop in a way that is fairly impervious to all but the most severe degradations in input (see literature reviewed in Oller 2000), there are recognizable differences in the “rhythmic feel” of babbling produced by infants acquiring different languages even during the first year or so of life (e.g., Whalen, Levitt & Wang 1991; Vihman 1993).

After presenting our understanding of what this literature says about development, we will then review salient differences in metrical structures across the languages being studied in the παιδολογος project (http://ling.osu.edu/~edwards), a cross-language investigation of phonological development in which we are eliciting and analyzing recordings from hundreds of children aged two years through five years who are acquiring one of several rhythmically-diverse languages, including (so far) Cantonese, English, Greek, and Japanese. We will then briefly present examples of how patterns of disfluent consonant productions differ across children acquiring different languages in this set, in ways that seem to be related to the differences in metrical organization across the languages.

2. Ontology of metrical structure

The starting point for our paper is the observation that a primary function of prosody is to provide a rhythmic scaffolding that specifies designated temporal
points of convergence and structural alignment among different components of the grammar. For instance, intonational phrase boundaries are points where talkers can align tones and gestures of consonants and vowels with respect to morphosyntactic clause boundaries. These alignment patterns are regular enough in read speech styles that some researchers (e.g., Selkirk 1984; Nespor & Vogel 1986; Gussenhoven 1992) have hypothesized a deterministic mapping between prosodic constituency and morphosyntactic constituency. Observations of other speech styles do not support the strongest version of this hypothesis (see, e.g., Schafer, Speer & Warren 2004). At the same time, the alignment patterns are regular enough that listeners can use their expectations about the mapping between phonology and syntax to parse otherwise ambiguous strings, as illustrated in (1). In most lab-speech readings of the sentences in (1), the distribution of intonational phrase boundaries (marked by a pause or by final lengthening, as well as by the tonal cues to the ends of potentially stand-alone intonation contours) resolves the morphosyntactic ambiguity resulting from the homophony between the adjective and pronoun her and between the noun pleas and the adverb please. That is, the intonational phrase break following the word pleas in (1a) coincides with the end of a syntactic phrase, making pleas the object of hear. In (1b), by contrast, the intonational phrase boundary immediately following her disallows this interpretation. This kind of alignment between edges of tonally marked phrases and edges of syntactic constituents has been observed in many languages (e.g., Lehiste 1973; Selkirk & Shen 1990; Kang & Speer 2003; Millotte, Wales & Christophe 2007).

(1)  
\begin{align*}
\text{a. } & [\text{When you hear her } \text{pleas}]_{\text{IP}} [\text{don't respond without thinking}]_{\text{IP}} \\
\text{b. } & [\text{When you hear } \text{her}]_{\text{IP}} [\text{please don't respond without thinking}]_{\text{IP}}
\end{align*}

In addition, metrical structure serves as the scaffolding for planning the pitch range relationships and the distribution of shorter-term melodic events that mark global and local discourse organization. For example, in English, pitch accents and phrase accents – the intonational morphemes that indicate local attentional foci within the discourse segment – are aligned to words and phrases by matching metrically strong positions to syntactically prominent constituents, such as lexical heads of the focus constituent, as illustrated in (2). The alignment of the pitch accent and phrase accent relative to the words in (2a) marks the apples as the focal constituent in the intonational phrase containing it, and suggests a discourse context in which the cook is about to start a round of pie-making and is considering different ingredients. By contrast, the distribution in (2b) makes pies the focal constituent of the phrase and suggests a context in which the cook has got a pile of apples and is considering various different desserts that could be made with them (see, e.g., Pierrehumbert & Hirschberg 1990; Rooth 1992). In Japanese, comparable differences in the relationship between the sentence and the larger discourse can be marked by differences in the intonational phrasing, the choice of boundary
pitch movement, and the relative pitch ranges of successive intonational phrases (see, e.g., Venditti, Maekawa & Beckman 2008).

(2) a. Please only make pies with [the apples]_{FOC}\\
\text{[} \left[ \sigma \right]_F \left[ \sigma_w \right]_F \text{]}_{IP}\\
\text{L+H* L- L%}

b. Please only make [pies]_{FOC} with the apples.\\
\text{[} \left[ \sigma \right]_F \left[ \sigma_w \right]_F \text{]}_{IP}\\
\text{L+H* L- L%}

Another important point about metrical structure is that, just within the phonological component, we find evidence of alignment between different structural elements. In cases such as (2), for example, the foot structure and associated syllable structures that are parsed for the monosyllabic prosodic word pies versus trisyllabic prosodic word the apples govern the alignment of the pitch accent and following phrase accent relative to the string of consonants and vowels in these words. Even more locally, consonant gestures coordinate with vowel postures, and differences in the coordination patterns reveal their alignment to different positions in the syllable and larger metrical structures. For instance, the production of consonant sequences will vary depending on metrical position, as schematized in (3).

(3) \[?\left[\text{æ}\right]_V\left[\text{p} \text{ l}\right]_l \left[\text{æ}\right]_V\left[\text{p} \text{ l}\right]_l \left[\text{æ}\right]_V\left[\text{p} \text{ l}\right]_l \left[\text{æ}\right]_V\left[\text{p} \text{ l}\right]_l \text{]}_{\omega} \left[\text{p} \text{ l}\right]_{\omega} \left[\text{p} \text{ l}\right]_{\omega} \left[\text{p} \text{ l}\right]_{\omega} \text{]}_{\omega}

\begin{align*}
\text{apple} & \quad \text{lips closed-open} \\
\text{applause} & \quad \text{tongue sides down-up} \\
\text{pleas} & \quad \text{glottis open-closed}
\end{align*}

When the sequence /pl/ is produced in foot-final position (as in apple) the glottal opening gesture and lip closure for the /p/ are aligned to overlap considerably less with the lingual posture of the /l/ as compared to the degree of overlap for the cluster in foot-initial position (as in applause) and word-initial position (as in pleas). Also, when the /pl/ sequence is foot-final, the tongue tip and tongue body gestures for the /l/ are sequenced in such a way that the velar constriction gesture dominates in the filter shape, resulting in the “dark” allophone of /l/. By contrast, when the sequence /pl/ is produced in foot-initial position, the tongue tip and tongue body gestures for /l/ overlap so that the “consonantal” tongue tip closure gesture can contribute to the filter shape. Evidence for this understanding of the role of metrical structure in facilitating consonant production comes from many sources, including studies of articulator kinematics in fluent read speech, such as Browman and Goldstein (1988), Krakow (1993, 1999), Byrd (1996), and
Loevenbruck et al. (1999). Some of the studies in this literature have shown layered effects of higher-level prosodic environment on consonant gestures at several levels, with longer constriction durations at intonational phrase edges compared to those at phrase-internal word edges as well as differences for different positions within a foot (Fougeron & Keating 1997; Cho & Keating 2001; Byrd et al. 2005; Bombien et al. 2006; Cho 2006). In this vein, Sproat and Fujimura (1993) have shown that the degree of overlap for foot-initial /l/ differs depending on how the foot edge is aligned to higher-level structures such as the prosodic word and intonational phrase, making for a continuum of more or less “light” allophones of /l/ at different initial positions.

The literature on speech errors provides a second source of valuable information regarding the role of metrical structure in consonant production. One key observation in this literature is that speech errors occur at different rates for consonants in different metrical positions (see, e.g., Dell 1985; Shattuck-Hufnagel 1987; Levelt 1989). For example, English errors often move or exchange consonants at foot beginnings, as shown in the example in (4), from Fromkin (1973). This type of error almost never occurs in other positions. That is, there are far fewer coda exchanges, and there are almost no exchanges between syllabic consonants and vowels (see, e.g., Fromkin 1971, 1973; Shattuck-Hufnagel 1979; MacNeilage 1998). This asymmetry between onset and coda positions is often interpreted as evidence for differences in metrical affiliation at the level of the syllable, as schematized in (5) for two competing accounts of English syllable-internal organization.

\[(4) \text{ slumber party} \rightarrow \text{lumber sparty} \]
\[
[[s [l[a] \text{V} m]_o [b [i] \text{V} ]_o ]_F \rightarrow \ [l [a] \text{V} m]_o [b [i] \text{V} ]_o ]_F
\]

\[(5) \text{ a.} \quad \text{F} \quad \text{F} \quad \text{b.} \quad \text{F} \quad \text{F} \]

Evidence that this asymmetry is an effect of metrical position per se (rather than a simpler sequential constraint on available cues in pre-vocalic versus post-vocalic position, as suggested implicitly by Steriade 1999) comes from cross-language comparison. Specifically, in Japanese, errors can target gestures in analogous mora positions in different syllables, regardless of whether the gestures are vocalic or consonantal, as shown by the example in (6). Kubozono (1989) uses this difference in error patterns between the two languages to argue for a difference in syllable-internal
metrical organization, as schematized in the tree diagram in (7), as contrasted to (5a), which is his account of English.

(6) paasento → pansento

Regardless of the structure that is posited, however, there is a notable fact about speech errors such as (4) and (6). When consonants move from one syllable to another in planning the utterance of a phrase, they are fluently aligned with the gestures of segments that are adjacent in the new environment, to make for the appropriate allophony. The glottal opening gesture of the /s/ in _lumber sparty_ in (4), for example, merges with the glottal opening gesture of the following labial stop so that the /p/ is now unaspirated, as appropriate for its no-longer foot-initial position. The /l/ that is left behind, conversely, is now fully voiced, since its lingual posture no longer is aligned relative to the glottal opening gesture of a preceding foot-initial voiceless obstruent. And analogous fluent positionally appropriate allophonic variation characterizes the pronunciation of the segments around the intrusive moraic nasal in the Japanese example in (6), as well. In particular, the velic opening gesture for the displaced moraic nasal overlaps heavily with the lingual posture for the preceding /a/ to make for a heavily nasalized vowel. Also, the nasality cannot extend into the following /s/ constriction, so that the transcription of [n] for the moraic nasal is much less appropriate here than it is in the second syllable of this word. We interpret these observations about allophonic appropriateness as evidence for the integral role of metrical structure in facilitating the fluent production of consonant gestures, particularly of gestures such as the labial release and glottal opening gestures of English /p/ which cannot even be heard unless they are timed so as to be co-produced appropriately with the gestures of other segments around a common metrical alignment point.

A third source of evidence for the integral role of metrical structure in consonant production comes from studies of disfluencies in the speech of stutterers. Following Levelt (1983), Shriberg (1999), and others, we understand disfluencies as arising from a speaker’s detection, and attempted correction, of an error in language production. Different types of errors suggest errors at different stages of planning. In the speech of stutterers, the distribution of disfluencies suggests a breakdown at the stage of assembling an articulatory plan, particularly in the planning of gestures around those metrical positions that present the most challenging constraints on alignment. For example, in English-speaking stutterers, disfluencies
Prosodic structure and consonant development

occur much more frequently in word-initial position (Brown 1938, 1945; Hahn 1942; Soderberg 1962; Taylor 1966; Weiner 1984; Hubbard 1998; Natke, Grosser, Sandrieser & Kalveram 2002), and in stressed words more often than in unstressed words (Brown 1938; Bergmann 1986; Weiner 1984; Wingate 1988; Prins, et al. 1991; Natke, Grosser, Sandrieser & Kalveram 2002). Moreover, there are also metrically-conditioned differences between different types of stressed words. Specifically, there are more disfluencies on stressed words that are aligned to pitch accents in the intonation contour of the utterance (Arbisi-Kelm 2006).

We attribute this last result to the exigencies of accent production in English. Consonant production in pitch-accented words is challenging for two reasons. The first is the set of linguistic conventions identified by researchers such as de Jong (1995) about the carefully “hyperarticulated” quality of the consonants and vowels associated to pitch-accented syllables in languages such as English. Even if there were no such conventions, however, we would expect accented words to be a locus of disfluency, because of the added burden of aligning the laryngeal and oral gestures for the consonants and vowels of the word together with the laryngeal and respiratory gestures for producing the tone pattern within the backdrop pitch range specified at that point in the discourse.

In summary, converging evidence from fluent and disfluent speech suggests to us that consonant production depends on a facility to coordinate gestures at multiple time scales, from the rapid sequence of raising and lowering movements that flick the tongue tip against the alveolar ridge to make the 20–30 ms closure for an alveolar tap to the gestures of the respiratory and laryngeal system that specify the pitch range over stretches of speech that can extend for 5–10 seconds.

Our understanding of the hierarchy of metrical structures for any given language is that these structures provide a conventionalized schema for organizing the planning of speech production in real time. Like Ferreira (1993) and Keating & Shattuck-Hufnagel (2002), that is, we interpret speech error data as telling us that metrical structures for larger prosodic constituents, such as accented positions within intonation phrases, are assembled relatively early in the production process, before word-specific gestural ensembles are retrieved and aligned relative to the rhythmic frames specified for the larger constituents. Moreover, we understand from the comparison of intonation systems and word-level prosodic templates across languages that the metrical hierarchy is highly conventionalized and language-specific.

This language specificity raises an important question: how exactly can the higher-level metrical structures be acquired so that the child can begin to produce the consonants that distinguish words of the ambient language? In the next section, we suggest one route by which language-specific metrical organization can emerge from infants’ developing motor control over their respiratory and laryngeal
systems, and over movements of their lips, tongue, and velum, in interaction with the voices and faces of speakers in the infants’ environment.

3. Ontogeny of metrical structure

Evidence for the emergence of rudimentary metrical organization can be found very early in speech development, during the pre-babbling and canonical babbling stages as identified by researchers such as Oller (1980; 1986; 2000), Stark (1980), and Koopmans-van Beinum & van der Stelt (1986). In the “phonation stage” (using Oller’s terms), infants practice their control of the most transparently autonomous articulatory gestures, exploring the motor space for different patterns of fundamental frequency or laryngeal source quality in conjunction with different vowel-like resonances. From around 2–3 months of age, during the “gooing stage”, children advance to articulations of consonant-like sounds, including vaguely [k]-like releases that can be produced with a raised jaw and tongue-filled oral tract. At these early stages, the dominant audible rhythm is not internal to the infant’s vocalization. Rather, it is the alternation of imitative turn-taking between the infant and the mother that can ensue if the mother responds to the baby’s coos with contingent imitation of the baby’s phonatory gesture and/or the baby’s resonance gesture (see, e.g., Papoušek & Papoušek 1989; Masataka 1993, 2003). Papoušek, Papoušek & Symmes (1991) even propose that there are cross-cultural commonalities in the melodic shapes that mothers use to engage an infant’s attention and invite a bout of vocal turn-taking at this age. They cite in support of this proposal their results showing that Mandarin-Chinese speaking mothers will suppress lexical content of their utterances to their infants in order to enable the expression of these tunes.

A bit later, starting at about 6 months, the rudimentary consonant-like releases of the “goo” stage become coupled to more rhythmically consistent mandibular oscillation patterns. At this stage of “canonical babble”, parents in homes where the babies are acquiring a spoken language are much more likely to describe their children’s vocalizations in terms of the consonants and vowels of the ambient language, as in Darwin’s (1877: 292) description of one infant’s progression from the earlier stages into the canonical babbling stage:

At 46 days old, he first made little noises without any meaning to please himself, and these soon became varied. An incipient laugh was observed on the 113th day, but much earlier in another infant. At this date I thought, as already remarked, that he began to try to imitate sounds, as he certainly did at a considerably later period. When five and a half months old, he uttered an articulate sound “da” but without any meaning attached to it.
Researchers have long noted that the onset of this canonical babble (or “reduplicative babble”) comes when the baby is maximally engaged in a regular rhythmic exploration in general, waving hands and feet, shaking rattles, and so on (e.g., Thelen 1979, 1991; see review in Ejiiri & Masataka 2001). While the timing of the mandibular cycle makes the infant’s vocalizations at this stage sound very much like consonant-vowel alternations, there are strong constraints on what consonants can combine with what vowels. These constraints suggest that the infant is still controlling tongue posture only at the whole-utterance level, in what Davis and MacNeilage (1995) refer to as “frame dominance” – a precursor to the “variegated babble” seen in many infants, when shorter term rhythms allow the infant to begin to control sequences of labial and lingual postures on a “syllable-by-syllable” basis.

Summarizing this work, we can say that observations of these stages of pre-linguistic babbling strongly support the idea that the intonation phrase and the core syllable are universal units of spoken language simply because they harness rhythmic structures that emerge universally in normal development. The intonation phrase (or “breath group” as Lieberman 1967, calls it) emerges as a very young infant explores the auditory consequences of coordinating the respiratory cycle with oral gestures for sustained egressive phonation. The core syllable, similarly, emerges as the somewhat older infant explores the auditory consequences of coordinating a basic mandibular oscillation with lingual and labial constrictions for a fluent sequence of consonants and vowels.

While these structures have a universal ontogenetic basis, however, the metric-al structures that eventuate do differ across languages. The tone shapes that mark the “breath group” are specific to each language variety, and so are the dominant syllable structures. Also, while there is some evidence that the core syllable continues to have privileged status even in languages with more complex syllable types, this privileged status is manifest in highly language-specific ways because of minimal word templates, among other things. For example, in English, where CVC is the single most common word shape, the privileged status of CV is evident in the way that CVC words are restructured in child-directed speech, as in *doggie* for *dog* and *kitty* for *cat*. In Japanese, by contrast, the privileged status of the core syllable is evident in the adult lexicon, where CVCCV is one of the most common word shapes. In child-directed speech, on the other hand, words with heavy initial syllables are favored, so that *inu ‘dog’* is realized as *wanwan* and *neko ‘cat’* as *nyanko*.

An obvious question that arises, then, is the following: how and when are these universal rhythms tuned to become the metrical structures specific to the phonological grammar of the ambient speech community? We believe that the universal rhythms are entrained to ambient language structures very early, because of the critical role of auditory feedback in normal development of motor control for
Production and perception capabilities that ultimately lead to speech are initially largely separate, but they begin to be coordinated (integrated) within the first few months of life. The integration of the two systems also interacts with the child’s linguistic background, such that the child’s exposure to the sounds around him/her eventually influences the child’s own pattern of vocalization.

Much research on early infant vocalization highlights the importance of auditory feedback. For example, Langlois, Baken and Wilder (1980) attributed the onset of a sustained exhalation phase of the respiratory cycle (and consequently longer vocalizations) entirely to anatomical factors – especially to the growth of the rib cage. However, the crucial role of audition becomes apparent when examining how utterance duration changes when there is little or no auditory feedback. Clement, Koopmans-van Beinum and Pols (1996) found that while both typically-developing and hearing-impaired children showed the expected increase in mean utterance duration around 3–4 months, the hearing-impaired children showed a significantly smaller increase. The authors ascribed this difference to a lack of auditory feedback, which is consistent with Lieberman’s (1986) suggestion that insufficient laryngeal muscle exercise during this period could result in reduced ability to manipulate sub-glottal air pressure. In other words, normal development of the universal basis of the “breath group” depends on auditory feedback at 3–4 months.

Analogous effects of auditory experience are also observed for the later development of canonical babbling rhythms. For example, in a study of two monozygotic twins – one profoundly hearing-impaired and the other with normal hearing – Kent, Osberger, Netsell and Goldschmidt Hustedde (1987) found key production differences at 8, 12, and 15 months. In addition to producing a smaller range of vowel formant frequencies at each of the three recording sessions, the hearing-impaired twin showed an overall later onset of canonical babbling. A larger study by Oller and Eilers (1988) of 21 infants with normal hearing and 9 deaf infants replicates this effect of input. All of the infants with normal hearing began canonical babbling between 6–10 months of age, while none of the deaf infants began this stage before 11 months.

Given this critical role of auditory input for the onset of canonical babbling, we might expect to see entrainment to ambient language speech patterns before the onset of language. And, indeed, influence of language-specific autosegmental content on infants’ vocalization patterns is found as early as 10 months. In their study of 10-month-old infants growing up in Arabic-, Cantonese-, English-, or French-speaking homes, de Boysson-Bardies, Hallé, Sagart and Durand (1984) found that
the distribution of formant frequencies measured in vowel-like intervals in each infant’s canonical babbling reflected the frequencies of different vowels in the lexicon of the ambient language. Transcribed consonant place frequencies in babbling reflect a similarly early influence of adult language input. In their cross-linguistic study of infant vocalization patterns, de Boysson-Bardies and Vihman (1991) found that 9–10-month-old infants growing up in English-, French-, Japanese-, or Swedish-speaking homes produced different proportions of labial consonants relative to lingual consonants, in keeping with the different distributions of labials in the adult lexicons of the four ambient languages.

Analogous results have been found for the time course of language-specific effects in the prosodic domain, as well. In a longitudinal study of babies growing up in French- or English-speaking homes, Levitt and Wang (1991) found no cross-language differences in “syllable” durations during the pre-canonical stages at 4–6 months. During the later reduplicative babbling stage, however, the English-acquiring children produced significantly shorter final syllables than did the French-acquiring infants, reflecting the predominantly trochaic pattern of English words, as opposed to the predominantly iambic rhythms of French. A companion study of the fundamental frequency patterns produced by these infants lends further support to this interpretation of the duration patterns in terms of the language-specific rhythmic patterns. Specifically, Whalen, Levitt and Wang (1991) report consistent differences in the pitch contours for two- and three-syllable reduplicative babbling trains, such that the French-acquiring babies produced a mix of both sequence-final rises and sequence-final falls, while the English-acquiring infants produced almost exclusively falling melodic contours. Parallel results for perception include the marked preference for trochaic words over iambic words in 9-month-old English-learning infants (e.g., Jusczyk et al. 1999).

Such cross-language prosodic differences in reduplicative babbling are important because canonical babbling provides the basic “vocal motor schemes” of the first words (Vihman et al. 1985; Davis et al. 2000; McCune & Vihman 2001). That is, the ability of infants to integrate the longer-term coordination of subglottal pressure modulation and laryngeal control for the global and local pitch patterns of intonation phrases together with the shorter-term control of jaw, tongue, and lip movements for vowels and consonants is prerequisite to the later fluent production of one-word utterances. Therefore, the fact that there are measurable effects of ambient language prosody on the tunes and rhythms of reduplicative babbling months before the infant makes the connection to meaning strongly supports the primacy of prosodic structure in constraining the development of motor control for later word production. Therefore, we are not surprised to find that the first recognizable words that young toddlers produce also reflect the prosodic structures of the languages that they are acquiring, even as they reflect more universal
constraints on the motoric complexity of early utterances. For example, Vihman, DePaolis and Davis (1998) note differences in the prosodic structures in early multi-syllabic word productions by French- and English-acquiring toddlers at 13–20 months. Mirroring the stress patterns of the ambient language, the English-acquiring babies produced only recognizable trochees, while the French-speaking babies produced recognizable iambics. The difference was especially striking in cases where a child truncated a long word or re-configured a word’s underlying syllabic structure to match the desired stress pattern, as in the French-acquiring baby’s production of [pəˈpjɪ] for the adult target *papillon* ‘butterfly’ or the English-acquiring baby’s production of [‘wənə] for the adult target *around*. While older toddlers tend not to show such dramatic prosodic reorganization, we might still expect to see constraints from the prosodic organizations that the child has mastered interacting with and affecting the child’s productions of one-word utterances. In Section 5, we will look in greater detail at several examples of disfluency patterns produced by young children acquiring English or one of three other languages, particularly with respect to how these error patterns are shaped and constrained by both language-specific and language-general aspects of metrical organization. In the next section, we describe the larger database of productions from which these examples are drawn.

4. The παιδολογος project – cross-linguistic research on phonological acquisition

The examples that we will discuss were recorded as part of a larger project in which we are comparing accuracy rates for productions of word-initial lingual obstruents in a range of following vowel contexts across a variety of languages. For each language, the subjects we record are twenty young adults (aged between 18 and 30 years) and 100 children aged from 2 years through 5 years. The target consonants are elicited using a picture-name repetition task that allows us to record them in real words that are not likely to be familiar to the youngest children, as well as in the highly familiar picturable real words used in most word-naming tasks. In this way, we can sample the consonants evenly across following vowel environments even when one coarticulatory context is relatively rare. The task even lets us elicit the target consonants in nonsense words so that we can compare the children’s accuracy rates across pairs of languages in all coarticulatory conditions of interest, whether the consonant-vowel context is phonotactically licit in both languages or just in one.

Although we have recently begun to record productions by children acquiring Korean, Mandarin Chinese, and French, the first languages that we chose to study
were Cantonese, English, Greek, and Japanese. These are the languages for which we have the most complete analyses to date, and therefore they are the languages from which we will draw all of our examples for this paper. These initial four languages (and the subsequent three) were chosen because we have access to children acquiring them and because each has a rich inventory of lingual obstruents which can be compared to identically transcribed consonants with different contrastive properties, different allophonic patterns, or different phonotactic frequencies in one or more of the other languages. For example, all of the languages have at least one sibilant fricative that is transcribed as /s/, and in all of the languages but Cantonese and Korean, this alveolar fricative contrasts with at least one other voiceless lingual fricative at a different place of articulation. In Greek and English, /s/ contrasts with a more anterior dental /θ/ as well as with a more posterior fricative, which is the dorso-palatal /ç/ in Greek and the coronal post-alveolar /ʃ/ in English, whereas in Japanese /s/ contrasts only with two more posterior fricatives – an alveolo-palatal /ɕ/ as well as the dorso-palatal /ç/. Moreover, in English, both /s/ and /ʃ/ occur readily in all following vowel environments, whereas in Japanese, /s/ does not occur before /i/ and /ɕ/ is only marginally attested before /e/. All the languages also have voiceless dental or alveolar and “plain” dorsal stops that can be transcribed as /t/ and /k/, as well as at least one other voiceless dorsal stop with a contrastive “secondary” labial or palatal articulation.

Our initial analysis is a categorical judgment (by a native speaker of that language who is a trained phonetician) of the accuracy of each target word-initial consonant and of the following vowel. That is, the transcriber opens the audio file for a recording session in Praat, along with a TextGrid file in which the interval for each word has been marked off and tagged with an initial broad phonemic transcription. The transcriber zooms in on each target word in turn, listens to the word production, and compares the word-initial consonant and the following vowel to their target transcriptions. If a sound is judged to be correct, the phonemic transcription is accepted. Otherwise, the transcriber analyzes the error to provide one of the types of tags listed in (8), the first three of which are types of “substitution” that require a transcription of the substituted sound. Both the automatically provided initial broad phonemic transcription of the target sounds and the transcribers’ analyses of the substitution errors use the WorldBet ASCII encoding of the IPA devised by Hieronymous (1994), and both adhere to conventions that sometimes are at odds with the usual phonemic analysis for the language. For example, the Greek front dorsal stop that occurs before /i/ and /e/ and that contrasts with “plain” /k/ before the other three vowels of the language in words such as /’kɔski/ ‘kiosk’ in (8) is transcribed as a palatalized stop rather than as a palatal stop, as in the usual phonemicization of Greek provided by Arvaniti (1999) and others. The English initial stop in words such as cute and key also is transcribed as
a palatalized velar, and the English initial stop in words such as *quake* is transcribed as a labialized velar.

(8) $ plus a broad transcription, for a within-inventory substitution error
e.g., $ts$ for a /ts/ for /kl/ substitution for Greek /kłoski/ 'kiosk'
+ plus a narrow transcription, if the substitution goes outside the
language's inventory
e.g., +tc] for a [tc] for /kl/ substitution for Greek /kłoski/ 'kiosk'
: separating some pair of IPA tags
e.g., +tc] $ts$ for a production intermediate between the above two
deletion for a segment that is simply not pronounced
distortion for an error that cannot be captured by any IPA symbol

This transcription of labialized or palatalized stops bears more explanation. By adopting these transcription conventions, we are equating the configuration of lingual and labial gestures at the beginning of English *quake* with the configuration of lingual and labial gestures at the beginning of Cantonese /kwa:/ 'melon' and we are equating the configuration of tongue body gestures at the beginning of Greek /kłoski/ 'kiosk' and of English *key* and *cute* with the configuration of tongue body gestures at the beginning of Japanese kippu /kippu/ 'ticket' and kyū /kyū/ 'nine'. That is, for the sake of cross-language comparison, we are analyzing the CV sequence in words such as *cute* as having the prosodic structure and alignment constraints in (9a). In this analysis, the front place of constriction of the dorsal stop in English *cute* is parsed as a property inherent to the consonant, just as in the contrastively palatalized dorsal stop in Japanese kyū. A more traditional analysis is shown in (9c). Here, the tongue fronting gesture is parsed as a feature of the first target in a following diphthong, which is co-produced with the consonant, yielding the “front allophone of /kl/” that is also seen in words such as *key*, as in (9b). The schemata in (9) are intended to highlight the idea that the difference between English *cute* and Japanese kyū is a prosodic difference rather than a difference in intrinsic featural content. Comparing these “palatalized velars” across the two languages gives us a way of evaluating the consequences of such cross-language differences in prosodic organization at the level of the syllable.

(9) a. b. c. tongue body
\[\begin{array}{c}
\text{kyū} \\
\text{key} \\
\text{cute}
\end{array}\]
The languages examined also have other, more obviously prosodic differences involving structures above the syllable. For instance, Greek and English both have intonational morphemes (pitch accents) that are constrained to anchor to syllables that are rhythmically prominent (i.e., have “lexical stress”). However, the word-level rhythms differ. English words are predominantly one or two syllables long, and words longer than one syllable show a trochaic bias, grouping alternating strong and weak syllables into bimoraic feet (Hayes 1980; Halle & Vergnaud 1987; Kager 1989). As a consequence, the most typical word shapes are a stressed syllable alone or a disyllable with initial stress, so that the consonant marks not just the beginning of the intonation phrase (in the citation form utterances we elicited) but also the beginning of a stress foot, with all of the concomitant constraints on precise alignment to coordinate the constriction gesture for the consonant with the lingual and labial postures specified for the following full (and typically pitch-accented) vowel.

This contrasts with Greek, which has almost no monosyllabic words, and which has a trisyllabic stress window aligned to the end of the word, with no phonetic evidence for stress alternations elsewhere (see Joseph & Philippaki-Warburton 1987, and other references cited in Arvaniti & Baltazani 2005). Many more words of Greek, therefore, have unstressed initial syllables. Furthermore, while in English only /ə/ and its variants are unstressed and subject to deletion, any vowel in Greek may be unstressed, and both /i/ and /u/ may be lenited in unstressed environments (Arvaniti 1994). This is illustrated in Figure 1. The word /ˈcilja/ ‘lips’ in the spectrogram on the left has a stressed initial syllable, with a fully realized vowel /i/, whereas the word /çiˈmonas/ ‘winter’ in the spectrogram on the right has stress on the second syllable, and the unstressed vowel /i/ is significantly reduced.

![Figure 1. Spectrograms and WorldBet transcriptions of a Greek adult speaker’s productions of a word beginning with stressed initial target /çi/ (left) versus one beginning with an unstressed and lenited initial target /çi/ (right)](image-url)
The consonantal contexts that favor this lenition of high vowels in Greek are very similar to those that promote “high vowel devoicing” in Tokyo Japanese, and the acoustic patterns that result are similar. The Japanese phenomenon differs prosodically, however, because Japanese does not have anything like stress in Greek. There are localized pitch events that Greek- and English-speaking second language learners of Japanese assimilate to the “pitch accents” in their native intonation systems. However, the pitch accents of Japanese are simply lexically specified tone patterns that are not associated to prominent syllables and which interact only very indirectly with the prosodic mechanisms for focus-marking (see Venditti, Maekawa & Beckman 2008, for a review). It is not at all uncommon for high vowels in lexically accented syllables to be deleted, as is clearly evident in the numbers reported for the Corpus of Spontaneous Japanese by Maekawa and Kikuchi (2005) as well as in many earlier studies of lab speech. Moreover, there is a rising initial boundary pitch movement covering the first one or two moras of every well-formed accentual phrase, and hence of every utterance, whether or not there is a lexical accent anywhere in the utterance. Therefore, the Japanese word-initial consonants in our database are not in a position of relaxed prosodic alignment constraints even in syllables with devoiced or deleted vowels.

Finally, Cantonese also has segmental lenition and even whole vowel deletion, in a phenomenon called “syllable fusion” (Wong 2004, 2006). This reduction phenomenon is not constrained by word-level prosodic prominence, since Cantonese does not have anything like the contrast between stressed and unstressed syllables of English and Greek. This description may make Cantonese syllable fusion seem like Japanese vowel devoicing. However, the two phenomena are quite different, as are the two prosodic systems more generally. Every syllable of Cantonese is prosodically strong in the sense that every syllable bears a full lexical tone. The lack of tonally unspecified syllables differentiates Cantonese not just from Japanese but also from most varieties of Mandarin Chinese as well as from Wu dialects such as Shanghai Chinese. In these other Chinese varieties, prosodic words typically are at least two syllables, and high vowel deletion like that seen in Greek occurs on weak syllables that have “neutral tone” – i.e., syllables that either are intrinsically not specified for tone, as in the second syllables of Mandarin dōufu ‘tofu’ and yìsi ‘meaning’, or that “lose” their tone specification in the word-formation process, as in the middle syllable of yào bu yào ‘want?’ (examples from Peng et al. 2005). Cantonese differs from these other varieties of Chinese in having far fewer disyllabic (and longer) words as well as in not having neutral tone syllables. Syllable fusion is much more strictly a post-lexical process, but even the most extreme cases of reduction and vowel deletion typically occur without tone loss. That is, even in cases where the medial consonants have been completely deleted and the two vowels
merged into a single sandhi form, both syllables’ lexical tone specifications are preserved, so that the syllable count is effectively unchanged (see examples and description in Wong, Chan & Beckman 2005).

Our original motivation for eliciting several productions of each word-initial lingual obstruent in each of several vocalic contexts was to be able to compare accuracy rates across the different languages. By comparing accuracy rates for ostensibly the same CV sequence across languages that have different phonotactics or different frequencies for the sequence, we can begin to tease apart the effects of language-specific phonotactic probabilities from any language-universal effects of the intrinsic difficulty of producing that particular constellation of consonant and vowel gestures (see Edwards & Beckman 2008). Given the differences in prosodic structure outlined in this section, we wonder whether the error patterns might also be informative. Even if children acquiring different languages make the same number of errors for a particular CV sequence, might they show differences in the types of errors that they make if the same autosegmental content is parsed differently by the prosodic organization of the language? For example, given the different prosodic analyses in (9), we might expect Japanese-acquiring children and English-acquiring children to latch onto different strategies for producing this difficult assemblage of lingual gestures. Moreover, even when there are no obvious differences at the level of the syllable, different demands for more or less precise coordination of consonant, vowel, and tone might lead to different patterns of disfluency. In order to be able to explore these possibilities, we encourage the transcribers to note examples of different types of disfluent productions on a “notes” tier, and we are developing consensus analyses and conventions for tagging recurring types. We have culled interesting examples of these recurring types which we present in the next section.

5. Disfluencies and deletions in the παιδολογος recordings

One of the disfluency tags on the “notes” tier is the tag “E” for an initially incorrect or very “effortful” production of a target fricative or affricate that eventually homes in on an acceptable target constriction. This is the disfluency type that seems most similar to the stereotype of adult stuttering. In the example in Figure 2, an English-speaking child produces an “E”-tagged disfluency on the initial /s/ in the trisyllabic nonword /ˈsɛviˌʃæʃ/. The “effortful” nature of the production is clearly evident in the acoustic pattern of an initially less strident interval followed by a momentary dip in amplitude as the child apparently repositions the tongue body or tongue blade to better direct the airstream to hit the edge of the incisors downstream of the constriction.
We originally devised the “E” label so that transcribers could analyze cases such as the /s/ in Figure 2 as different from fluent productions of target fricatives and affricates, while still recognizing that the child eventually achieves some configuration of lingual and laryngeal postures that is recognized as the target strident sound. The “E” tag, therefore, is reserved for target fricatives and affricates, where visible (and audible) changes in amplitude and spectral shape over the course of the turbulent interval can reveal the child’s ongoing struggle with the difficult aerodynamic requirements of these sounds.

Figure 3 is an example of a different analysis category – the “split-CV” – which we now think reflects a similarly effortful struggle to coordinate the lingual and laryngeal gestures for a target stop with the tongue posture for the following vowel. In contrast to the rapid succession of more or less fluent stop releases that define the stereotypical English-speaking adult stuttering pattern, the children’s disfluent stop productions seem much more effortful and less organized rhythmically. In Figure 3, for example, the child successfully releases the dorsal stop closure into the lingual posture for the target /a/, but fails to initiate voicing. There is a nearly 300-ms period of “aspiration” which would be interpreted, if this were Japanese, as a fluent devoiced vowel. The child then overshoots the laryngeal adduction target for voicing, so that there is a 180-ms pause before another short interval of “aspiration” and then finally the full stressed target vowel. This creates the percept of an epenthetic vowel followed by a glottal stop, with a correspondingly increased syllable count.
In English, such “split-CV” cases occur especially frequently when the child attempts the particularly demanding coordination involved in sequencing the lingual and laryngeal gestures for a fluent aspirated palatalized velar stop before the labial and lingual gestures for a back vowel, either in the legal sequence /kʰu/ – traditionally analyzed as /kʰ/ followed by /ju/ as in (9c) – or in the “illegal” sequence */kʰo/ which we could elicit only in nonwords. Figure 4 shows such a case. There is the same apparent epenthetic syllable as in Figure 3, and also a perceived fronting substitution of /tʰ/ for the target dorsal.

![Figure 4](image)

Figure 4. A “split-CV” production of */kʰo/ in the English non-word /ˈkʰoˌzəm/

The error patterns that Japanese-acquiring children produce for the analogous sequences in their language show two interesting differences from the English error patterns in Figure 4, both of which are illustrated in Figure 5. First, the more common place error corresponding to the fronting to [tʰ] in Figure 4 is the substitution of an alveolopalatal affricate, as in the production shown in the left panel of Figure 5. Tsurutani (2004) also notes this substitution pattern, which mirrors a frequent substitution pattern for the palatalized velar in the less challenging sequence /kʰi/ in studies such as Nakanishi, Owada and Fujita (1972). Thus, while in our data the English palatalized velar stop errors are almost without exception moraic, analogous target sequences transcribed for the forty Japanese-acquiring children to date are predominantly consonant substitutions.

Second, in cases that are analogous to the “split-CV” aspect of the example in Figure 4, the mistiming does not involve the laryngeal gesture. Instead, the palatal place-of-articulation gesture is extended into the voiced vocalic interval, creating the percept of an [i] vowel before the transition to the unrounded back vowel, as in the production shown in the right panel of Figure 5. Although this mistiming increases the perceived syllable count, the prosodic reorganization does not sound nearly as disfluent as the voiceless vowel followed by glottal stop that the English-native-speaker transcriber tagged with the “split-CV” label in Figure 4. These differences between the two languages suggest that it is partly the added requirement of controlling the alignment of the laryngeal gesture for the English aspiration contrast in word-initial stressed syllables that makes the English sequence more challenging.
Further support for this suggestion comes from Cantonese disfluencies that are analogous to the English cases in Figures 3 and 4. Figure 6 gives an especially illuminating example. The child backs the initial /t/ to [k] and then seems to hesitate momentarily before initiating voicing half-way through the second target in the following diphthong. Although there is a 130-ms interval where /a/-like formants are excited by soft [h]-like turbulence, followed by a 200-ms interval where /u/-like formants are excited by even softer turbulence, the Cantonese transcriber does not perceive this as an aspiration error. Instead, she tags it as a deletion of the first (short /a/) target in the diphthong. That is, since there is voicing to carry the tone only during (the second half of) the interval where the tongue body and lips are postured for /u/, this is the only part that can be counted as a vowel. Or, to put it another way, because Cantonese, unlike English, does not have unstressed syllables that license vowel devoicing, the hesitation after the release of the stop cannot be perceived in terms of a prosodic reorganization that increases the syllable count.

A similar cross-linguistic difference in prosodic licensing patterns is also observed in child productions of labialized velar stops. As illustrated in the spectrogram in Figure 7, the labio-velar sequence /k^w^h/ – traditionally transcribed as /k^h/ followed by /w/ – of the English word *quake* is here produced with a centralized vowel epenthesized between the velar and labial gestures. Like in the “split-CV”
cases, the transcriber again interprets this as the addition of an extra syllable – a repair strategy also observed in the speech of aphasics in Buchwald et al. (2007).

Such error types have been fairly common in the data from the English-learning children, particularly among the youngest children. By contrast, none of the Cantonese-learning children transcribed so far have been noted as making this type of vowel insertion error. The predominant error for /kw/ in our database (as in norming studies for Cantonese phonological acquisition tests such as So 1993) is simply to delabialize, substituting [k] for /kw/. As in the Cantonese example in Figure 6, the absence of tone prohibits the licensing of a rounded schwa release as a syllable in Cantonese.

![Figure 7. A case of vowel insertion in the production of */kʰweɪk/ for the English target word */kʰweɪk/*](image)

All of these differences in interpretation of ostensibly identical misalignment patterns across languages highlight the close dependency between fluent production of segments and mastery of the language-specific prosodic constraints on gestural coordination. They also raise an important question about the act of transcription. As suggested earlier in our discussion of Japanese kyū and English cute in (9), consonant and vowel gestures do not carry their prosodic affiliations on their sleeves. Assigning a particular segmental transcription to a child’s production implies a particular prosodic analysis, and different transcribers can disagree on the prosodic analysis. The production in Figure 8 is a case in point.

![Figure 8. Vowel devoicing in a production of English nonword */ʃuˌɡɪmɛɡ/*](image)
This production of the English nonword /ʃuɡɨˌmeɡ/ is another case of an “E” disfluency. Unlike the example in Figure 2, however, here the effortful struggle to produce the aerodynamic conditions necessary for strident turbulence results in a failure to initiate voicing after the (eventually successful) configuration of gestures for the /ʃ/ constriction is released into the wider oral passage for the following vowel. One English speaker (the second author) interpreted this failure to voice the /u/ in terms of a prosodic reorganization that displaced the stress to the second syllable. Another English speaker (the first author) picked up on residual cues to the stress pattern, such as the high intensity during the fricative in alternation with the low intensity on the weak vowel in the second syllable, to perceive this disfluency instead as an illegal vowel lenition in a stressed syllable.

We might be tempted to ascribe such disagreements to the different backgrounds that any two transcribers necessarily bring to the task of transcription, even when they share a common native language. However, the examples in Figures 9 and 10 show that there can be truly ambiguous cases, where there is no good basis for choosing between two different prosodic analyses for two different transcriptions. In our Greek data, cases of “effortful” productions of fricatives were most frequent in four-syllable non-words, which necessarily have initial unstressed syllables, given the three-syllable window for stress. Such four-syllable forms with antepenultimate stress are very rare (although non-initial stress is not rare), and the children find the non-words with this shape particularly challenging, as illustrated in these two figures.

**Figure 9.** Deletion of initial unstressed syllable in Greek nonword /di’samonis/

**Figure 10.** Ambiguous case that can be analyzed either as deletion of the initial unstressed syllable, or as vowel deletion with merger of the features of the resulting /ds/ in a disfluent production of the Greek nonword /da’samonis/
In the example in Figure 9, the child appears to have deleted the entire first syllable of the target form /diˈsamonis/, and then produced a slightly distorted /s/ from the second syllable. By contrast, in the example in Figure 10, since there is clear evidence of an initial stop burst, the interpretation of the disfluency is ambiguous. Does this [ts] sequence represent a complete initial CV deletion, and the substitution of the affricate [ts] for the now word-initial target /s/? Or does the [ts] instead come about via a less substantial reduction of the first syllable – i.e., a “devoicing” of the vowel as well as of the initial /d/? And if so, is the original syllable count preserved so that the apparent [ts] sequence is interpreted as a sequence of two syllable onsets, as it would be if the “devoiced” vowel were /i/ or /u/? Or does the unexpected lenition of a target low vowel (observed sometimes in Japanese but never reported in Greek) effectively change the syllable count, so that the resulting [ts] is now a merging of the onset consonants into the affricate [ts]? Our Greek transcriber entertained all three possibilities, but did not find any one of them more compelling than the others.

6. Summary and conclusion

A key insight of Autosegmental-Metrical theory is that metrical structures are parsed from the way in which they license the choice of autosegmental content from the language-specific inventory of paradigmatic contrasts, and govern the syntagmatic alignment of different autosegmental content specifications with each other. As children acquire the ambient spoken language, they must learn the metrical structures as well as the inventory of autosegmental content specifications specific to their target language. In this paper, we have reviewed a few examples of young children’s disfluent productions of consonants in positions where the gestural alignment patterns or the prosodic interpretation of the gestural alignments is particularly challenging. We have tried to show how such a comparison of young children’s mis-productions of consonants cross-linguistically can illuminate the interaction between ostensibly universal constraints imposed by immature motor control systems and the language-specific metrical structures and autosegmental content inventories. In future work, we hope to further uncover these processes and interactions, in order to gain a greater understanding of the mechanisms available to a young speaker in planning and generating utterances.

References


Prosodic structure and consonant development


Prosodic structure and consonant development 135


Rhythmic and prosodic contrast in Venetan and Sicilian Italian*

Laurence White¹, Elinor Payne² & Sven L. Mattys¹
¹University of Bristol/²University of Oxford

We compared the Italian of speakers from the Veneto, in the north of Italy, and from Sicily, in the far south, looking for evidence of rhythmic and prosodic differences. We found no reliable differences in scores for rhythm metrics (VarcoV, %V, VarcoC) for Venetan and Sicilian, with both varieties having scores similar to French and indicative of a greater durational marking of stress than Spanish. However, we found much stronger prosodic timing effects in Sicilian Italian, with stressed vowels in nuclear utterance-final position twice as long as in prenuclear utterance-medial position. We also found evidence of differential patterns of vowel reduction: Sicilian showed greater modulation of F1 and F2 values according to stress and prosodic position, indicating greater vowel centralisation in prosodically-weak contexts than in Venetan Italian. Overall, the results indicated greater prosodic contrast in southern Italian, and suggest that multiple factors contribute to the perception of rhythmic differences.

1. Introduction

1.1 Rhythm metrics and rhythmic typology

Alternation between stressed and unstressed syllables has been held to underpin listeners’ perception of speech rhythm and of rhythmic differences between languages, at least for the well-studied languages of Europe. It is certainly the case that Romance languages such as Spanish have less durational contrast between stressed and unstressed syllables than Germanic languages such as English (e.g. Dauer, 1983). Germanic languages tend to have more complex syllable onsets and codas than do syllable-timed languages, particularly in stressed syllables. They also tend to have more marked vowel reduction in unstressed syllables, associated

*This research was supported by British Academy grant SG-42911 to Sven Mattys. We would like to thank Maja Roch, Sebastiano Grasso and Mari De Agostino for help with recordings, the organisers and delegates of Phonetics and Phonology in Iberia 2007 for interesting discussions, and three anonymous reviewers for comments on previous drafts of this paper.
with a greater difference in stressed-unstressed vowel duration. Rhythm metrics such as VarcoV (standard deviation of vocalic interval duration divided by the mean) and %V (proportion of total utterance duration comprised of vocalic intervals) exploit this variation to differentiate languages like Spanish and French from Dutch and English (Ramus et al. 1999; Dellwo & Wagner 2003; White & Mattys, 2007a). Such results offer some support for the “syllable-timed” vs. “stress-timed” typological distinction (Pike 1945), but gradient variation in scores between languages within “rhythm classes” suggests that rhythm cannot simply be classified dichotomously, at least from a production perspective. Furthermore, it is clear that isochrony of syllables in “syllable-timed” languages or of stress-delimited feet in “stress-timed” languages, which was originally held to underpin this categorical distinction, is not observed in speech production (e.g. Dauer 1983).

Gradient variation in rhythm scores is also evident within languages. For example, Singapore English, which has been described as relatively “syllable-timed” (e.g., Tongue 1974), was shown to have a lower vocalic pairwise variability index (nPVI-V: mean durational difference between successive pairs of vocalic intervals) than standard southern British English (Low, Grabe & Nolan 2000). A subsequent study indicated, however, that the difference in scores was relatively small compared to that between, for example, Spanish and English (Grabe & Low 2002). In other cases, the rhythmic contrast within languages has been found to be more marked: Welsh Valleys English, suggested to be relatively “syllable-timed” (e.g. Mees & Collins 1999), was shown to have VarcoV and %V scores intermediate between standard southern British English and Castilian Spanish (White & Mattys 2007b). Sometimes differences between accents of a language defy simple classification along a stress-timed vs. syllable-timed continuum of rhythm scores, however: for example, European Portuguese has been shown to have higher variation in consonantal interval duration than Brazilian Spanish, but lower variation in vocalic interval duration (Frota & Vigário 2001).

Differences in rhythm scores have been shown to be predictive of listeners’ discrimination. Ramus, Dupoux and Mehler (2003) utilised resynthesised speech, comprising sequences of monotone *sasasa* syllables with the durational values of the original vowels and consonants. They showed that listeners can discriminate *sasasa* speech from languages with distinct rhythm scores (e.g., English and Spanish), but not languages with similar rhythm scores (e.g., English and Dutch). By comparing *sasasa* utterances from different varieties of the same language, White, Mattys, Series and Gage (2007) controlled for the possibility of perceptual distinctions being cued by cross-linguistic differences in the syntagmatic arrangements of stressed and unstressed syllables – a property not explicitly captured by rhythm metrics – and found that listeners can discriminate Welsh Valleys and standard southern British English. The test was more stringent than that provided
by Ramus et al.: speech rate was normalised and all utterances were trimmed to the same number of syllables, thus removing utterance-edge durational information, such as might be provided by differential patterns of final lengthening. Discrimination performance was poorer than in Ramus et al.’s study, but still reflected differences in rhythm scores.

Contrasts in the duration of stressed and unstressed vowels, as measured by VarcoV, and in the relative distribution and duration of vowels and consonants, as measured by %V, appear therefore to be significant factors in the perception of rhythmic contrasts. However, the fact that the minimal task utilised by White et al. (2007) was more difficult than the similar discrimination experiments of Ramus et al. (1999), indicates the importance of additional temporal information that is not reflected in the durational contrasts of stressed and unstressed vowels and consonants. Furthermore, there are many sources of non-durational information that may contribute to listeners’ perception of rhythm, including intonation and intensity variation, as well as segmental cues to stress.

1.2 Temporal and non-temporal correlates of rhythm

Spanish and English, at least among languages so far studied with rhythm metrics, seem the most opposed in terms of the durational marking of lexical stress. As well as a high durational contrast between stressed and unstressed syllables, English has substantial lengthening of pitch-accented (i.e. phrasally-stressed) syllables, which extends to other syllables within the pitch-accented word (e.g. Turk & White 1999). In addition, onset consonants are longer at the start of words than word-medially (e.g., Oller 1973) and vowels and coda consonants are longer at the end of phrases than phrase-medially (e.g. Wightman, Shattuck-Hufnagel, Ostendorf & Price 1992). Outside of the temporal domain, English reduced vowels, which occur in almost all unstressed syllables, are greatly centralised, in addition to being much shorter than stressed vowels.

Lexical stress in Spanish is also associated with a degree of lengthening, particularly in word-final syllables, but the size of the durational contrast between stressed and unstressed syllables is small (e.g., Delattre 1965). In a recent study, for example, Ortega-Llebaria and Prieto (2007) report 7 ms stress-related lengthening for penultimate syllables and 15 ms for final syllables, much less than the durational difference between full and reduced syllables in English. Despite the small degree of temporal stress contrast, it should be noted that Ortega-Llebaria and Prieto (2009, this volume) find that Spanish speakers use duration as a cue to stress, contrasting with English speakers for whom vowel quality is a stronger stress cue. With regard to vowel quality in Spanish, Ortega-Llebaria and Prieto found a small effect of stress on the realisation of word-final vowel [o], which was slightly more
centralised in unstressed syllables, but no effect on penultimate [i], which lacked any contrast in its realisation between stressed and unstressed syllables. Higher-level durational effects are attenuated in Spanish: in the same study, Ortega-Llebaria and Prieto found word-final syllables lengthened very slightly (by 6 ms on average) when pitch-accented, and penultimate syllables showed no significant lengthening at all. Furthermore, Frota, D’Imperio, Elordieta, Prieto and Vigário (2007) found that only 40% of intonational-phrase boundaries were marked by final lengthening in Castilian Spanish, compared with, for example, 100% of boundaries for Catalan.

As Frota et al. acknowledged, controlled studies are required for a true comparison both of the frequency and magnitude of boundary-related lengthening in Romance languages. However, it may be noted that Catalan has significant vowel reduction in unstressed syllables, unlike Spanish, and has been considered a less “syllable-timed” language, a claim empirically supported by the single-speaker data of Grabe and Low (2002). Frota et al. also found that 100% of IP-boundaries were marked in the Italian of Neapolitan speakers, which, as discussed below, is one of the southern varieties of Italian that have been held to be “stress-timed”. Thus, greater evidence for final lengthening was found in languages (Catalan, Neapolitan Italian) held to be less “syllable-timed” than Spanish.

Results for rhythm metrics such as nPVI-V and VarcoV (e.g. Grabe & Low 2002; White & Mattys 2007a, 2007b) clearly point to gradient distinctions in rhythm between and within languages. In addition, differences between languages suggested by rhythmic typology, and supported by rhythm scores, appear to correlate – particularly at the English and Spanish extremes – with differences in the overall use of duration as a cue to structure: languages like English utilise a high durational contrast in delimiting lexical and higher-level prosodic structure, whereas languages like Spanish utilise low durational contrast for the same purpose.

As discussed above, the dichotomous typology of “stress-timed” vs. “syllable-timed” and the isochrony this typology implies are empirically unsupported. It may be more useful to think of languages differing in the degree to which they utilise durational and segmental contrast to indicate both stress and higher-level prosodic structure.

1.3 The phonetic basis of Italian rhythm

Here we examine the contribution of durational and segmental factors to the perceived rhythmic difference between northern and southern Italian. We look for evidence of variation in the exploitation of contrast, both durational contrast and contrast in vowel quality. Italian, although another Romance language generally held to be “syllable-timed”, differs from Spanish in having vowels that vary significantly in duration as a result of lexical stress and other phonological factors.
Italian vowels are longer in stressed open syllables than in closed syllables or in unstressed open syllables, and this is especially the case when this stressed open syllable is in word-penultimate position (D’Imperio & Rosenthall 1999). Indeed, Bertinetto (1980) claims that vowel duration is the most important parameter for the perception of stress in Italian. As suggested by Russo and Barry (2004), this phonetic lengthening of stressed vowels is likely to influence vocalic variability scores (e.g. VarcoV) away from the extreme low of Spanish. Likewise, the existence of geminate consonants in Italian leads one to expect a relatively high degree of variation in consonantal interval durations, and a lower %V score than would otherwise be predicted for a Romance language.

Despite the durational cues to stress, standard Italian (e.g. that of Tuscany) has generally been described as “syllable-timed”, as have varieties of northern Italian. In contrast, several sources have suggested that southern Italian tends towards “stress-timing” (e.g. Grice, D’Imperio, Savino & Avesani 2004; Russo & Barry 2004 also cite Romito & Trumper 1993; and Trumper, Romito & Maddalon 1991), although objective perceptual tests of such rhythmic differences appear to be lacking.

Various phonetic processes in southern Italian serve to increase contrast between stressed and unstressed syllables. Russo and Barry (2004) report gradient variation in the degree of reduction of word-final unstressed vowels, which ranges from shortening and centralisation to devoicing to deletion. The elision of voiced vowels also serves to increase the complexity of consonant clusters, creating heavy or superheavy syllables not licensed in standard Italian phonology (Russo & Barry 2004). A similar process is observed in European Portuguese, as contrasted with Brazilian Portuguese where epenthetic vowels may actually interrupt and thereby simplify consonant clusters (Frota & Vigário 2001). Consonant interval durations are made more variable by the fact that lexical geminates are longer in southern Italian, and by *raddoppiamento fonosintattico*, a post-lexical gemination process not present in northern varieties of Italian, whereby word-initial consonants are significantly lengthened in certain circumstances (e.g. after certain words, or less consistently, when following a final stressed vowel – cf. Lopocaro 1997; Payne 2005).

1.4 Experimental aims

We are not aware of any systematic attempt to quantify rhythmic differences between northern and southern Italian using rhythm metrics such as VarcoV and %V, beyond a preliminary study (Barry, Andreeva, Russo, Dimitrova, & Kostadinova 2003) on pre-labelled corpora of spontaneous speech. A range of rhythm metrics were applied to compare Italian speech from Bari and Naples (southern) and Pisa (central, close to standard Italian), together with varieties of German and Bulgarian. Their results were inconclusive, however, partly as a
result of speech rate differences between the corpora and also perhaps due to labelling discrepancies between corpora.

Here we report a three-part investigation into rhythmic and prosodic differences between northern and southern varieties of Italian. Our working hypothesis was that southern Italian speakers should show higher indices of contrast on all of the following measures.

**Rhythm scores.** We utilised rhythm metrics to quantify the degree of variation in vocalic interval duration (VarcoV) and the relative balance of vocalic and intervocalic intervals (%V). We chose these metrics as they have been shown to be the most discriminative between languages (White & Mattys 2007a). These metrics have also distinguished accents of English held to differ rhythmically (White & Mattys 2007b). Importantly, they are robust to variation in speech rate (White & Mattys 2007a), at least for languages so far studied.

In contrast, metrics of consonantal interval duration have been shown to be problematic. Scores for non-rate-normalised metrics (ΔC: standard deviation of consonantal interval duration; rPVI-C: mean durational difference between successive pairs of consonantal intervals) have clear inverse correlations with speech rate. Rate normalisation, for example, in VarcoC (standard deviation of consonantal interval duration divided by the mean) appears to remove much of the linguistically-relevant differences between samples (White & Mattys 2007a), as had been previously suggested by Grabe and Low (2002). Despite this caveat, we thought it necessary to include a rate-insensitive metric of consonantal interval variation – VarcoC – given the differences in distribution and magnitude of geminate consonants between northern and southern Italian.

**Prosodic timing analysis.** We compared the degree of structurally-determined lengthening in northern and southern Italian. Specifically, we measured stressed vowel duration in utterance-medial words carrying a prenuclear pitch accent and utterance-final words carrying a nuclear pitch accent.

**Vowel reduction analysis.** We investigated the degree of centralisation of stressed and unstressed vowels in the two varieties. We measured the formant structure of stressed, pre-stress and post-stress vowels for two different levels of prominence in the phrase: prenuclear utterance-medial and nuclear utterance-final.

### 2. Method

#### 2.1 Participants

We recorded six native speakers of Italian from the Veneto, in the north east of Italy, and six from Sicily, in the far south. All speakers had been brought up in the
regions in which they were recorded, and none reported any speech or hearing impediments. The Veneto speakers were recorded by the first author at the University of Padua and the Sicilian speakers were recorded by the second author at the University of Palermo and the University of Catania. All speakers were paid a small honorarium for their participation.

2.2 Materials

**Rhythm sentences.** To obtain rhythm scores, we recorded each speaker reading the same set of five sentences. Following the methodology used in White and Mattys (2007a, 2007b), we designed these sentences to exclude approximants as far as possible. This was to facilitate the segmentation of the recorded utterances into vocalic and consonantal intervals. The sentences were:

\begin{quote}
Davide insegna matematica come tutti i suoi amici.
“David teaches maths like all of his friends.”

Quando c’è così tanta gente in città vado in montagna.
“When there are so many people in the city I go to the mountains.”

Giovedì ho visto Donata Zanzetti e Amanda Baggio in tivù.
“One Thursday I saw Donata Zanzetti and Amanda Baggio on the telly.”

Hanno mangiato una zuppa di cozze e un piatto di pesce fritto.
“They ate mussel soup and a plate of fried fish.”

Se non dici dove sono i cappotti, non possiamo uscire.
“If you don’t say where the coats are, we can’t go out.”
\end{quote}

**Prosodic-timing sentences.** To obtain measurements of the magnitude of prosodic timing effects, we contrasted a series of target words in prenuclear utterance-medial position and nuclear utterance-final position. The target words (with position of lexical stress illustrated) were: ‘bada, cand’i tissimi, ‘fata, fa’ tale, ‘fatta, ‘fichi, ‘mitti. Because there was the possibility of sentences being realised as two or more intonational phrases, we used capital letters to indicate the desired placement of phrasal stress. We hoped thereby to forestall the possibility of a major prosodic boundary being realised after the target word in the prenuclear utterance-medial context. The target words carried prenuclear accent in this context, but, as described below, we successfully avoided the realisation of a phrase-final nuclear accent on the target words in these tokens. These are the sentences, with target words shown here in bold (they were not presented in bold for participants):

\begin{quote}
Non bada TANTO al Signor GADDA.
“He’s not so concerned about Mr Gadda.”

Ai suoi canditissimi SORRISI non ci BADA.
“He takes no notice of her very clear smiles.”
\end{quote}
As can be seen, because of the way the materials were constructed, several sentences contained both an utterance-medial prenuclear target word and a (different) utterance-final nuclear target word.

**Vowel-reduction sentences.** To obtain estimates of vowel reduction effects, we recorded eight sentences which had been designed to contain the vowels /i/ and /a/ in various degrees of lexical and phrasal prominence. The word-level stress contrasts are illustrated in Table 1.

<table>
<thead>
<tr>
<th>Pre-stress</th>
<th>Stressed</th>
<th>Post-stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>f I ch i s i m i</td>
<td>f I ch i</td>
</tr>
<tr>
<td>/a/</td>
<td>b A d a r e / b A d a t e</td>
<td>b A d a</td>
</tr>
</tbody>
</table>

The phrase-level prominence conditions were: (i) prenuclear accent, utterance-medial position; (ii) nuclear accent, utterance-final position. These are the sentences, with target words shown here in bold (once again, they were not presented in bold for participants):

*Non badate TANTO a diventare RICCHI.*
“Don’t be too concerned about becoming rich.”

*Per diventare ricchi DAVVERO non ci dovete BADARE*
“To get really rich, you shouldn’t care so much.”
Non bada TANTO al Signore GADDA.
“He’s not so concerned about Mr Gadda.”
Ai suoi candidissimi SORRISI non ci BADA.
“He takes no notice of her very clear smiles.”
Solo i fichissimi ragazzi della CITTA’hanno capelli FITTI.
“Only the really cool city kids have thick hair.”
Quel badalucco INGLESE non aprezza questini arancini FICHISSIMI.
“That English waste of space doesn’t appreciate these really cool ‘arancini’.”
Questi fichi D’INDIA sono veramente dei MITI.
“These prickly pears are really legendary.”
È fatale DAVVERO, mangiare troppe FICHI.
“Eating too many figs is really fatal.”

Some of these sentences and target words were also used in the prosodic timing analyses. As discussed above, block capitals were used to indicate the desired placement of the main phrasal stresses.

2.3 Recordings

All speakers read all sentences. The rhythm sentences were presented on a single sheet of paper, preceded by another five sentences which were not analysed for this experiment. The prosodic-timing and vowel-reduction sentences were printed together in random order on a separate sheet of paper, plus two other sentences.

The prosodic-timing and vowel-reduction sentences were read first. These were followed by a task in which participants described a route around a fictional map. The participants then read the rhythm sentences plus fillers. Finally, the prosodic-timing and vowel-reduction sentences were read again. The second reading of the sentences was not used in the analyses reported here.

Prior to reading the sentences, participants were instructed to take as much time as they wished to familiarise themselves with the sentences, and then to read the sentences out loud, at a normal rate and in a conversational style. They were told to avoid pausing within sentences, but to make a brief pause between sentences and to repeat any sentence in which they made an error. For the prosodic-timing and vowel-reduction sentences, participants were directed to the words in capital letters as being the most important words in the sentences. This was intended to prevent utterance-medial targets carrying nuclear accents or being followed by phrase-boundaries.

The experimenter monitored the productions and asked speakers to repeat sentences when they were misread. The experimenter gave no other instructions regarding the reading of the sentences.
All recordings took place in a quiet rooms, using good quality microphones, and were made directly to disk at sample rates of 32 kHz or higher.

2.4 Measurements

*Rhythm analysis.* We followed the methodology of White and Mattys (2007a, 2007b), dividing each utterance into vocalic and intervocalic intervals based on inspection of the waveform and wideband spectrogram using Praat (http://www.fon.hum.uva.nl/praat/). Full details of the segmentation criteria used are given in White and Mattys (2007a).

The vowel and consonant interval durations for each utterance of each speaker were measured. We used these durational data to calculate the rhythm metrics VarcoV, %V and VarcoC. VarcoV is the standard deviation of vocalic interval duration divided by the mean (and multiplied by 100). VarcoC is the standard deviation of consonantal interval duration divided by the mean (and multiplied by 100). %V is the proportion of utterance duration comprised of vocalic intervals. We also calculated the overall speech rate for each utterance, as the number of syllables divided by the utterance duration.

*Prosodic-timing analysis.* Using Praat, we measured the duration of the vowel in the stressed syllable of each target word in the two prosodic contexts through inspection of the waveform and wideband spectrogram.

*Vowel-reduction analysis.* Using Praat, the speech recordings were segmented and values for F1 and F2 were measured using the formant-tracking function of the software and checked by hand against the spectrogram for accuracy. Since the identity of preceding and following consonants influences formant transitions into and out of the vowel in question, formant values were taken only at the middle of the vowel, where consonantal effects were expected to be less strong.

3. Results

3.1 Results: Rhythm scores

Table 2 shows the mean rhythm scores for VarcoV, VarcoC and %V, together with the mean speech rates. Results of By-Subjects comparisons for each measure are also shown. None of the expected differences in rhythm scores are evident in Table 2. If Sicilian Italian were more “stress-timed” than Venetan Italian, it should have higher VarcoV and VarcoC scores and lower %V. The only reliable difference was that the speech of Sicilian Italian was significantly faster than that of Venetan Italian.
As shown in Figure 1, reliable differences have been previously been found between and within languages using, as in the current study, six speakers for each group reading five sentences each. In particular, using this methodology, Welsh Valleys English was shown to have significantly higher %V and significantly lower VarcoV than standard southern British English (White & Mattys 2007b). Thus, the lack of significant differences here is unlikely to be an issue of statistical power.

Pairwise comparisons indicate that both Venetan Italian and Sicilian Italian had higher VarcoV scores than Castilian Spanish [Venetan vs. Castilian: t(10) = 4.48, p = .001; Sicilian vs. Castilian: t(10) = 2.37, p < .05]. VarcoV scores for both Venetan and Sicilian Italian were not significantly different from those for French, and clearly much lower than those for standard English and Dutch. The %V scores for Venetan and Sicilian Italian were not significantly different from those for Spanish or French, but clearly higher than those for Dutch and English. Thus, VarcoV is the more discriminative metric here, suggesting a gradient of rhythm, with lesser temporal stress contrast in Spanish than French or either variety of Italian, but with no temporal rhythmic distinction between the Italian varieties.

VarcoC scores, while numerically supportive of greater contrast in Sicilian Italian (48.2 compared with 45.7 for Venetan Italian), were not significantly different. This is somewhat surprising given the greater incidence of post-lexical gemination in southern Italian, and the greater magnitude of geminates in general. It may be that different materials eliciting more lexical and post-lexical geminates would produce reliable differences in scores, although VarcoC has not previously been shown to manifest predicted discrimination patterns (e.g., White & Mattys 2007a).

Thus, rhythm scores did not support any differentiation between Venetan and Sicilian in terms of their degree of temporal stress contrast, with both varieties appearing rhythmically similar to French. What then underpins the perception of so-called “stress-timing” in southern Italian varieties, such as Sicilian? In the following sections, we consider localised rather than global measures of temporal contrast, as well as non-temporal contrasts in vowel realisation.
Results: Prosodic timing

As described above, stressed-vowel duration was measured for seven words in prenuclear utterance-medial and nuclear utterance-final vowels. Table 3 shows mean stressed vowel duration for these target words in Venetan and Sicilian Italian. A By-Subjects repeated measures ANOVA showed a main effect of Position (prenuclear utterance-medial, nuclear utterance-final) \( [F(1,10) = 171.97, p < .001] \), but no main effect of Accent (Venetan, Sicilian) \( [F(1,10) = 0.12, p > .10] \). There was a significant interaction between Position and Accent \( [F(1,10) = 23.17, p = .001] \).

Table 3. Mean stressed vowel duration in ms (standard errors in parentheses) according to prosodic position and accent of Italian

<table>
<thead>
<tr>
<th></th>
<th>Prenuclear utterance-medial</th>
<th>Nuclear utterance-final</th>
</tr>
</thead>
<tbody>
<tr>
<td>Venetan Italian</td>
<td>95 (7)</td>
<td>130 (7)</td>
</tr>
<tr>
<td>Sicilian Italian</td>
<td>72 (5)</td>
<td>146 (10)</td>
</tr>
</tbody>
</table>

The nature of the interaction is evident from inspection of the means in Table 3. In utterance-medial position, stressed vowels were 33 ms longer in Venetan than in Sicilian Italian, reflecting the finding reported above that the speech rate of the Sicilian speakers was faster than that of the Venetan speakers. In utterance-final
position, however, the position was reversed: vowels in utterance-final syllables were 16 ms longer in Sicilian than in Venetan Italian. Both varieties showed vowel lengthening in nuclear utterance-final words, but the magnitude of the effect was much greater in Sicilian (75 ms vs. 35 ms in Venetan).

![Graph](image)

Figure 2. Mean durational difference (ms) between stressed vowels in prenuclear utterance-medial position and nuclear utterance-final position

Figure 2 illustrates the nuclear final-lengthening effect on a word-by-word basis. As can be seen, for all words, the durational difference in stressed-syllable vowel duration between medial and final position was greater for Sicilian than Venetan speakers. A By-Subjects two-tailed t-test on the mean differences confirmed the reliability of this pattern \(t(10) = 4.81, p = .001\).

The difference in the proportional nuclear accent/final lengthening effect was particularly marked. As reported above, the rate of speech was greater in Sicilian and so the utterance-medial vowels were much shorter. As a proportion of utterance-medial duration, the nuclear accent/final lengthening of Sicilian stressed vowels was 104%, compared with 38% for Venetan stressed vowels.

Thus, although the results of the rhythm metrics analysis did not support the hypothesised distinction between northern and southern varieties of Italian, the prosodic-timing analysis does support the more general contrast hypothesis. There is much greater prosodic lengthening – in nuclear-accented utterance-final position – in Sicilian than in Venetan Italian. This echoes the finding, discussed above, of much greater prosodic lengthening effects in high stress-contrast English than in low stress-contrast Spanish. (Direct numerical comparison between these
studies is not possible, due to differences in experimental design.) Unlike English and Spanish, the difference in prosodic lengthening between Venetan and Sicilian was not paralleled by differences in the rhythm scores.

3.3 Results: Vowel reduction

Both Venetan and Sicilian Italian showed some evidence of phonetic vowel reduction in unstressed vowels when compared with stressed vowels. Figure 3 shows mean values for F1 and F2 in /i/ and /a/ according to stress condition and prosodic position for Venetan Italian; Figure 4 shows the same data for Sicilian Italian. We assessed the effect of three independent variables – Stress (stressed, pre-stress, post-stress), Position (prenuclear utterance-medial, nuclear utterance-final) and Accent (Venetan, Sicilian) – separately for F1 and F2, using By-Subjects repeated measures ANOVAs. Post-hoc Tukey HSD pairwise comparisons were calculated separately for each accent group, as differences in formant frequencies between accent groups are not theoretically interesting for this study.

**Vowel /i/**. Overall, the trend was for a higher F1 and lower F2 in pre-stress and post-stress /i/ compared with stressed /i/, suggesting a more centralised production for unstressed /i/ (in Figures 3 and 4, grey triangles and squares are closer to the intersection than grey circles). With regard to the effect of prosodic position, vowels in nuclear-accented words, other things being equal, have more extreme articulation than those in prenuclear, phrase-medial words (grey filled symbols are further from the intersection than are grey open symbols).

For F1, there was a main effect of Stress \[F(2,20) = 19.39, p < .001\], a main effect of Position \[F(1,10) = 32.97, p < .001\], and a main effect of Accent \[F(1,10) = 24.17, p = .001\]. There were no significant interactions, but pairwise comparisons revealed that the only reliable difference in F1 frequency was in nuclear position for Sicilian, where F1 was significantly lower in stressed syllables than in post-stress syllables \(p < .05\). In the same position, the difference in F1 between Sicilian stressed and pre-stress syllables almost attained significance at the \(p < .05\) level. Thus, for Sicilian but not for Venetan, stressed /i/s in nuclear position tend to have more extreme articulation than pre-stress or post-stress vowels. This is in line with the contrast hypothesis: the southern variety shows greater articulatory distinction according to prosodic position.

For F2, there were main effects of Stress \[F(2,20) = 140.95, p < .001\], Position \[F(1,10) = 143.46, p < .001\] and Accent \[F(1,10) = 24.04, p = .001\]. There were significant interactions between Stress and Accent \[F(2,20) = 11.36, p = .001\] and between Position and Stress \[F(2,20) = 3.65, p < .05\]. Pairwise comparisons showed that, for Venetan, there was a distinction between prenuclear and nuclear stressed vowels, with nuclear stressed vowels having higher F2 \(p < .05\), suggestive
of more extreme articulation. Sicilian showed a similar nuclear vs. prenuclear distinction for stressed vowels (p < .01) and a distinction in pre-stress vowels as well, with higher F2 values in nuclear than prenuclear position (p < .01).

Within prenuclear vowels, Venetan distinguished stressed vowels and post-stress vowels (p < .05), with the difference between stressed and pre-stress vowels almost attaining significance (p < .10); Sicilian distinguished stressed vowels from both pre-stress and post-stress vowels (both p < .01). Within nuclear vowels, Venetan distinguished stressed vowels from pre-stress and post-stress vowels (both p < .01); Sicilian likewise distinguished these categories (both p < .01). In all cases, stressed /i/ vowels had higher F2, indicative of more extreme articulation. Additionally, in Sicilian, pre-stress vowels had higher F2 than post-stress vowels (p < .01).

![Figure 3. Mean vowel positions in F1/F2 plane for Venetan Italian. Filled symbols: nuclear utterance-final; open symbols: pre-nuclear utterance-medial. Circles: stressed syllables; triangles: pre-stress syllables; squares: post-stress syllables. Hairline intersection = vowel plane centre](image)

The F2 results for /i/ suggest more differentiation according to stress condition and utterance position, for both Venetan and Sicilian, than do the F1 results. However, there are larger absolute F2 differences for Sicilian and reliable differentiation across more categories, as suggested by the interaction between Stress and Accent (Sicilian vs. Venetan).

**Vowel /a/**. Overall, the trend was for a lower F1 and higher F2 in pre-stress and post-stress /a/ compared with stressed /a/, suggesting a more centralised production for unstressed /a/ (in Figures 3 and 4, black circles are further from the intersection than are black triangles and squares). There was also a general trend for post-stress vowels to be more centralised than pre-stress vowels (black triangles are further from the intersection than black squares). With regard to the effect of prosodic position, vowels in nuclear utterance-final words, other things being equal, have more extreme articulation than those in prenuclear
utterance-medial words (black filled symbols are further from the intersection than black open symbols).

For F1, there was a main effect of Stress \[F(2,20) = 949.10, p < .001\], a main effect of Position \[F(1,10) = 586.28, p < .001\] and a main effect of Accent \[F(1,10) = 721.33, p < .001\]. There was also a significant interaction between Position and Accent \[F(1,10) = 23.93, p = .001\]. The effect of Position is evident in the post-hoc comparisons between prenuclear and nuclear position: for both Venetan and Sicilian, /a/ vowels in nuclear-accented words have higher F1, indicative of more extreme articulation, than equivalent vowels in prenuclear words, whether stressed (Ven: p < .05; Sic: p < .01), pre-stress (Ven: p < .05; Sic: p < .01) or post-stress (Ven: p < .01; Sic: p < .01).

![Figure 4. Mean vowel positions in F1/F2 plane for Sicilian Italian. Filled symbols: nuclear utterance-final; open symbols: pre-nuclear utterance-medial. Circles: stressed syllables; triangles: pre-stress syllables; squares: post-stress syllables. Hairline intersection = vowel plane centre](image)

There was an effect of Stress within nuclear words for both Venetan and Sicilian, with stressed vowels having higher F1 than pre-stress or post-stress vowels (p < .01 for all comparisons). The same stress distinctions – stressed vowels having higher F1 than pre-stress or post-stress vowels – were also found in prenuclear words (p < .01 for all comparisons). The only point of between-accent differences for F1 in the /a/ vowel was that Sicilian speakers also distinguished pre-stress and post-stress vowels in prenuclear position, with the former having higher F1 (p < .01).

Thus, results for /a/ F1 do not greatly distinguish Venetan and Sicilian, with both accents marking stressed vowels with more extreme articulation, and all types of vowels in nuclear position having more extreme articulation than equivalent prenuclear vowels.

For F2, as with the other formant measures, there were main effects of Stress \[F(2,20) = 196.26, p < .001\], Position \[F(1,10) = 125.11, p < .001\] and Accent \[F(1,10) = 354.29\]. There were also interactions between Position and Accent
[F(1,10) = 67.83, p < .001] and Stress and Accent [F(2,20) = 45.49, p < .001], and a three-way interaction between Position, Stress and Accent [F(2,20) = 4.40, p < .05]. For /a/ F2, the only distinction made by Venetan speakers was in nuclear position, where stressed vowels had lower F2 than either pre-stress or post-stress vowels (both p < .01), suggesting the latter were more centralised. Sicilian speakers also showed these distinctions in nuclear position (both p < .01), where, in addition, pre-stress vowels had lower F2 than post-stress vowels (p < .01).

The full range of distinctions according to stress position was also made in prenuclear words by Sicilian speakers: stressed vowels had lower F2 than both pre-stress and post-stress vowels (both p < .01); pre-stress vowels had lower F2 than post-stress vowels (p < .01). Finally, nuclear position was marked by lower F2 than prenuclear position for all stress conditions – stressed, pre-stress, post-stress – by Sicilian speakers (all p < .01).

![Figure 5. Mean F2 values in Hz for /a/ according to stress and prosodic position](image)

The results for F2 for the /a/ vowel provide the strongest support for the contrast hypothesis: stress was marked with more extreme articulation in both nuclear and prenuclear positions by Sicilian speakers, who also marked nuclear accent with more extreme articulations for both stressed and unstressed vowels. In contrast, Venetans only marked stress by more extreme articulation in nuclear position and did not further distinguish vowels in nuclear and prenuclear words. These patterns are illustrated in Figure 5.

Overall, these results strongly suggest that Sicilian phonetically distinguishes different parts of the intonational phrase to a greater extent than does Venetan. The greater centralisation of vowels in prenuclear position correlates with our
findings on greater phrase-final lengthening in Sicilian, a relationship that may have a biomechanical interpretation. There is strong evidence for a linear relationship between duration and formant displacement, with shorter vowels tending to undergo more formant displacement – in the direction of centralization – than longer vowels (Lindblom 1963; Moon & Lindblom 1994). With less time available for the realization of the vocalic gesture, hypo-articulation occurs, resulting in target undershoot. Ortega-Llebaria and Prieto (2007:173) find a similar correlation of vowel duration and centralization in Spanish, although the distinctions according to lexical stress and prosodic position are relatively small compared with those found here for Italian, in particular for Sicilian.

4. Discussion

Southern Italian (e.g., Sicilian) has been frequently described as more “stress-timed” than northern Italian (e.g., Venetan). Evidence from previous studies suggests, however, that differences in language rhythm may be a matter of degree rather than categorical. In addition, there is evidence, particularly from the comparison between English and Spanish, that high durational contrasts between stressed and unstressed syllables co-occur with large contrasts in prosodically-governed vowel centralisation and with the existence of large prosodic lengthening effects. We therefore looked for evidence of rhythmic, prosodic and segmental contrasts between Sicilian and Venetan Italian.

Perhaps surprisingly, we found no evidence from VarcoV and %V scores of Sicilian Italian being rhythmically distinct from Venetan Italian, even though the same measures (and sample size) highlighted clear differences between and within languages in previous studies. Furthermore, the fact that Venetan Italian had rhythm scores that are close to French, and with somewhat higher temporal stress contrast than Spanish, seems to accord with expectations based on the phonetic characteristics of stress production in these languages. Thus, the result for Sicilian Italian, which was located in the same VarcoV vs. %V rhythm space as Venetan Italian, may be regarded as reliable.

It is possible that there may be greater differences in rhythm scores between other varieties of Italian than those we have chosen. For example, discussing “stress-timing” in southern Italian, Russo and Barry (2004) presented evidence from the Italian of Bari and Naples. Most natives of these regions speak both Italian and the local Baresse or Neapolitan dialects, and these dialects – which greatly influence the form of spoken Italian in these regions – form part of the “Upper-Meridional” group of dialects in Italy, which are distinct from the dialects of the far south, including Sicilian. Further investigation of a broader spectrum of
standard Italian varieties and, for comparison, dialects, is needed to provide a clearer picture of rhythm variability in Italian.

Despite the lack of evidence for high temporal stress contrast between Sicilian and Venetan Italian, we did find a marked amplification in prosodic lengthening for Sicilian Italian. Stressed vowels in nuclear utterance-final words were more than twice as long as the same vowels in prenuclear utterance-medial words. For the same prosodic contrast, stressed vowels in Venetan Italian were lengthened by just over a third of their utterance-medial duration. It seems likely that this difference in prosodic timing would confer a more contrastive quality to Sicilian, although perceptual tests, for example using the *sasasa* resynthesis technique, would be required to determine whether such differences alone allow listeners to discriminate the two varieties. Of course, it would be expected that prosodic contrast in vowel duration would be reflected in a measure of such variation, with the greater nuclear accent/final lengthening in Sicilian contributing to a higher VarcoV score than for Venetan. However, nuclear accent/final lengthening only applies once per utterance, and thus may not have a major impact on the overall standard deviation of vocalic interval duration.

Furthermore, it may be noted that the precise nature of the lengthening effect observed here is uncertain. Because we did not want to constrain speakers’ productions excessively – in the interest of generating naturalistic speech – two possible sources of lengthening were confounded: the contrast between prenuclear-accented and nuclear-accented syllables, and the contrast between utterance-medial and utterance-final syllables. This may have some relevance to rhythmic typology. Prosodic timing effects can clearly be divided into domain-head effects (e.g., stress-related lengthening, accentual lengthening) and domain-edge effects (e.g., word-initial lengthening, phrase-final lengthening). Extrapolating from the case of lexical stress, Beckman (1992) speculates that head effects tend to be specific to “stress-timed” (i.e. high temporal stress contrast) languages, whereas edge effects may be more or less ubiquitous. This contention makes distinct predictions from our contrast hypothesis, which predicts that high temporal stress contrast languages should show more marked prosodic lengthening effects across the board (i.e. both domain-head and domain-edge effects). An obvious direction for future research would be to test the predictions of the contrast hypothesis and the heads/edges hypothesis by comparing the magnitude of domain-head and domain-edge effects in rhythmically-distinct languages or varieties.

Data from Frota and Vigário (2001) suggest that the hypothesised lower contrast (“syllable-timed”) Brazilian Portuguese actually has more widespread final lengthening than higher contrast (“stress-timed”) European Portuguese, a result apparently opposed to the contrast hypothesis. However, although consonant interval duration is indeed more variable in European Portuguese, scores for vowel
duration variation are actually higher for Brazilian Portuguese, indicating not only that the factors contributing to rhythmic perception may be partially independent, but also that the magnitude of prosodic timing effects may be calibrated in relation to vowel rather than consonant duration patterns.

Our contrast hypothesis is quite well supported by the results for vowel reduction. We looked for evidence of more centralised articulations for unstressed vowels compared with stressed vowels and for vowels in prenuclear utterance-medial words compared with nuclear utterance-final words. We found evidence for both – centralisation in unstressed vowels and centralisation in prenuclear position – from F1 or F2 values for both /i/ and /a/, but the pattern was not wholly consistent. In all cases where we found reliable differences, however, there was greater and more reliable contrast in formant frequencies in Sicilian Italian than in Venetan Italian. In Sicilian Italian, there was also evidence for an articulatory contrast between pre-stress and post-stress syllables – greater centralisation in the latter – which was absent in Venetan.

English-speaking linguists have inferred from the relative prominence of stressed syllables in Germanic languages that stress is more perceptually salient than in Romance languages. However, as Arvaniti (1994) points out, stressed syllables are in fact highly salient to native speakers of languages like Spanish, Italian or Greek, who are sensitive to misaccentuations in minimal pairs such as the Spanish ‘como (I eat) vs. co’mo (she ate). The hypothesis that rhythmic, prosodic and segmental contrasts are all attenuated in such languages fits with this observation: native speakers may have become attuned during language acquisition to expect relatively small durational differences to be linguistically meaningful, whereas speakers of high stress-contrast languages such as English or Dutch cannot perceive such fine durational distinctions in speech. This perceptual hypothesis awaits further testing.

5. Conclusions

Our results suggest that not all perceptually-salient rhythmic differences are captured by duration-based rhythm metrics, and provide a further challenge to the notion that rhythmic variation is based on categorical differences in relative durations. Furthermore, they suggest that the perception of rhythmic variation may be determined by a combination of factors including syllable structure, segmental and prosodic timing, and the relationship between prosodic structure and vowel and consonant fortition/lenition. Syntagmatic differences in stress distribution between languages, not examined here, may also be a covariate. The cumulative effect of these factors produces rhythmic templates which are variety-specific but may cluster, at least perceptually, around rhythmic types.
References


Stem boundary and stress effects on syllabification in Spanish*

Teresa Cabré & Maria Ohannesian
Universitat Autònoma de Barcelona

Spanish stress shows a uniform pattern in verbal forms. For stressed roots in particular, the unmarked position is the last vowel. Nevertheless, first conjugation roots ending in a high vowel display two opposite behaviors: while one group follows the unmarked pattern (e.g., envío ‘I send’), the other group keeps the high vowel unstressed, thus always becoming a glide (e.g., cambio ‘I change’). While nominals related to the latter group exhibit the same stress position as the first conjugation (e.g., cambio [kámbjo] ‘change’, cambio [kámbjo] ‘I change’), this is not true of nominals related to the former group, since we find some cases with stress shift (e.g., amplio [âmpljo] ‘large’, amplio [amplio] ‘I enlarge’) and other cases without this change (e.g., envío [embio] ‘shipment’, envío [embio] ‘I send’).

The goal of this paper is to account for these facts (including the lack of pairs such as *amplio [amplio] ‘large’ and *amplio [âmpljo] ‘I enlarge’) and analyze the prosodic and morphological factors determining glide formation.

1. Introduction

Glide formation in vowel sequences of rising sonority has been dealt with widely in the literature on Spanish (Navarro Tomás 1948; Harris 1983; Harris & Kaisse 1999; Colina 1999; Hualde 1999, 2005; Cabré & Prieto 2007). Nevertheless, none of these studies has focused on a comparison of first conjugation verbs whose

*Partial and previous versions of this work were presented at Toulouse – Le Mirail at the 4th International Colloquium on Morphology “Décembrinettes” (2005), at the University of Barcelona at the 7th Congress of General Linguistics (2006) and at PaPI’07 Phonetics and Phonology in Iberia. A related paper was published in Cuadernos de Lingüística XIV (2007). We are grateful to the audience at those events, and especially, to James W. Harris, Eulàlia Bonet, José Ignacio Hualde, Michael Kenstowicz, Joan Mascaró, Pilar Prieto and Iggy Roca for helpful comments on certain parts of the work. We also thank the reviewers whose useful comments helped improve the paper. All errors are ours. This research was funded by grants 2005SGR-00753 from the Generalitat de Catalunya and HUM2006-13295-C02-01/FILO (EJE C/B) and HUM2006-01758/FILO from the Spanish Ministry of Science and Education.
stems end in a high vowel with the lexically related nominal forms. The aim of this study is to account for the syllabic behaviour of rising sonority sequences in this set of lexically related items.

Harris and Kaisse (1999) pointed out that verbs and related nominals may or may not display the root stress in the same position. If they do, the high vowel may be either unstressed (cámbio ‘change’, cámbian ‘they change’) or stressed (lío ‘confusion’, lian ‘they confuse’). If they do not, the stress is moved to the last vowel in verbal roots. Thus, in cases such as ampl[io] ‘large’ or vár[io] ‘varied’ the final vowel is lexically unstressed, whereas the high vowel is stressed in the corresponding verbal forms ampl[i.á]n ‘they enlarge’ or var[i.á]n ‘they vary’. Strikingly, we find no examples in Spanish of the reverse phenomenon, e.g., *amplio ‘large’/ *ámplio ‘I enlarge’. The following table shows the different stress positions that are found.

<table>
<thead>
<tr>
<th>Noun/Adjective /i/</th>
<th>Verb /i/</th>
</tr>
</thead>
<tbody>
<tr>
<td>cámbio</td>
<td>cámbian</td>
</tr>
<tr>
<td>lío</td>
<td>lian</td>
</tr>
<tr>
<td>ámlio</td>
<td>amplían</td>
</tr>
</tbody>
</table>

Our goal is to offer a unified analysis of the phenomena presented above within the Optimality Theory framework, based on the relationship between the prosodic and morphological factors that play some role in favoring or blocking glide formation on the one hand (Cabré & Prieto 2006), and the strategies determining stress assignment on the other (Ohannesian 2004).

The syllabification of high vowels adjacent to another vowel has had different solutions among the Romance languages. The tendency for unstressed high vowels to become glides in this context is cross-linguistically attested. In addition, Romance languages like Spanish have undergone a historical process whereby stressed mid vowels have tended to become diphthongs in specific contexts (e.g. tener ‘to have’ / tiene ‘she has’; morir ‘to die’ / muero ‘I die’). Nevertheless, there are some other contexts where Spanish preserves hiatus syllabification. The glide formation process has been amply dealt with in the literature on Spanish (Navarro Tomás 1948; Harris 1983; Harris & Kaisse 1999; Colina 1999; Hualde 1999, 2005; Cabré & Prieto 2007; Chitoran & Hualde 2007). It has been claimed that morpheme boundaries and stressed high vowels tend to block glide formation in morphologically related words (env[i.á]r ‘to send’, env[i.ó] ‘I send’; punt[u.á]r ‘to punctuate’, punt[u.ó] ‘I punctuate’), but Cabré & Prieto (2007) point out that the stressed high vowel in nominal cases such as navio ‘ship’ or policía ‘police’ does not prevent glide formation in derived forms such as nav[j]ero ‘shipping’ or polic[j]al ‘police-related’. In fact, verbal forms tend to maintain a uniform shape in the paradigm, including
deverbal nominals, e.g. cámbj, cambjámos, cambjánte; confi.o, confi.ámos, confi.anza. Otherwise, confi.amos or confjamos alternates respectively with confi.ánza or confjanzá, depending on the speaker.

Our study focuses on the resolution as diphthong or hiatus of the rising sonority sequences at the stem boundary in lexically related items. Specifically, it deals with first conjugation verbs and related nominals (including adjectives) whose stems end in a high vowel (desafío ‘challenge’ – desafiar ‘to challenge – desafiante ‘challenging’; cambio ‘change’ – cambiar ‘to change – cambiante ‘changing’). The aim of this paper is to give a unified explanation as to the outcomes of this subset of lexical items and show the prosodic and morphological reasons that allow for its surface syllabification.

Following Harris & Kaisse (1999), who include in their analysis the first conjugation verbs that exhibit a high vowel at the right edge of the stem, first conjugation Spanish verbs can be classified into three patterns according to the relationship between the stress position in verbal and related nominal stems and the stress position itself. As we see in Table 1 below, verbs and related nominals showing patterns 1 and 2 reveal the same stress position: the high vowel is always unstressed in 1, whereas it is always stressed in 2. In contrast, pattern 3 words show high vowels that are stressed in the verbal stem but unstressed in related nominals. Unexpectedly, pattern 4, the hypothetical opposite of pattern 3, with high vowels unstressed in verbs but stressed in related nominals, does not exist. In this table and elsewhere in this article, in order to facilitate interpretation, high vowels are phonetically transcribed and stressed syllables are marked independently of the conventional orthography.

(1)

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Verbs</th>
<th>Related nominals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>cámbján</td>
<td>cámbjó</td>
</tr>
<tr>
<td>2</td>
<td>rocíán</td>
<td>rocíó</td>
</tr>
<tr>
<td>3</td>
<td>ansíán</td>
<td>ánjsíá</td>
</tr>
<tr>
<td>4</td>
<td>*ánsján</td>
<td>*ansíá</td>
</tr>
</tbody>
</table>

Our analysis focuses on three issues: (a) why verbs in group 1 in the table do not shift the stress position in verbal forms as predicted by the general rule of the language (cambio N / *cambio V vs. ansía N / ansío V); (b) why there is an apparent gap in the system, i.e. group 4 in the table does not exist (*ansía N / *ánssio V); and finally, (c) why derived nouns, which exhibit a theme vowel before the suffix, display the same behavior as their related verbs in terms of the syllabic condition of the high vowel: cambjamos, cambjante *cambi.amos, *cambi.ante vs. confi.amos/ confjamos, confi.anza/confjanzá (Cabré & Prieto 2007).
Our goal is thus to provide a unified analysis and a feasible explanation for the gap presented above, the preservation of the historical stress position and, finally, the uniform behavior of deverbals in contrast to denominals.

The paper is organized as follows. Section 2 presents the data. Section 3 examines the related prosodic factors, such as the stress position, and morphological processes, such as the formation of the diminutive that exhibit analogous behavior. Section 4 presents a unified analysis within the Optimality Theory framework of the morphological and prosodic factors that control the hiatus/diphthong realization, and Section 5 includes conclusions and further implications.

2. The data

Spanish verbal forms are built on the following morphological constituents: root (R), theme vowel (TV), tense, aspect and mode morpheme (TAM) and number and person morpheme (NP). We have adopted the structure presented in Alcoba (1991, 1999:4924) in which the root and theme vowel make up what is labeled the verbal theme (‘Tema’) while the inflexional suffix (‘Flex’) includes the tense, aspect and mode morpheme and the number and person morpheme, as shown in (2). The verbal theme component serves as the base for inflexional forms and derived nominals.

\[
(2) \quad [[[R][TV]]_{Tema}[[TAM][NP]]_{Flex}]_{v}\\
 [[[[cambi][á][TV]_{Tema}[[ba]_{TAM} [mos]_{NP}]]_{Flex}]_{v} '1pp Ind.Imp.Past'\\
 [[[cambi][á][TV]_{Tema}[[Ø]_{TAM} [mos]_{NP}]]_{Flex}]_{v} '1pp Ind.Pres.'
\]

Our analysis is focused on the three persons singular and the third person plural of Present Tense, the only forms in which the root can be stressed. Taking into account those first conjugation verbs whose roots end in a high vowel, we find that verbs fall into two broad groups depending on the stress position in Present Tense forms. Verbs in group 1 – the larger one – (see 3a) keep the high vowel unstressed throughout the paradigm, and it always surfaces as a glide. In contrast, verbs in group 2 (see 3b) stress the final high vowel and preserve the hiatus across the paradigm.

\[
(3) \quad a. \quad \text{Group 1} \quad \begin{align*}
\text{cambjár} & \text{ ‘to change’} & - \text{ cámbjo ‘I change’ 1ps IP} \\
\text{limpjár} & \text{ ‘to clean’} & - \text{ límpjo ‘I clean’ 1ps IP} \\
\text{silencjár} & \text{ ‘to silence’} & - \text{ siléncjo ‘I silence’ 1ps IP} \\
\text{alivjár} & \text{ ‘to relieve’} & - \text{ alívjo ‘I relieve’ 1ps IP} \\
\text{entibjár} & \text{ ‘to cool’} & - \text{ entibjo ‘I cool’ 1ps IP} \\
\text{odjár} & \text{ ‘to hate’} & - \text{ ódjo ‘I hate’ 1ps IP} \\
\text{acaricjár} & \text{ ‘to caress’} & - \text{ acarícjo ‘I caress’ 1ps IP}
\end{align*}
\]
If we now consider the relationship of these verbs with their nominal forms, we note that the verbs in group 1 (3a) here – which coincides with pattern 1 in Table 1 – differ from those showing patterns 2 and 3. While nouns and adjectives of group 1 always diphthongize (e.g., cambjár – cámbjo ‘change’, limpjár – limpjjo ‘clean’, silencjár – siléncjo ‘silence’), words following patterns 2 and 3 show two different solutions, as we can see in (4a) and (4b) (note that group 2 shown in (3b) follows pattern 2 in Table 1).

(4) a. Root-final high vowel is stressed in both verbal and related nominal forms (corresponds to pattern 2 from Table 1):

- espían ‘they spy’ — espía ‘spy’
- rocían ‘they sprinkle’ — rocio ‘dew’
- envían ‘they send’ — envío ‘shipment’
- enfrián ‘they cool’ — frío ‘cold’

b. Root-final high vowel is stressed in verbal forms but unstressed in related nominals (corresponds to pattern 3 from Table 1):

- ansíán ‘they long for’ — ánjsja ‘worry’
- amplíán ‘they enlarge’ — ampljo ‘large’
- contrarián ‘they oppose’ — contrárjo ‘opposed’
- perpetúán ‘they perpetuate’ — perpétwo ‘perpetual’
- continúán ‘they continue’ — continwo ‘continuous’
Furthermore, there is a striking asymmetry between the two high vowels. While the final /i/ is maintained in all related lexical items, final /u/ is present in the related nominals of some verbs (see 5a) but not present in the nominals of other verbs (see 5b), which show allomorphy:

(5) a. menguar ‘to diminish’ — mèngua ‘waning’
    continuar ‘to continue’ — continuo ‘continuous’
    perpetuar ‘to perpetuate’ — pertépto ‘perpetual’
    atenuar ‘to attenuate’ — ténue ‘thin’

    b. atestiguar ‘to testify’ — testigo ‘witness’
    puntuar ‘to punctuate’ — punto ‘point’
    graduar ‘to graduate’ — grado ‘degree’
    efectuar ‘to effect’ — efecto ‘effect’

It is also worth noting that the verbs showing a different stress position from that of their corresponding nominals (i.e. ámpljo ‘large’ vs. yo amplío ‘I enlarge’) are all derived from that nominal form, as is shown in (6).

(6) ámpljo ‘large’ → amplí.an ‘they enlarge’
    grádo ‘degree’ → gradú.an ‘they graduate’
    contrárjo ‘opposite’ → contrarí.an ‘they oppose’
    pátrja ‘motherland’ → repatrí.an ‘they repatriate’
    perpétwo ‘perpetual’ → perpetú.an ‘they perpetuate’

Among the remaining verbs, we can find both directions of derivation, as shown in (7a) – nominals derived from verbs – and (7b) – verbs derived from nominals:

(7) a. renunciar ‘to give up’ → renúncia ‘abdication’
    despreciar ‘to scorn’ → desprécio ‘scorn’
    enviar ‘to send’ → envío ‘shipment’
    desafiar ‘to challenge’ → desafío ‘challenge’

1. All verbs with root-final /u/ tend to stress this vowel independently of the stress position of related nominals (even in cases such as adecú.o, against the preceptive form adécwo ‘I adjust’), except for verbs with voiced velar obstruent /g/ (frágwan ‘they forge’, atestígwan ‘they testify’):

   acénto ‘stress’ — acentú.an ‘they stress’
   perpétwo ‘perpetual’ — perpetú.an ‘they perpetuate’
   ténwe ‘thin’ — atenú.an ‘they attenuate’
   continwo ‘continuous’ — continú.an ‘they continue’
b. siléncio ‘silence’ → silenciar ‘to silence’
envidia ‘envy’ → envidiar ‘to envy’
rocio ‘dew’ → rociar ‘to sprinkle’
frío ‘cold’ → enfriar ‘to cool’

3. Related issues

In this section we will examine those factors that can help us to understand the syllabic resolution of the sequences under study. Certain morphological processes, such as the formation of diminutives, can shed light on the issue because they show the same distributional properties. We will start with nominal stress assignment and its consequences on syllabification in derived forms, and then go on to deverbal nominals and diminutive formation in the groups presented above.

3.1 Stress assignment

Spanish stress is limited to the last three syllables of the word in what is known as the three-syllable window (Harris 1983, and Roca 1988, 2005, among others). Depending on the position of the stress, words are classified into oxytones (canción ‘song’, hindú ‘Hindu’), paroxytones (casa ‘house’, árbol ‘tree’) and proparoxytones (sábana ‘sheet’, régimen ‘diet’). Stress assignment has been widely discussed in the literature and has been considered either syllable-quantity-dependent (Harris 1983; Hammond 1995; Oltra & Arregi 2005) or quantity-independent, either totally (Ohannesian 2004) or partially (Roca 1988, 2007). We assume here the analysis of Ohannesian (2004), who establishes three accentual patterns in Spanish, according to the number of syllables between the right edge of the stressed syllable and the right edge of the stem. The stress position is conditioned by morphological categories, namely, by the stem and prosodic word boundaries in the unmarked pattern. In marked and ultramarked cases, lexical marks determine the stress position.

As we can see in the table in (8) below, the last vowel of marked stems and the last two vowels of ultramarked stems have a lexical mark of zero or null prominence (underlined vowels), e.g., árbol ‘tree’, sában-a ‘sheet’, lúgubre ‘lugubrious’, régimen ‘diet’. These marked vowels cannot be stress bearers.2

2. The table in (8) shows only primitive words. Derivational suffixes also follow the same lexical marks, thus a suffix such as /ik/ bears a lexical mark of null prominence (e.g., mito ‘myth’, mitíco ‘mythic’). In contrast, inflectional suffixes lack this kind of mark. For a detailed discussion of the advantages of lexical marking of unstressed vowels instead of tonic vowels, see Ohannesian (2004).
The unmarked pattern contains words in which the right edge of the stressed syllable coincides with the right edge of the stem. If the word ends in a terminal element or epenthesis it will be paroxytone (cása, madre); if the word lacks these elements, it will be oxytone (canción, hindú). The marked pattern includes words with only one marked vowel between the right edge of the stressed syllable and the right edge of the stem. The presence of a terminal element or epenthesis yields proparoxytones (sábana, lúgubre), whereas the absence of such elements results in paroxytones (árbol). The ultramarked pattern is formed by words with two marked vowels between the right edge of the stressed syllable and the right edge of the stem (régimen). All the words belonging to this group are proparoxytones.

3.2 Verbal stress

Romance languages tend to regularize the stress position at the right edge of the verbal stem, which inhibits specific accentual patterns in verbal paradigms. Spanish is no exception and has standardized the stress position on the last stem vowel, thus blocking proparoxytones in Present Tense forms. The stress in proparoxytone nominal forms has moved to the last vowel of the stem in related verbs, as shown in the examples in (9) below.

Verbs following pattern 3 in Table 1 (exemplified in 4b) also follow this pattern and standardize the stress position.

(10) Nominal form       Verbal form

várjo ‘varied’       yo vario ‘I vary’
pátrja ‘motherland’  yo repatrio ‘I repatriate’
contínwo ‘continuous’ yo continúo ‘I continue’
eféceto ‘effect’      yo efectúo ‘I effect’

It is important to highlight that all the verbs in (9) and (10) derive from their corresponding nominals and stress movement is a consequence of this fact. Crucially, stress shift is limited to this direction of derivation, i.e. from nominal forms to verbal forms. This does not imply that whenever we have noun-to-verb derivation we also have stress shift (e.g., tibío / entibío, frío / enfrió). However, whenever we have stress shift we necessarily have noun-to-verb derivation. In contrast, verb-to-noun derivation can never entail any stress movement (e.g., renunciar ‘to give up’/ renúncio ‘I give up’ → renúncia, *renuncía ‘abdication’).

If the only possibility for stress shift is noun-to-verb derivation, this means that when nouns and verbs display different stress positions, the verbs are derived from the nouns and not vice versa. This is one of the factors that underlie the absence of hypothetical pattern 4 in Table 1. For example, a pair like *amplio A / *yo ámplio V would imply a derivation from verb to noun and stress movement to the unmarked position. The derivational pattern illustrated by the words in (9) and (10) interacts with the general tendency of the language to homogenize verbal stress to an unmarked position. In short, in noun-verb pairs, nouns always preserve their stress idiosyncrasy while verbs do not.
The verbs in group 1 (3a) and their related nominals do not follow the general tendency of the language and fail to shift the stress to the last stem vowel. Nevertheless, the diphthong resolution blocks antepenultimate stress in Present Tense forms, submitting the outcome to the general rule of the language (e.g., ódjo ‘hate’ yo ódjo ‘I hate’). The verbs in group 2 (3b) do not present any deviation because among these words the last high vowel is always stressed (e.g., amnistía ‘amnesty’ / yo amnistío ‘I amnesty’).

3.3 Syllabification in derived forms

Derived nominals with theme vowels preserve the syllabification of the sequence as it exerts in the related verb. Conversely, derived nominals with stressed suffixes diphthongize independently of the syllabic condition of the high vowel in the primitive root (Cabré & Prieto 2007). The examples in (11) below show the syllabic heritage of nominals derived from verbs in opposition to the absence of this kind of heritage in (12), with nominals derived from nouns.

(11) Verbal form — Nominal form
    cambjár — cambjánte ‘changing’
    limpjár — limpjáble ‘cleanable’
    silencjár — silencjáble ‘waning’
    averigwár — averigwáble ‘ascertainable’
    confiár — confiánza ‘trust’
    esquiár — esquiáble ‘suitable for skiing’
    punctuár — punctuáble ‘evaluable’
    fluctuár — fluctuánte ‘fluctuating’

(12) Nominal form — Nominal form
    ánnsja ‘anxiety’ — ansjóso ‘anxious’
    pátrja ‘motherland’ — patrijóta ‘patriot’
    siléncjo ‘silence’ — silencjóso ‘silent’
    ágwja ‘water’ — acwóso ‘watery’
    valja ‘value’ — valjósso ‘valuable’
    manja ‘foible’ — manjático ‘finicky’
    sangría ‘bleeding’ — sangríento ‘bloody’
    navjio ‘ship’ — navjéro ‘shipping’

4. Other Romance languages such as Occitan show similar behaviour, e.g., cámbjes ‘you change’, envies ‘you send’ (Sauzet & Ubaud 1995).

5. We distinguish the term theme vowel from word marker. Following Harris (1991), word marker or terminal element (TE) is that element which only appears at the right-hand edge of a word in its singular form and cannot be followed by any suffix except a plural marker. Such elements may be made up of any of the five vowels, either alone or followed by s, or –s itself. Theme vowel must be interpreted as the vowel that follows the verbal root.
3.4 Diminutives

As is well known, the most common Spanish diminutive suffixes are \(-it+\text{Terminal Element (TE)}\) and \(-cit+\text{TE}\) (see Footnote 5). Leaving aside special cases, the marked form \(-cito/a\) is used in bases without a terminal element (see 13a) and bisyllabic words whose stem ends in an unstressed high vowel. In these cases, the diminutive formation process preserves the syllabification of the base by the insertion of an epenthetic \(e\), as we see in (13b) below. Otherwise the unmarked form \(-ito/a\) is used.\(^6\)

\[
(13) \begin{align*}
\text{a.} & \quad \text{jardín} \text{‘garden’} & \rightarrow & \text{jardincito} \\
\text{madre} \text{‘mother’} & \rightarrow & \text{madrecita} \\
\text{b.} & \quad \text{labjo} \text{‘lip’} & \rightarrow & \text{labjecito} \\
\text{indja} \text{‘Indian’} & \rightarrow & \text{indjecita} \\
\text{c.} & \quad \text{casa} \text{‘house’} & \rightarrow & \text{casita}, ^{*}\text{casecita} \\
\text{comadre} \text{‘kinswoman’} & \rightarrow & \text{comadrita}, ^{*}\text{comadrecita} \\
\text{pintalabjos} \text{‘lipstick’} & \rightarrow & \text{pintalabitos}, ^{*}\text{pintalabjecitos (cf. labjecito, *labito)}
\end{align*}
\]

It is important to point out that the unmarked form \(-ito/a\) can also be accepted in some cases that ought to yield \(-cito/a\), such as \(\text{jardín} \rightarrow \text{jardinito}\), but the inverse situation is always ungrammatical, e.g., \(\text{casa} \rightarrow ^{*}\text{casecita}\) (see 13c).

We present in (14) some possible diminutive forms of nouns related to the three patterns described above (exemplified in 3a, 4a and 4b). The peculiarities of this process, i.e. the spontaneity conditions of its production, involve considerable variability, whether dialectal or idiolectal. Nevertheless, some forms are clearly rejected by all speakers and others are clearly preferred. In this regard, it is noteworthy that groups 1 and 2 follow their respective verbal/nominal syllabification patterns in the formation of diminutives. Interestingly, however, the diminutives of words in group 3 do not follow the pattern seen in (13b) above, but rather obey the same pattern as the stressed nominals of group 2. Diminutive formation in the three groups is illustrated in (14a), (14b) and (14c) below.

As we see in (14a), group 1 diminutive forms with \(-ecito/a\), which allow the glide to be preserved, are preferred over forms with \(-ito/a\), which are pronounced with hiatus. Conversely, forms from group 2 surface with hiatus. In addition, any form with a glide is clearly rejected (see 14b).

\[
(14) \begin{align*}
\text{a.} & \quad \text{Group 1} \\
\text{plágjo} \text{‘plagiarism’} & \rightarrow \text{plagjecito}, ^{*}\text{plagiíto (cf. plagiár, plágjan, plagiában)} \\
\text{limpjo} \text{‘clean’} & \rightarrow \text{limpito}, ^{*}\text{limpiíto (cf. limpjár, límpjan, limpjában)} \\
\text{génjo} \text{‘genious’} & \rightarrow \text{genjecito}, ^{*}\text{geniíto (cf. congenjár, congénjan)}
\end{align*}
\]

prémjó ‘award’ — premjécto, *premiító (cf. premjár, premján, premjában)
cópja ‘copy’ — cópjécta, *copiíta (cf. copjár, cópjan, copjában)
b. Group 2
espía ‘spy’ — espííta, *espjécta (cf. espiár, espían, espiában)
crio ‘kid’ — criíto, *criícto (cf. criár, crián, criában)
estría ‘groove’ — estriíta, *estriécta (cf. estriár, estrián, estriában)
rocío ‘dew’ — rociíto, *rociícto (cf. rociár, rocián, rociában)
c. Group 3
ánsoja ‘worry’ — ansiíta, ??ansjécta (cf. ansiár, ansían, ansiában)
ámpljo ‘large’ — ampliíto, ??ampljécto (cf. ampliár, amplían, ampliában)
ágrjo ‘sour’ — agriíto, ??agrjécto (cf. agriár, agrían, agriában)
pátrja ‘motherland’ — patriíta, ??pátrjécta (cf. repatriar, repatrían, repatriaban)

The behavior of nominals from group 3 is identical to that of group 1: a high vowel surfaces as a diphthong in both cases. However, strikingly, the diminutivization process works as in group 2. These diminutive forms with hiatus realization are preferred over gliding, as we may see in (14c).

As this examination of diminutives reveals, the distributional patterns of the verbs under study have a larger scope than it seemed at first sight. They are not limited to primitive nouns and verbs, nor to deverbal nominals. Diminutive forms thus give us additional evidence that can help improve our analysis.

4. Analysis

Based on the analysis by Ohannesian (2004), we will now describe the Spanish stress assignment conditions that predict stress position. We will be starting from nouns – and not from verbs – because of the lexical marks of null prominence, which can be deleted in verbs. As we have seen in the preceding section, the stress position in verbal roots is fixed when they are stressed on the last syllable.

Our analysis is built on the combination of alignment constraints (McCarthy 1993, 2004) with a faithfulness constraint (Prince & Smolensky 1993). The alignment constraints account for morphological categories such as stems or words intervening in the stress assignment, whereas the identity constraint imposes faithfulness to lexical marks of vowel prominence.

Alignment constraints are applied categorically, i.e. violations are not gradient. For example, the form régimen has only one mark of violation of ALIGN STRESS R STEM R, even though there is a distance of two syllables.
Leaving aside the constraints that are not relevant for our purposes in this paper, the first alignment constraint Align Foot Right Prosodic Word Right, defined in (15), prevents the foot that holds the stress from moving away from the right edge of the word, while the second constraint, Align Stress Right Stem Right, defined in (16), favors the coincidence of the stressed syllable with the stem.

(15) Align Foot R PrWd R
The right edge of the foot must coincide with the right edge of the prosodic word.

(16) Align Stress R Stem R
The right edge of the stressed syllable must coincide with the right edge of the stem.

An identity constraint Faithfulness Null Prominence imposes faithfulness to lexical vowel prominence marks. This constraint must be ranked higher than alignment constraints. Faithfulness must be respected because Spanish nominal stress assignment gives priority to preserving the underlying form over respecting alignment, as can be seen in the unmarked pattern.

(17) Faithfulness null Prominence
The output must respect the vowel prominence marks of the input.

These statements yield the hierarchy of constraints shown in (18):

(18) Faithfulness null Prominence >> Align Foot R PrWd R >> Align Stress R Stem R

The tableau in (19) shows an example of the unmarked pattern. Stems are given between square brackets and feet between parentheses. Candidate b. respects neither the foot alignment nor the stress alignment constraints. Candidate c. violates the stress alignment constraint.

(19)

<table>
<thead>
<tr>
<th></th>
<th>Faith null Prominence</th>
<th>Align Foot R PrWd R</th>
<th>Align Stress R Stem R</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[casca(bél)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>![]</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>c.</td>
<td>![]</td>
<td>![]</td>
<td></td>
</tr>
</tbody>
</table>

An example of the marked pattern is shown in (20), and in (21) we present an example of the ultramarked pattern. Both candidates a. in (20) and (21) are

8. The hierarchy Align Foot R PrWd R >> Align Stress R Stem R prevents undesirable outputs in verbal forms, like *cántamos – [(cánt)a]mos, stressed at the right edge of the stem cant-. In contrast, the correct form cantámos - [can(t)a]mos fulfils the foot alignment but not the stress alignment constraint.
optimal because they respect the faithfulness constraint at the top of the hierarchy. Candidates \( b \) and \( c \) are ungrammatical in spite of the fact that they respect alignment constraints.

(20)

<table>
<thead>
<tr>
<th>[sabán]a</th>
<th>Faith null Prominence</th>
<th>Align Foot R PrWd R</th>
<th>Align Stress R Stem R</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ( ) [sába]n\a</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. [sa(bán)a]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(21)

<table>
<thead>
<tr>
<th>[regimen]</th>
<th>Faith null Prominence</th>
<th>Align Foot R PrWd R</th>
<th>Align Stress R Stem R</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ( ) [régimen]</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. [re(gímen)]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. [regi(mén)]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.1 Verbal stress assignment

In relation to verbal stress assignment, we have to take into account those marks of prominence that are held by the nominals lexically related with verbs. For example, \( \textit{cambio} \) ‘change’ has the final high vowel of the stem marked as unstressed: \( \textit{cambí}_N \). So the final high vowel of the verbal stem must also be marked: \( \textit{cambí}_V \) (Cabré & Ohannesian 2007). The following two tableaux illustrate the difference between the marked pattern in (22) and the unmarked pattern in (23).

(22)

<table>
<thead>
<tr>
<th>cambí_V+o</th>
<th>Faith null Prominence</th>
<th>Align Foot R PrWd R</th>
<th>Align Stress R Stem R</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ( ) [cám.bí\Jo]</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. [cam(bí)o]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. [(cambí)].o</td>
<td>*!</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

(23)

<table>
<thead>
<tr>
<th>vací_V+o</th>
<th>Faith null Prominence</th>
<th>Align Foot R PrWd R</th>
<th>Align Stress R Stem R</th>
</tr>
</thead>
<tbody>
<tr>
<td>( ) a. [va(cí)o]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. [(vácj)o]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. [(váci)].o</td>
<td>*!</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>
Unstressed high vowels followed by another vowel surface as glides. As noted by Cabré & Prieto (2004:133), glide formation “is triggered by a general instantiation of the Onset Principle. Within OT, Onset expresses the general prosodic restriction that every syllable must have an onset and motivates the strong preference for CV syllables rather than V syllables’.

As we have pointed out, verbs presented in (3a) and (3b) (e.g., cambiar ‘to change’ or confiar ‘to trust’), can exhibit a uniform pronunciation of the raising sonority sequences throughout the paradigm. Verbs such as confiar display hiatus pronunciation even in forms with the high vowel unstressed (e.g., conf[i.á]mos). Thus, as noted, paradigm uniformity arises as another factor to determine diphthong or hiatus pronunciation. In order to account for this effect, we have to resort to a new constraint (OPMaxv), presented in Cabré & Prieto (2007), according to which the competing candidates are the whole paradigms themselves (McCarthy 2005).

(24) \[\text{OPMax}_v\]
All forms belonging to one paradigm must be realized in a uniform way.

The tableau in (25) shows how the high-ranked OPMaxv (unordered with respect to Faith) accounts for the output resolutions of the whole verbal paradigm. In the subset of verbs represented by cambiar, the correct candidate forms a diphthong, and Onset selects the optimal candidate (25a) over (25d). For ease of presentation we have not included the whole paradigm but only the forms that are relevant to the hierarchy.

(25)

<table>
<thead>
<tr>
<th>Form</th>
<th>OPMaxv</th>
<th>Faith null Prominence</th>
<th>Align Stress R Stem R</th>
<th>Onset</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. cámbjo, cambjár, cambjába, cambjámós</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. cambi.o, cambi.ár, cambi.ába</td>
<td></td>
<td>*!</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>c. cámbjó, cambi.ába</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. cambi.o, cambi.ába</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*!</td>
</tr>
</tbody>
</table>

On the other hand, in the case of verbs like confiar, without null prominence lexical marks, Align Stress R Stem R prefers the candidate that stresses the high vowel, because it stays at the right-hand edge of the stem, as we see in (26). In these cases Onset is not relevant.
The theme vowel plays a crucial role in maintaining the uniformity effects of gliding not only across the verbal paradigm but also in derived contexts. As we have seen in Section 3.3, suffixed nominals derived from verbs preserve the same syllabification, e.g., cámbjo – cambjánte ‘changing’, confi.o – confi.ánza ‘trust’. It seems that the stressed high vowel of the stem forces the hiatus outcome in the derived nominal. This correspondence is reinforced in comparison with nouns or adjectives derived from other nouns: derived nominals with stressed suffixes diphthongize independently of the syllabic condition of the high vowel of the primitive root (Cabré & Prieto 2007): vali.a ‘value’ – valjóso ‘valuable’, sangrí.a ‘bleeding’ – sangrjénto ‘bloody’.

As we have seen, the presence of the theme vowel in derived nominals constitutes the unquestionable evidence of their deverbal condition. These facts demonstrate that, at least in the pronunciation of raising sonority sequences, we must extend the scope of paradigmatic effects to deverbal forms. Kenstowicz (1996:382) proposes the “Uniform Exponence” constraint, which “minimizes the differences in the realization of a lexical item (morpheme, stem, affix, word)”. On the basis of this proposal, which evaluates sets of morphologically related words for prosodic similarity, we reinterpret the paradigmatic constraints of McCarthy (2005) to include all derived forms with theme vowel. We have illustrated our model in (27) and (28).
Boundary effects on syllabification in Spanish

4.3 Prosodic effects

In the case of longer words, the diphthong pronunciation is generalized. In order to explain this, following Cabré & Prieto (2007), we resort to a faithfulness condition called PROSODIC PROMINENCE which agglutinates three prominence conditions that apply to syllables in terms of acoustic duration: (1) syllables in stressed position are more prominent than syllables in unstressed position; (2) syllables in word-initial position are more prominent than syllables in non-initial position; and (3) syllables in short stems or words are more prominent than syllables in longer stems or words. The prominence level of a given syllable is obtained through a computation of these three pairs of prominence levels. If the syllabic prominence obtained is high, this is a clear indicator that glide formation will be blocked.\(^9\) We have indicated in italics those forms that violate PROSODIC PROMINENCE in any of these three conditions.

\[(29)\]

<table>
<thead>
<tr>
<th></th>
<th>PROSODIC PROMINENCE</th>
<th>OPM(_x)</th>
<th>OPM(_y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>fi.o, fi.ámos, fi.arémos</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>b.</td>
<td>fi.o, fi.ámos, fi.arémos</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>c.</td>
<td>fi.o, fi.ámos, fi.arémos</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>a.</td>
<td>confi.o, confi.ámos, confi.arémos</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>b.</td>
<td>confi.o, confi.ámos, confi.arémos</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>c.</td>
<td>confi.o, confi.ámos, confi.arémos</td>
<td>*!</td>
<td>*</td>
</tr>
</tbody>
</table>

\(^9\) Although a certain amount of variation is found, our analysis is based on the predominant tendency.
Having analyzed the consequences of **Prosodic Prominence** on the verbal paradigm, we will now offer a feasible account of the regularization of stress position in verbal stems. The regularization of stress position in verbal stems is morphologically driven. As Harris (1987:64) says, “each inflectional paradigm has a characteristic fixed stress pattern that admits no variation, however minimal, among individual lexical items.” As we have seen, those verbs that shift the stress from nominals to the right edge of the stem violate **Faith null Prominence** because we assume that all lexically related items share the same prominence marks (e.g., \( fábric\)a\(_N\), \( fábric\)o\(_V\), *\( fábric\)o\(_V\)). So we need a new markedness constraint that dominates **Faith null Prominence** in order to account for the fact that all verbal stressed stem forms ending in a consonant carry the stress on the last vowel (* \( \cdot VC\)\(_{VbSt}\)). The stressed stem forms are 1, 2, 3, and 6 Present Tense forms, that is, all the singular forms and the third person plural of all present tenses. This constraint allows for the emergence of the unmarked pattern in those verbal stems that hold lexical marks of null prominence.

(30)  
* \( \cdot VC\)\(_{VbSt}\) (1, 2, 3, 6 Present)

The rightmost vowel of stems ending in a consonant must be stressed in 1, 2, 3, and 6 persons of the Present tense forms.

(31)

<table>
<thead>
<tr>
<th></th>
<th>( \cdot VC)(_{VbSt})</th>
<th><strong>Faith null Prominence</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>ñy fábric(j)o</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>fábric(j)o</td>
<td>*!</td>
<td></td>
</tr>
</tbody>
</table>

Verbs of pattern 3 in Table 1 such as *ampliar also have the regularized stress position exemplified by *fabricar regardless of the presence of prominence marks. According to Harris & Kaisse (1999:124) “unmarked high vocoids surface as
peaks when there is no vocoid of greater sonority next to them; when adjacent to a non-high vocoid, they surface as glides. Syllabic [i u] in hiatus is the special case, which we represent (…) /i./ and /u./.” Following Harris & Kaisse we will mark the final high vowel of these verbs. We propose a new markedness constraint for this specific shape (* ˘V[.]_{VbSt}).

The added dot is a lexical mark that indicates the syllabic condition of the high vowel. In (33) and (34) we show how the two constraints work.

(32)  * ˘V[.]_{VbSt} (1, 2, 3, 6 Present)
The rightmost vowel of stems ending in a syllabic high vowel (marked with a dot) must be stressed in 1, 2, 3, and 6 persons in the Present Tense forms.

<table>
<thead>
<tr>
<th>verb</th>
<th>* ˘VC[ ]_{VbSt}</th>
<th>* ˘V[.]_{VbSt}</th>
<th>Faith null Prominence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ampli-</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>amplijo</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

(33)

<table>
<thead>
<tr>
<th>verb</th>
<th>* ˘VC[ ]_{VbSt}</th>
<th>* ˘V[.]_{VbSt}</th>
<th>Faith null Prominence</th>
</tr>
</thead>
<tbody>
<tr>
<td>amplijo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>amplijo</td>
<td></td>
<td></td>
<td>!</td>
</tr>
</tbody>
</table>

The examples in (31), (33) and (34) have the last vowel of the stem lexically marked with a null prominence and consequently this syllable cannot accept stress in verbal forms. The constraint * ˘VC[ ]_{VbSt} accounts for the stress shift to the last vowel of the verbal stem in fábrico, whereas the constraint * ˘V[.]_{VbSt} is necessary to explain the same stress shift for verbal stems from group 3, such as amplío.

However, Harris’s claim about the uniformity of inflectional paradigms cannot be categorical because the first group of verbs (camb[jo]N, camb[jo]V) fails to shift the stress to the last vowel of the stem. All words lexically related to this group keep the root-final vowel unstressed, invariably exhibiting a diphthong pronunciation. We exemplify the contrasting behavior of verbs from groups 1 and 3 in the following tableau.

10. Roca (2007) uses a similar strategy to mark syllabic high vowels.
This lexical marking gives us a feasible solution to the asymmetry that groups 1 and 3 show. Group 1 always diphthongizes because its high vowel does not have the hiatus mark. As we have seen above, Spanish diminutive formation provides additional empirical evidence due to the different realization of high vowels: génjo ‘genius’ – genjecito, *geni.ito (cf. congenjár, congénjan); ágrjo ‘sour’ – agri.ito, *agri.ecito, ??agrjecito (cf. agri.ár, agri.an, agri.ában).

Our proposal also takes into account the non-existence of pattern 4 in Table 1 (*ampl[jo]A – *yo ampl[jo]V), that is, verbs with stress marks lexically related to nouns lacking the same mark. For Harris & Kaisse (1999), the absence of a pair like this can be explained by the fact that in a word such as cóp[j]a ‘copy’ the high vowel is an unmarked vowel, whereas in án[s]i.a the high vowel is a marked vowel, similar to cases such as vac[i]o. A word like *vác[j]o is not possible because adjective and verb would derive from two distinct underlying representations, marked for the adjective but unmarked for the verb. The problem of their analysis is that ansi.a is predicted, leaving it unclear as to how the usual pronunciation of án[s]i.a can be accounted for. In our study, in contrast, Onset prevents ansi.a from being the optimal candidate (see 25).\(^{11}\)

Our analysis, thus, explains this gap with reference to the analogous forms in verbal environments. It is well known that derived contexts tend to display uniform patterns. This can be considered an example of the emergence of the unmarked (McCarthy & Prince 1994). Verbs of pattern 3 in Table 1 are always derived forms and this is what blocks stress shift to the left. Therefore, it is not possible for a primitive form with an unmarked stress position such as vacío to exhibit a marked stress position typical of derived forms: *vácjo. We account for this by resorting to the requirements presented above, particularly the constraint *\( \text{V}_{\text{vbSt}} \), which blocks the presence of an unstressed high vowel (in the corresponding verbal forms) that

\(^{11}\) For more details about such cases, see Cabré & Prieto (2007).
carries a dot in the stem at the lexical level. This constraint is ranked higher, thus forcing the violation of faithfulness and blocking the possibility of pattern 4.

5. Conclusion

In this paper we have offered a unified analysis of Spanish first conjugation verbs whose stem ends in a high vowel, framing the phenomenon in the general context of the language and avoiding undesirable predictions.

We have thus addressed the three basic questions raised at the beginning of the paper: (a) we have answered why verbs of the first and largest group do not regularize verbal stress position; (b) we have described the crucial role of the theme vowel in clarifying the behavior of derived forms; and (c) we have provided conclusive evidence denying the possibility of a hypothetical pattern 4. We have contextualized all these facts in the general processes of the language and reinforced our argument by means of additional evidence from the formation of diminutives.

References


The purpose of this study was to test prosodic and segmental effects on vowel intrusion duration in Spanish /ɾC/ clusters. For each cluster, I measured the acoustic duration of the intervening intrusive vowel and then analyzed the mean intrusive vowel duration under the scope of seven hypotheses based on prosodic and segmental factors. The current study consisted of twenty-nine participants across six countries and I obtained a total of 496 intrusive vowels. The study suggests that one prosodic factor, namely across a word boundary, and one segmental factor, order of constriction location, significantly affect intrusive vowel duration; data analysis for prosodic stress, heterorganic vs. homorganic, C₂ voicing, and manner and place of articulation did not evidence significant results. Finally, I discuss the findings in theoretical terms, using Articulatory Phonology (Browman & Goldstein 1989, et seq.), including the prosodic (π-) gestural model (Byrd & Saltzman 2003; Byrd et al. 2006).

1. Introduction

A current challenge in recent work on Spanish sound structure is to understand the conditions that determine vowel intrusion and the consequences that vowel intrusion may have on Spanish phonology.¹ In Spanish, an intrusive vowel (henceforth, IV) occurs between an alveolar tap, /ɾ/, and its adjacent consonant,² as in (i) tautosyllabic, /Cɾ/ clusters (e.g. o.tro [tʰɾ] “other”) or (ii) heterosyllabic, /ɾC/ clusters (e.g., par.te [ɾʰ.t] “part”); there also exists a third environment in which the IV occurs after a tap (Araujo 1894; Gili Gaya 1921), as in (iii) utterance final position.

¹I would like to thank Joaquín Romero for his insightful suggestions and Travis G. Bradley and two anonymous reviewers for their helpful comments on a previous version of this paper.

²See Davidson (2003), Hall (2003, 2006), and Riera, Romero, & Parrell (2009, this volume) for non-Spanish studies.

²To my knowledge, Malmberg (1965:44) is the only study that suggests that an IV can occur in Spanish in any consonant cluster (e.g. resbalar [resʰ.βa.lár] “to slide”).
(or ‘word final, before a pause’): following /ɾ/ (e.g. Voy a escribir [ɾʷ] “I am going to write”). Observe in these examples and throughout the paper that I follow previous phonetic representation of the IV in Spanish (Lenz 1892; Malmberg 1965; Bradley 2004, 2005, 2007; Bradley & Schmeiser 2003; Kilpatrick, Kirby, & McGee 2006; Schmeiser 2007) by denoting the IV as the superscripted schwa, [ᵊ].

Current work (Colantoni & Steele 2005, 2007; Kilpatrick, Kirby, & McGee 2006; Ramirez 2002, 2006; Schmeiser 2006, 2007) has discussed which prosodic and segmental factors have significantly affected IV duration in /Cr/ clusters. These studies have greatly enhanced our understanding that mean IV duration is not as random as previously thought (Gili Gaya 1921) and that it is conditioned by prosodic and segmental factors. For example, all previously-mentioned studies concur in that significantly longer IV mean duration is evidenced after a voiced consonant as opposed to a voiceless one. However, rather curiously, there exists a striking deficiency in empirical research of the phenomenon in the /rC/ environment. Only three known studies (Bradley 2004; Romero 1996, 2008) have included discussion on vowel intrusion in Spanish /rC/ clusters; two of which (Bradley 2004; Romero 1996) discussed theoretical representations and one examined perception (Romero 2008). Thus, there exists a gap in our current understanding of what factors, prosodic or segmental, affect IV duration in /rC/ clusters.

The current study seeks to fill this gap in three distinct ways. First, it measures the prosodic and segmental effects on IV durational variability in an environment that has not been tested in Spanish. Second, it is the first known study in Spanish intrusive vowel research to offer empirical data from subjects from six countries. Finally, the current study is particularly relevant to the linguistics community in that its findings offer new theoretical implications for the prosodic (π-) gestural model (Byrd & Saltzman 2003; Byrd et al. 2005; Byrd et al. 2006). IV analysis in Spanish /rC/ clusters offers a particularly ideal testing environment for the prosodic (π-) gestural model because we are able to examine IV duration across both a syllable and a word boundary. This is in stark contrast to /Cr/ clusters, which are always tautosyllabic.

The layout of the rest of the paper is as follows: §2 offers a survey of earlier work and lists the hypotheses posited in the current study; §3 lays out the experimental design; §4 presents the data results; §5 discusses the data results within the prosodic (π-) gestural model and §6 concludes.

2. The intrusive vowel

2.1 Why IVs merit special investigation

IVs represent a unique category of vowel insertion and differentiate themselves from epenthetic vowels in three ways. First, IVs in /Cr/ clusters adopt the formant
structure of the nuclear vowel (Quilis 1970)\textsuperscript{3} and therefore when an IV achieves phonological full-vowel status diachronically, we can predict which vowel it will be (e.g., *pereces < preces “prayers”). It should be noted, however, that the formant structure of an IV in the heterosyllabic context (i.e. /rC/ clusters) has not been measured. I also observe that all known diachronic cases of an IV that achieves full-vowel status only exist in /Cr/ clusters. Though it is not within the parameters of the current study to analyze IV formant structure, future research with specific regard to /rC/ clusters is required to ascertain any differences compared to its tautosyllabic counterpart.

Second, unlike epenthetic vowels, IVs are not part of the syllabic make-up of a lexical item. Thus, a word like, parte “part”, illustrates the syllabic division:

\begin{enumerate}
  \item [páɾ.te] parte “part”
  \item *[páɾ.ɾ.te] parte “part”
\end{enumerate}

We see in (1a) that the syllable boundary is between the tap and the adjacent dental stop, /t/, in the disyllabic word. However, note that the IV is never the nucleus of the syllable as in *(1b).

Finally, we may examine this restriction in terms of prosodic stress in paraparoxytonic words, such as ártico [ár.ti.ko] “arctic”. Main stress in Spanish is confined without exception to a three-syllable window at the right-edge of the morphological word (Harris 1995:869). If the IV in the word, ártico [ár.ti.ko] “arctic”, created a new syllable, the end result would be *[á.r.ɾ.ti.ko], which would require extending beyond the three-syllable window. A shift in stress, such as *[a.r.ɾ.ti.ko] (or *[a.r.ɾ.ti.ko]) has not been attested.

\subsection{2.2 The intrusive vowel in Spanish /rC/ clusters}

Recall that whereas /Cr/ clusters are tautosyllabic, the second environment is comprised of /rC/ clusters, which are (almost)\textsuperscript{4} always heterosyllabic. In Spanish /rC/ clusters, the consonant following the tap may be one of the following:

\begin{enumerate}
  \item voiceless fricative [f, x, θ, s]
  \item voiceless stop [p, t, k]
  \item voiced stop [b, d, g]
  \item voiced approximants [β, ð, y]
  \item nasals [n, m]
  \item liquids [l]
\end{enumerate}

\footnotesize
3. See Colantoni & Steele (2005) for an analysis that is contrary to this claim.
4. There are six words in Spanish which evidence /rC/ clusters as a tautosyllabic, complex coda cluster (e.g. perspectiva [pers.pek.tíβa] “perspective”).
I point out that not all of the possibilities listed in (2) were tested in the current study, due to limitations of the chosen text read by each subject. Future researchers will want to control for all the possibilities listed in (2).

In light of previous research, Gili Gaya (1921) is the first to take durational measurements in /rC/ clusters. In his study, he notes that IV duration in all three environments ‘varies within the same word uttered by the same speaker’ (279). Of the 103 /rC/ tokens analyzed, 61.2% contain IVs and all clusters in which a tap was realized\(^5\) evidenced an IV.

Blecua’s (2001) study on Spanish rhotics in Peninsular Spanish reveals significant results for voicing and manner of articulation. That is, rhotic duration is longer before a voiced consonant and a pattern is exhibited in terms of manner of articulation in that the duration is shortest before fricatives, followed by stops, approximants, nasals (slightly shorter than approximants) and finally laterals. Her data seem to suggest that segmental factors might play a role in IV durational variability as well.

Other recent studies discuss realizations of an /rC/ cluster in gestural terms. Romero (1996) discusses the realization of [ɾd̪] in gestural terms, noting that the cluster is pronounced in sequential fashion, as opposed to the [ld] cluster, which incurs more gestural overlap. Bradley (2004) attributes the sequential realization of the [ɾn̪d̪]\(^6\) to the ‘execution of the release phase of the rhotic, which separates the tap constriction from that of the following consonant, thereby ensuring an open transition between the two’ (198). In his study of perception of an /rC/ cluster, Romero (2008) adds that Spanish IVs for any /rC/ cluster are ‘epiphenomena, by-products of the temporal overlap between articulatory gestures, not the result of a deliberate process of insertion’ (61).

Heterosyllabic /rC/ clusters differ from tautosyllabic /Cr/ both in terms of syllable structure and in terms of phonotactic restrictions on the adjacent consonant (16 possible adjacent consonants for /rC/ clusters, as opposed to 10 for /Cr/ clusters). In short, to deepen our understanding of vowel intrusion in Spanish and its potential implications for Spanish phonology, we must add to the previously-mentioned growing body of work in /Cr/ clusters by also assessing what factors affect IVs in /rC/ clusters.

2.3 The aim of the current study

The aim of the current study was to test the effect of prosodic and segmental factors on IV duration in Spanish /rC/ clusters. Based on previous findings for /Cr/ clusters

---

5. Keep in mind that the tap, /ɾ/, though the basic realization, is not the only possibility. The /ɾ/ may be realized as an approximant, [ʃ], a trill, [r], or even elided, [ø].

6. Observe that Bradley includes the IV in the sequence, whereas Romero (1996) does not.
(Colantoni & Steele 2005, 2007; Kilpatrick, Kirby & McGee 2006; Ramírez 2002, 2006; Schmeiser 2006, 2007), coupled with Blecua’s (2001) findings regarding rhotic duration, I posited the seven hypotheses in Figure 1:

1. /ɾC/ clusters across a word boundary will evidence longer IVs than word-internal ones.
2. Stressed /VeɾC/ demisyllables will evidence longer IVs than unstressed ones.
3. /ɾC/ clusters with a front-to-back order of constriction location (i.e., /ɾk/, [ɾɣ]) will evidence longer IVs than ones with a back-to-front order (i.e. /ɾp/, /ɾm/).
4. Heterorganic /ɾC/ clusters (i.e. /ɾt/, /ɾp/, /ɾb/, /ɾk/, /ɾg/, /ɾm/, /ɾθ/, /ɾx/) will evidence longer IVs than homorganic ones (i.e. /ɾt/, [ɾð], /ɾm/).
5. /ɾC/ clusters in which C is voiced will evidence longer IVs than ones in which C is voiceless.
6. Stops (/ɾt/ and /ɾk/) will evidence the shortest IV duration, followed by approximants ([ɾβ], [ɾð], [ɾɣ]) and finally, nasals (/ɾn/ and /ɾm/).
7. Labials (/ɾp/, /ɾb/, /ɾm/) will evidence the shortest IV duration, followed by coronals (/ɾt/, [ɾð], /ɾn/) and dorsals (/ɾk/ and [ɾɣ]).

Figure 1. The seven posited hypotheses

The first two hypotheses tested prosodic factors and the last five tested segmental ones. The first hypothesis tested the effects of word position; the second tested the effects of prosodic stress; the third tested the effect of order of constriction in the vocal tract; the fourth tested the effect of heterorganic vs. homorganic clusters; the fifth tested the effects of the [voice] feature of the following consonant; the sixth and seventh tested manner and place of articulation, respectively. For each hypothesis, the data were tested both across all subjects and across country of origin.

3. Experimental design

3.1 Subjects

The current study was comprised of twenty-nine subjects, 19 male and 10 female, across six countries: Spain, Mexico, Argentina, Guatemala, Colombia, and Ecuador. Subjects ranged from four to six in number for each country. The selection of corpus subjects was guided by the criterion of rhotic standard varieties. The overall corpus consists of subjects from over twenty-five different regions throughout the Spanish-speaking world. All subjects in the current study are native speakers of Spanish and produce the Spanish tap (as opposed to an approximant, [ɾ], a trill, [r], or deletion, [ø]) in /ɾC/ clusters as the canonical realization.
3.2 Data collection

Each subject read a passage of approximately 420 words in length. The passage contained 23 /rC/ clusters and 33 /Cr/ clusters (not treated in the current study). The current corpus elicited 589 /rC/ clusters, from which 496 IVs were present and were spectrographically and aurally analyzed. The subjects were originally recorded on reel-to-reel tape,\(^7\) and the recordings were later digitized and stored in CD in MPEG format at a sample rate of 22,050 Hz and sample size of 16-bit. Spectrographic analyses and waveforms were taken using Speech Analyzer 2.6. The spectrogram in Figure 2 illustrates a canonical /rC/ cluster with an intervening IV, marked with an arrow:

![Spectrogram](image)

**Figure 2.** A canonical /rC/ cluster in partes [ar^tʰ.e] “parts”

Observe in Figure 2 the tap, followed by the IV and finally the voiceless dental stop (with a short burst, transcribed here as [ʰ]); on each side of the cluster, we note the flanking vowels, which are /a/ and /e/, respectively.

3.3 Data analysis

Given the exploratory nature of the study, coupled with the notion that IVs do not necessarily surface in every consonant cluster, an imbalance of speakers and/or tokens was to be expected for any given hypothesis; the analysis of unbalanced

---

7. The corpus was made available by John Dalbor at the Pennsylvania State University. It has since then been digitized under the supervision of Eric Baković at the University of California, San Diego.
data was corrected by the orthogonal sums of squares (also known as ‘Type III sums of squares’). Using a regression framework, durational means were taken across subjects and single-factor, within-subjects ANOVAs were employed to determine significant results (p<0.05). Finally, all data were analyzed using the statistics software program, R.

4. Results

Of the 589 /ɾC/ clusters collected, 84.2% of the clusters elicited an IV for a total of 496 tokens. Table 1 offers the number (of subjects), and the IV durational mean, range and standard deviation for the data from each country. The number of tokens collected and their percentage of occurrence in /ɾC/ clusters are also included.

Table 1. Overall IV breakdown by country

<table>
<thead>
<tr>
<th>Country</th>
<th>Number, Mean, Range, and Standard Deviation</th>
<th>Number of IVs (496)</th>
<th>% of IV occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spain</td>
<td>6 24.10 21.22–27.73 2.45 128</td>
<td>94.1%</td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td>4 22.84 19.69–27.32 3.55 78</td>
<td>84.78%</td>
<td></td>
</tr>
<tr>
<td>Mexico</td>
<td>4 22.93 21.71–25.71 1.87 73</td>
<td>81.1%</td>
<td></td>
</tr>
<tr>
<td>Guatemala</td>
<td>5 30.17 27.55–34.13 2.45 56</td>
<td>80.0%</td>
<td></td>
</tr>
<tr>
<td>Colombia</td>
<td>5 23.48 15.07–26.50 4.75 69</td>
<td>75%</td>
<td></td>
</tr>
<tr>
<td>Ecuador</td>
<td>5 25.44 21.71–31.96 4.31 92</td>
<td>84.4%</td>
<td></td>
</tr>
</tbody>
</table>

From the data in Table 1, we detect that the data from Spain evidenced both the highest number of tokens and the highest percentage of IVs. The lowest percentage was found in the data from Colombia. Finally, I note that the data from Ecuador evidenced a percentage closest to the average of all countries examined.

4.1 Across all subjects

The first part of the results consists of data taken across all subjects, irrespective of country of origin. In the following Table 2 includes the token count, mean, range, and standard deviation for /ɾC/ clusters across all subjects. For the first five hypotheses, there are two factors tested, but take note that the sixth and seventh hypotheses test three factors. For each hypothesis, a significant difference (p<0.05) in IV duration between factors is marked with mean figures in bold and marked with ‘***’. In the third hypothesis, I abbreviate the order of constriction factors, back-to-front (i.e. BTF) and front-to-back (i.e. FTB); I also abbreviate manner of articulation (i.e. MOA) and place of articulation (i.e. POA) for hypotheses 6 and 7. Finally, Table 3 follows and offers the p values for each hypothesis:
Table 2. Token count, mean, range, and standard deviation for /ɾC/ clusters across all subjects

<table>
<thead>
<tr>
<th>H#</th>
<th>Factor</th>
<th>Factor 1</th>
<th></th>
<th></th>
<th>Factor 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>Mean</td>
<td>Range</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>1</td>
<td>Word-internal vs. Across word boundary</td>
<td>419</td>
<td>33.60**</td>
<td>16.47–46.31</td>
<td>6.56</td>
<td>79</td>
<td>40.05**</td>
</tr>
<tr>
<td>2</td>
<td>Stressed vs. Unstressed</td>
<td>312</td>
<td>34.75</td>
<td>21.53–46.24</td>
<td>6.34</td>
<td>184</td>
<td>35.37</td>
</tr>
<tr>
<td>3</td>
<td>BTF vs. FTB</td>
<td>84</td>
<td>31.25**</td>
<td>21.00–44.47</td>
<td>6.74</td>
<td>147</td>
<td>36.59**</td>
</tr>
<tr>
<td>4</td>
<td>Heterorganic vs. Homorganic</td>
<td>224</td>
<td>34.69</td>
<td>22.71–9.23</td>
<td>6.20</td>
<td>272</td>
<td>35.11</td>
</tr>
<tr>
<td>5</td>
<td>Voiced vs. Voiceless</td>
<td>191</td>
<td>34.83</td>
<td>23.58–46.30</td>
<td>5.52</td>
<td>295</td>
<td>35.22</td>
</tr>
</tbody>
</table>

Table 3. P values from ANOVA (two-factor with replication) across all speakers for /ɾC/ clusters

<table>
<thead>
<tr>
<th>H#</th>
<th></th>
<th>Factor 1</th>
<th></th>
<th></th>
<th>Factor 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>Mean</td>
<td>Range</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>H1</td>
<td>Across All</td>
<td>&lt; 0.001**</td>
<td>0.332</td>
<td>0.003**</td>
<td>0.891</td>
<td>0.481</td>
<td>0.003**</td>
</tr>
</tbody>
</table>

Pertaining to position relative to the word, mean IV duration was significantly longer (40.05 ms) across a word boundary than word-internally (33.60 ms). A front-to-back order of constriction elicited a significantly longer IV (36.59 ms) when compared to a back-to-front one (31.25 ms). Both manner and place of
articulation exhibited significant results. In the case of manner of articulation, mean IV duration was significantly longer before approximants (36.96 ms), followed by stops (35.22 ms) and finally nasals (31.18 ms); in terms of place of articulation, mean IV duration was significantly longer before dorsals, followed by coronals, and finally labials. In Figure 3, which summarizes these findings, we observe that the greatest difference was found in the prosodic factor of word position; we also note the continuum that forms as concerns both manner and place of articulation:

![Mean duration across all speakers](image)

**Figure 3.** Hypotheses that exhibited significant findings

In lieu of hypotheses that did not evidence significant findings, there was slightly longer mean IV duration in unstressed demisyllables (35.37 ms) than in stressed ones (34.75 ms). With regard to whether two segments share the same place of articulation, homorganic (35.11 ms) and heterorganic (34.69 ms) clusters exhibited the highest p value (0.891) of all the hypotheses posited. Finally, unlike the findings for /Cr/ regarding voicing (Colantoni & Steele 2005, 2007; Kilpatrick, Kirby, & McGee 2006; Ramírez 2002, 2006; Schmeiser 2006, 2007) which strongly suggest longer mean IV duration after a voiced consonant, the same does not appear to be true for its /rC/ counterpart. That is, a slightly longer mean IV duration was found before a voiceless consonant in /rC/ clusters, but no real trend emerged from the current study.

4.2 By country of origin

For convenience, I have included all of the results by country of origin in Appendix 1. With direct regard to position within the word, longer IV duration was found across a word boundary in the data from all countries except those from Spain, which
Benjamin Schmeiser had a slightly longer, insignificant mean IV duration in word-internal cases. Data results for position within the word by country of origin are illustrated in Figure 4:

![Figure 4](image-url)

**Figure 4.** Mean IV duration (ms) by country for word-internal and across word boundary

The data from three countries were significant, namely Argentina, Mexico, and Guatemala, whose data evidenced the greatest difference in mean IV duration.

In light of order of constriction location, the data from all countries evidenced a trend of longer mean IV duration in a front-to-back sequence for the data from all six countries, though rather curiously I found the data from only one country (Colombia) to be significant, with the noted data from Spain as marginally significant. Data results for order of constriction location by country of origin are illustrated in Figure 5.

In terms of manner and place of articulation, however, an analysis by country of origin revealed few tendencies. In the case of manner of articulation, the shortest mean IV duration was found before nasals in the data of all six countries. After that, the data from four countries suggest shorter IV duration before stops, with the data from Mexico and Guatemala suggesting shorter duration before approximants. In addition, only the data from one country, namely Ecuador,
were significant. As concerns place of articulation, the shortest IV duration was found before labials in the data of all six countries. After that, however, no tendency emerged as the data were split regarding coronals and dorsals. Finally, the data from two countries were significant, namely those from Spain and Colombia, though they exhibited different results; that is, the data from Spain suggested shortest mean IV duration before labials, followed by dorsals, and finally coronals. To the contrary, the data from Colombia suggested shortest mean IV duration before labials, but then followed by coronals, and finally by dorsals. In short, an analysis of manner and place of articulation revealed dialectal variation, few significant results and few tendencies.

With direct regard to those hypotheses which did not evidence significant findings across all subjects, no real tendencies emerged. In the case of prosodic stress, longer mean IV duration was found in stressed demisyllables in the data from Guatemala and Colombia, though results were insignificant; longer mean IV duration was found in unstressed demisyllables in the data from Spain, Argentina, Mexico, and Ecuador, with only the data from Argentina exhibiting significant results. Pertaining to sharing the same place of articulation, longer mean IV duration was evidenced in heterorganic clusters for the data from two countries, namely
Spain and Mexico and longer mean IV duration was evidenced in homorganic clusters for four countries, namely Argentina, Guatemala, Colombia, and Ecuador. I note, however, that the data from all six countries exhibited insignificant results. Finally, mean IV duration was split with regard to C₂ voicing. That is, longer mean IV duration was found before a voiced consonant in the data from Spain, Argentina, and Ecuador; that withstanding, longer mean IV duration was found before a voiceless consonant in the data from Mexico, Guatemala, and Colombia. I again note that none of the data taken was significant.

In short, an analysis by country of origin revealed few tendencies, but suggested dialectal variation with regard to what factors affect IV duration. I conclude the current section by offering the p values for each hypothesis by country in Table 4:

<table>
<thead>
<tr>
<th>Country</th>
<th>H1</th>
<th>H2</th>
<th>H3</th>
<th>H4</th>
<th>H5</th>
<th>H6</th>
<th>H7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spain</td>
<td>0.838</td>
<td>0.308</td>
<td><strong>0.065</strong></td>
<td>0.194</td>
<td>0.821</td>
<td>0.493</td>
<td><strong>0.045</strong></td>
</tr>
<tr>
<td>Argentina</td>
<td><strong>0.003</strong></td>
<td><strong>0.015</strong></td>
<td>0.668</td>
<td>0.367</td>
<td>0.816</td>
<td>0.294</td>
<td>0.590</td>
</tr>
<tr>
<td>Mexico</td>
<td><strong>0.041</strong></td>
<td>0.814</td>
<td>0.636</td>
<td>0.863</td>
<td>0.366</td>
<td>0.310</td>
<td>0.681</td>
</tr>
<tr>
<td>Guatemala</td>
<td><strong>0.007</strong></td>
<td>0.856</td>
<td>0.256</td>
<td>0.444</td>
<td>0.518</td>
<td>0.780</td>
<td>0.503</td>
</tr>
<tr>
<td>Colombia</td>
<td>0.361</td>
<td>0.367</td>
<td><strong>0.011</strong></td>
<td>0.712</td>
<td>0.440</td>
<td>0.156</td>
<td><strong>0.013</strong></td>
</tr>
<tr>
<td>Ecuador</td>
<td>0.076</td>
<td>0.863</td>
<td>0.644</td>
<td>0.939</td>
<td>0.934</td>
<td>0.011**</td>
<td>0.894</td>
</tr>
</tbody>
</table>

**=statistically significant with alpha set 0.05
*=marginally statistically significant with alpha set between 0.051 and 0.065

In conclusion, Table 4 illustrates significant results for the data from every country in at least one hypothesis. The first hypothesis exhibited the highest number (3) of countries whose results evidenced significant data and the fourth and fifth hypotheses evidenced the lowest number with no significant data. In the following section, I determine which of the original hypotheses posited may be deemed supported, based on the results.

4.3 Supported hypotheses

I define a supported hypothesis as one for which the data results exhibited significant results across all subjects and evidenced a trend in the analysis by country of origin. I define a trend as data from at least five of the six countries that corroborate the given hypothesis; additionally, these data must be (marginally) significant in at least two of the countries tested. Using these definitions, I assess each hypothesis to determine if it is one that is supported. For the reader’s convenience, I repeat the original hypotheses from §2.3:
1. /ɾC/ clusters across a word boundary will evidence longer IVs than word-internal ones.
2. Stressed /VeɾC/ demisyllables will evidence longer IVs than unstressed ones.
3. /ɾC/ clusters with a front-to-back order of constriction location (i.e., /ɾk/, [ɾy]) will evidence longer IVs than ones with a back-to-front order (i.e., /rp/, /rm/).
4. Heterorganic /ɾC/ clusters (i.e. /ɾt/, /ɾp/, /ɾb/, /ɾk/, /ɾɡ/, /ɾm/, /ɾθ/, /ɾx/) will evidence longer IVs than homorganic ones (i.e. /rt/, [ɾð], /rm/).
5. /ɾC/ clusters in which C is voiced will evidence longer IVs than ones in which C is voiceless.
6. Stops (/t/ and /k/) will evidence the shortest IV duration, followed by approximants ([β], [ð], [ɣ]) and finally, nasals (/n/ and /m/).
7. Labials (/p/, /b/, /m/) will evidence the shortest IV duration, followed by coronals (/t/, [ð], /n/) and dorsals (/k/ and [ɣ]).

Figure 6. The seven posited hypotheses

In lieu of word position, the results across all subjects and the data from five countries (data from three countries were significant) suggested longer mean IV duration across a word boundary; hypothesis 1 is supported. Note that I do not discuss hypotheses 2, 4 and 5 in more depth, given that across all subjects data were not significant. In terms of order of constriction location, the results across all subjects and the data from all six countries (data from two countries were (marginally) significant) suggested longer mean IV duration in a front-to-back sequence; hypothesis 3 is supported. As concerns manner of articulation, the results across all subjects suggest a different ordering than what was hypothesized. Moreover, analysis by country of origin exhibited a trend of four countries that shared the same ordering, however, the data from only one was significant; hypothesis 6 is not supported. Finally, pertaining to place of articulation, though data across all subjects supported the hypothesis, by country of origin analysis revealed no trends; hypothesis 7 not supported.

In conclusion, hypotheses 1 and 3 were deemed supported in the current study; hypotheses 2, 4–7 were deemed unsupported. In the following section, I discuss the supported hypotheses in theoretical terms.

5. Discussion

5.1 A gestural approach

For a theoretical discussion of the supported hypotheses, I utilize a gestural approach. Articulatory Phonology (Browman & Goldstein 1989, et seq.) is rooted
in basic units, called gestures, which are dynamically defined articulatory movements that produce a constriction in the vocal tract. Articulatory gestures have internal duration, a property represented abstractly in terms of a 360° cycle and phonetic timing is thus intrinsic to the phonological representation (Browman & Goldstein 1989). Adjacent gestures are temporally coordinated with respect to each other and may exhibit varying degrees of overlap. Finally, consonantal articulations are superimposed on vocalic gestures, which are themselves articulatorily adjacent (Gafos 1999).

In terms of timing patterns of CC (consonant-consonant) clusters, recent research (Byrd 1996a,b; Krakow 1999; Byrd et al. 2000; Browman & Goldstein 2001; Byrd & Saltzman 2003; Nam & Saltzman 2003; Goldstein, Byrd, & Saltzman 2006) has noted that timing is more stable in tautosyllabic CC clusters than heterosyllabic ones. Nam & Saltzman’s (2003) empirical results suggest a typology of intergestural coupling, namely that CC intergestural timing is least variable in onsets, more variable in codas and most variable across a word (or syllable) boundary (2256). This is particularly relevant to the current study in that, when π-gestures are present, they induce greater decreases of intergestural overlap (i.e. greater IV duration) in CC structures with the less stable relative phasing (see §5.2 for further discussion).

Apropos gestural timing and /rC/ clusters in Spanish, Romero (1996) discusses the blending of lingual gestures, including the tap. A possible explanation for lack of assimilation in [ɾð] is due to duration. Minimal overlap in such clusters would facilitate the ballistic movement required for successful articulation of the extra-short /ɾ/ in this context. Bradley (2004) offers a gestural account of rhotic realizations, using both a gestural approach and Optimality Theory. What is of particular relevance to the current study is the gestural representation he offers in Figure 7:

![Figure 7](image-url)
Observe in Figure 7 that the tap and the stop are superimposed onto the vowel gestures; in the transition between the two consonants, the listener perceives the underlying vowel in a very short duration. The data from the current study suggests that the speaker alters intergestural timing in terms of a prosodic boundary and/or order of constriction location. In the following sections, I account for the changes in intergestural timing in both environments.

5.2 Word position

To account for significantly longer mean IV duration across a word boundary, I utilize the prosodic ($\pi$-) gestural model (Byrd & Saltzman 2003; Byrd et al. 2005; Byrd et al. 2006). There are four important points to consider in this model. First, $\pi$-gestures extend over an interval at a juncture to slow the time course of articulatory gestures that are active during that interval (Byrd et al. 2005). Second, a stronger $\pi$-gesture slows down the clock (i.e. local speaking rate) more than a weaker one. Third, a $\pi$-gesture becomes stronger as we move up the prosodic hierarchy. Finally, with direct regard to coda release, Byrd’s et al. (2005) findings suggest that the release was longer along a phrase edge than it was in phrase internal position.

For the current study, as the clock rate of /ɾC/ clusters is slowed down within the $\pi$-gesture and the tap coda release is lengthened, the underlying vowel gesture becomes (even) more perceptible than, for example, in a /Cɾ/ cluster which involves a tighter constriction. Figure 8, shown on the following page, illustrates greater gestural disassociation as the $\pi$-gesture strengthens. Observe that the figure contains an arrow to call the reader’s attention to the prosodic tier; additionally, the stronger $\pi$-gesture in Figure 8(b) is illustrated in bold and with a thicker line for reader convenience:

In Figure 8(a), shorter IV duration is evidenced in word-internal clusters, given that a syllable boundary is a lower prosodic domain than a word boundary, and that the concomitant ‘weaker’ $\pi$-gestures associated with the syllable boundary will induce a smaller decrease in gestural overlap. On the contrary, the data from the current study suggest that a word boundary, given that it is a higher prosodic domain than a syllable boundary, constitutes a ‘stronger’ $\pi$-gesture. This stronger $\pi$-gesture, which is illustrated in Figure 8(b), slows down the clock rate, which consequently results in less temporal coproduction (i.e. less gestural overlap), and in the listener perceiving, to a greater degree, the underlying vowel gesture.

5.3 Order of constriction location

As concerns order of constriction location, recall that longer mean IV duration was evidenced in a /ɾ/-to-back order of constriction location (e.g. porque [pórɛ.ke] “because”) than in a /t/-to-front ordering (e.g. transforma [tɾ̃rans.fóɾɛ.ma]
“transforms”). The results are not surprising, given certain restrictions in the vocal tract. More specifically, the back of the tongue moves slower than the tongue tip or the lips (Hardcastle 1973:266), and thus, we are not surprised to see less gestural overlap from the apico-alveolar tap to the velar consonant (in this case, /k/ or [ɣ]).

6. Conclusion

The current study sought to test prosodic and segmental effects on intrusive vowel duration in Spanish /rC/ clusters. An analysis across all subjects exhibited significant results for four of the hypotheses posited, namely word position, order of constriction location, and manner and place of articulation. An analysis by country of origin, however, revealed much dialectal variation.

Based on the findings from both analyses, I defined a supported hypothesis and according to this definition, one hypothesis based on prosody (word position)
Prosodic and segmental effects on vowel intrusion duration in Spanish /ɾC/ clusters and one based on segmental factors (order of constriction location) were deemed supported. These results differ from data results taken from the same corpus and hypotheses for /Cr/ clusters, which suggest that prosodic factors do not play a role in Spanish IV duration, and that, except for one hypothesis (#4) segmental factors do play a role (Schmeiser 2006). In short, the results corroborate the notion that gestural overlap in the Spanish, heterosyllabic context should be viewed differently from the tautosyllabic context.

Finally, I discussed the supported hypotheses in theoretical terms, framing the empirical results on word position in terms of the prosodic (π-) gestural model (Byrd & Saltzman 2003; Byrd et al. 2005; Byrd et al. 2006) and suggesting phonetic explanations for the effect of order of constriction location. In sum, given that in the current study (i) there is so much variability across Spanish dialects and (ii) IV duration is sensitive to certain prosodic and segmental factors, the results support the phonetic nature of the vocalic elements that appear in /ɾC/ clusters. The implications of the current study for Spanish phonology, then, point toward viewing IVs in /ɾC/ clusters, as Romero (2008) states, ‘contextually conditioned articulatory transitions’ (61).

Future research should seek to define the apparent “tug-of-war” that exists between prosodic and segmental effects. That is, though certain segmental factors offered significant results, it is possible that prosodic factors might outweigh them. In terms of prosody, only the syllable and word boundaries are tested in the current study. In addition, a limitation of the current study is that not all consonants are tested and further research must address this by including all consonants in the inventory, listed in (2) in §2.2.

References

Araujo, Fernando. 1894. Estudios de Fonética Castellana. Toledo, Spain.


Appendix 1 Analysis by country of origin

<table>
<thead>
<tr>
<th>H1 Factor</th>
<th>Word-internal</th>
<th>Across word boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Word-internal vs. Across word boundary</td>
<td>Spain</td>
<td>113</td>
</tr>
<tr>
<td>Argentina</td>
<td>66</td>
<td>27.26**</td>
</tr>
<tr>
<td>Guatemala</td>
<td>45</td>
<td>32.99**</td>
</tr>
<tr>
<td>Colombia</td>
<td>55</td>
<td>35.00</td>
</tr>
<tr>
<td>Ecuador</td>
<td>78</td>
<td>36.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H2 Factor</th>
<th>Stressed</th>
<th>Unstressed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Stressed vs. Unstressed</td>
<td>Spain</td>
<td>86</td>
</tr>
<tr>
<td>Argentina</td>
<td>48</td>
<td>26.58**</td>
</tr>
<tr>
<td>Mexico</td>
<td>48</td>
<td>32.65</td>
</tr>
<tr>
<td>Guatemala</td>
<td>31</td>
<td>38.12</td>
</tr>
<tr>
<td>Colombia</td>
<td>41</td>
<td>36.98</td>
</tr>
<tr>
<td>Ecuador</td>
<td>58</td>
<td>36.35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H3 Factor</th>
<th>Back-to-front</th>
<th>Front-to-back</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Front to-back vs. Back-to-front</td>
<td>Spain</td>
<td>21</td>
</tr>
<tr>
<td>Argentina</td>
<td>11</td>
<td>26.41</td>
</tr>
<tr>
<td>Mexico</td>
<td>8</td>
<td>30.26</td>
</tr>
<tr>
<td>Guatemala</td>
<td>10</td>
<td>33.78</td>
</tr>
<tr>
<td>Colombia</td>
<td>14</td>
<td>25.63**</td>
</tr>
<tr>
<td>Ecuador</td>
<td>20</td>
<td>35.52</td>
</tr>
</tbody>
</table>

**=statistically significant with alpha set 0.05
* =marginally statistically significant with alpha set between 0.051 and 0.065
### H4: Prosodic and Segmental Effects on Vowel Intrusion Duration in Spanish /ɾC/ Clusters

<table>
<thead>
<tr>
<th>Factor</th>
<th>Heterorganic</th>
<th></th>
<th></th>
<th>Homorganic</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>Range</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Heterorganic vs. Homorganic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td>56</td>
<td>37.89</td>
<td>27.46–45.73</td>
<td>6.32</td>
<td>72</td>
<td>35.15</td>
</tr>
<tr>
<td>Argentina</td>
<td>35</td>
<td>27.53</td>
<td>22.71–33.02</td>
<td>4.28</td>
<td>43</td>
<td>29.52</td>
</tr>
<tr>
<td>Mexico</td>
<td>30</td>
<td>33.60</td>
<td>26.26–38.43</td>
<td>5.28</td>
<td>43</td>
<td>32.40</td>
</tr>
<tr>
<td>Guatemala</td>
<td>27</td>
<td>35.45</td>
<td>23.30–49.22</td>
<td>9.27</td>
<td>29</td>
<td>39.06</td>
</tr>
<tr>
<td>Colombia</td>
<td>31</td>
<td>34.78</td>
<td>33.00–36.33</td>
<td>1.29</td>
<td>38</td>
<td>36.08</td>
</tr>
<tr>
<td>Ecuador</td>
<td>45</td>
<td>36.61</td>
<td>29.54–41.35</td>
<td>4.93</td>
<td>47</td>
<td>36.77</td>
</tr>
</tbody>
</table>

### H5: Voice vs. Voiceless

<table>
<thead>
<tr>
<th>Factor</th>
<th>Voiced</th>
<th></th>
<th></th>
<th>Voiceless</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>Range</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Voiced vs. Voiceless</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td>53</td>
<td>36.70</td>
<td>29.60–46.30</td>
<td>5.77</td>
<td>65</td>
<td>36.16</td>
</tr>
<tr>
<td>Argentina</td>
<td>28</td>
<td>28.68</td>
<td>23.58–37.15</td>
<td>6.00</td>
<td>50</td>
<td>28.36</td>
</tr>
<tr>
<td>Mexico</td>
<td>25</td>
<td>32.75</td>
<td>28.83–39.60</td>
<td>4.97</td>
<td>48</td>
<td>33.43</td>
</tr>
<tr>
<td>Guatemala</td>
<td>19</td>
<td>38.62</td>
<td>29.15–45.76</td>
<td>7.03</td>
<td>37</td>
<td>38.83</td>
</tr>
<tr>
<td>Colombia</td>
<td>25</td>
<td>33.75</td>
<td>30.13–37.13</td>
<td>2.87</td>
<td>44</td>
<td>36.49</td>
</tr>
<tr>
<td>Ecuador</td>
<td>41</td>
<td>37.23</td>
<td>35.28–39.65</td>
<td>1.69</td>
<td>51</td>
<td>36.11</td>
</tr>
</tbody>
</table>

### H6: Stops vs. Nasals

<table>
<thead>
<tr>
<th>Factor</th>
<th>Stops</th>
<th></th>
<th></th>
<th>Nasals</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>Range</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>MOA</td>
<td>Spain</td>
<td>75</td>
<td>36.16</td>
<td>23.67–42.85</td>
<td>7.18</td>
<td>24</td>
</tr>
<tr>
<td>Argentina</td>
<td>50</td>
<td>28.36</td>
<td>21.84–32.86</td>
<td>4.76</td>
<td>10</td>
<td>25.18</td>
</tr>
<tr>
<td>Mexico</td>
<td>48</td>
<td>33.43</td>
<td>22.26–46.85</td>
<td>10.13</td>
<td>9</td>
<td>28.80</td>
</tr>
<tr>
<td>Guatemala</td>
<td>37</td>
<td>38.83</td>
<td>28.16–53.22</td>
<td>10.06</td>
<td>6</td>
<td>38.15</td>
</tr>
<tr>
<td>Colombia</td>
<td>44</td>
<td>36.49</td>
<td>32.84–41.49</td>
<td>3.99</td>
<td>9</td>
<td>28.25</td>
</tr>
<tr>
<td>Ecuador</td>
<td>51</td>
<td>36.11</td>
<td>27.29–43.13</td>
<td>6.01</td>
<td>19</td>
<td>31.18</td>
</tr>
</tbody>
</table>

** = statistically significant with alpha set 0.05
* = marginally statistically significant with alpha set between 0.051 and 0.065
## H6 Factor Approximants

<table>
<thead>
<tr>
<th>Factor</th>
<th>Spain</th>
<th>Argentina</th>
<th>Mexico</th>
<th>Guatemala</th>
<th>Colombia</th>
<th>Ecuador</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>29</td>
<td>18</td>
<td>18</td>
<td>13</td>
<td>15</td>
<td>22</td>
</tr>
<tr>
<td>Mean</td>
<td>38.45</td>
<td>30.68</td>
<td>32.84</td>
<td>38.68</td>
<td>36.60</td>
<td>42.44**</td>
</tr>
<tr>
<td>Range</td>
<td>31.92–42.96</td>
<td>27.27–37.74</td>
<td>28.88–39.60</td>
<td>32.43–43.00</td>
<td>30.13–41.00</td>
<td>38.16–48.55</td>
</tr>
<tr>
<td>SD</td>
<td>5.13</td>
<td>4.77</td>
<td>4.88</td>
<td>4.61</td>
<td>4.55</td>
<td>4.03</td>
</tr>
</tbody>
</table>

## H7 Factor Labials

<table>
<thead>
<tr>
<th>Factor</th>
<th>Spain</th>
<th>Argentina</th>
<th>Mexico</th>
<th>Guatemala</th>
<th>Colombia</th>
<th>Ecuador</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>21</td>
<td>11</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>Mean</td>
<td>33.49**</td>
<td>26.41</td>
<td>30.26</td>
<td>33.78</td>
<td>25.63**</td>
<td>35.52</td>
</tr>
<tr>
<td>Range</td>
<td>24.05–40.05</td>
<td>22.10–32.70</td>
<td>22.70–37.43</td>
<td>28.20–44.47</td>
<td>21.00–30.35</td>
<td>27.73–42.30</td>
</tr>
<tr>
<td>SD</td>
<td>6.39</td>
<td>4.60</td>
<td>7.37</td>
<td>7.37</td>
<td>3.96</td>
<td>6.51</td>
</tr>
</tbody>
</table>

## H7 Factor Coronals

<table>
<thead>
<tr>
<th>Factor</th>
<th>Spain</th>
<th>Argentina</th>
<th>Mexico</th>
<th>Guatemala</th>
<th>Colombia</th>
<th>Ecuador</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>21</td>
<td>11</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>Mean</td>
<td>42.32**</td>
<td>29.52</td>
<td>32.40</td>
<td>36.08**</td>
<td>36.77</td>
<td>36.77</td>
</tr>
<tr>
<td>SD</td>
<td>9.22</td>
<td>5.09</td>
<td>8.08</td>
<td>9.61</td>
<td>4.43</td>
<td>3.10</td>
</tr>
</tbody>
</table>

## H7 Factor Dorsals

<table>
<thead>
<tr>
<th>Factor</th>
<th>Spain</th>
<th>Argentina</th>
<th>Mexico</th>
<th>Guatemala</th>
<th>Colombia</th>
<th>Ecuador</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>35</td>
<td>24</td>
<td>22</td>
<td>16</td>
<td>21</td>
<td>28</td>
</tr>
<tr>
<td>Mean</td>
<td>40.59**</td>
<td>28.06</td>
<td>33.94</td>
<td>37.62</td>
<td>39.09**</td>
<td>37.21</td>
</tr>
<tr>
<td>SD</td>
<td>8.35</td>
<td>6.44</td>
<td>6.19</td>
<td>10.48</td>
<td>3.30</td>
<td>4.00</td>
</tr>
</tbody>
</table>

** = statistically significant with alpha set 0.05
* = marginally statistically significant with alpha set between 0.051 and 0.065
PART III

Interactions between segments and features
Acoustic and aerodynamic factors in the interaction of features
The case of nasality and voicing*

Maria-Josep Solé
Universitat Autònoma de Barcelona

This paper presents an account of the physical factors responsible for cross-linguistically common patterns of co-occurrence between values of the features [voice] and [nasal]. Specifically, it offers explanations for why nasals are typically voiced and why voiced obstruents are often accompanied by nasalization, or in terms of features, why [+voice] and [+nasal] co-occur so often and in such a variety of ways. First, it addresses the acoustic-auditory factors responsible for glottal vibration favoring the perceptibility of nasalization. Second, it examines the aerodynamic factors responsible for nasality facilitating glottal vibration. In particular, it suggests that nasal leakage is a maneuver to facilitate voicing in the stop and to preserve the voicing contrast. The paper also argues that if the interaction between the two features can be explained by phonetic principles, then there is no need to encode the patterns of co-occurrence as redundancy rules or constraints in universal grammar. Furthermore, phonological representations that assign the nasal valve and the larynx to separate nodes cannot capture the interaction between nasality and voicing and the co-occurrence patterns.

1. Introduction

This paper addresses the dependency relations of features within a segment and when segments follow one another. Specifically, it focuses on the dependency between nasalization and voicing, that is, the likelihood that the two features combine into segments and that they occur in contiguous segments. We review experimental data and sound patterns illustrating the interaction of the two features and present an account of the physical factors responsible for such interaction.

*This study was funded by a grant HUM2005-02746 from the Ministry of Science and Technology, Spain, and by the research group SGR 2005-2008 of the Catalan Government. The comments of an anonymous reviewer are gratefully acknowledged.
We argue that abstract feature specifications, e.g., [+nasal], [–voice], or [+cont], devoid of detailed phonetic content cannot adequately account for how features combine into segments and how they affect each other when they occur in contiguous segments, for example, in context-dependent phonological processes or sound change. In particular, we argue that restrictions on the combination of features, in particular the features [nasal] and [voice], are partly determined by phonetic factors.

Phonological approaches within generative linguistics, and Articulatory Phonology, view independent articulators, such as the oro-nasal valve, the laryngeal valve and the oral articulators as belonging to different nodes or tiers. Since velic opening and closing is generally independent of the movements of the oral articulators and the action of the vocal folds, different velum positions (open/closed) can in principle occur simultaneous to different articulatory constrictions and glottal states. All the logical possibilities of combinations of feature specifications, however, do not occur in languages, nor are all features equally likely to combine. There seem to be dependency relations between independently controlled gestures which fail to be captured by formal representations that separate features by assigning them to different nodes, as convincingly argued by Ohala (2005). Such dependency relations are partly due to physical (e.g., aerodynamic or biomechanical) interactions between articulatory gestures, or the acoustic-auditory consequences of such interactions. Prior work on phonetically-based restrictions on the combination of features includes the aerodynamic voicing constraint (Ohala 1983; Westbury & Keating 1986) that accounts for the difficulty involved in maintaining voicing during an obstruent, particularly a long or a back articulated obstruent; the dependency between voicelessness and frication, by which sonorants and high vowels become fricatives when devoiced (Ohala 1983, 1997), and the perceptual consequences on vowel height of vowel nasalization (such as those giving rise to the lowering of nasal vowels in French, e.g., *fin* [fɛ̃] “the end” vs. *finir* [fɛ̃iʁ] “to finish”; *brun* [brœ̃] “brown” vs *brune* [bryn] “brown-haired” (fem.); Beddor, Krakow & Goldstein 1986), amongst others.

Focusing on nasalization, Ohala & Ohala (1993) provide a thorough account of the aerodynamic and acoustic factors underlying how nasality interacts with other features, in particular manner and place features. In the following sections we will examine the phonetic basis for the interaction between nasality and voicing in the light of recent research and will argue that what happens at the glottis affects nasal coupling and perceptibility of nasalization and what happens at the velopharyngeal port influences the continuation or extinction of voicing.

The dependency between nasalization and voicing, or in terms of features, why [+nasal] and [+voice] co-occur so often and in such a variety of ways, has received considerable attention in the phonological literature. For example, dependency
relations between the two features within segments, such as the tendency for nasals to be voiced, have been accounted for by ‘redundancy rules’ (such as ‘nasals are redundantly voiced’; Jakobson, Fant & Halle 1951:43) and Optimality Theory constraints (e.g., NAS/VOI: A nasal must be voiced (Itô, Mester & Padgett 1995); Not Co-occurring (+sonorant, –voiced): Sonorants must be voiced (Stemberger & Bernhard 1999)). Similarly, the dependency relations between nasality and voicing across segments – such as postnasal voicing and prenasalization of voiced but not voiceless obstruents – as been captured by the ‘IDENTICAL CLUSTER CONSTRAINT [VOICE]: a sequence of consonants must be identical in voicing’ (Pulleyblank 1997:64, 69ff) and the more specific *N∁ constraint which disallows nasal-voiceless obstruent clusters (Pater 1999). Possibly the most extreme claim of the dependency between nasality and voicing is made by Nasukawa (2005), who within an Element Theory approach – which posits a restricted set of prime elements capturing the properties of both consonants and vowels – proposes a single nasal-voice category, {N}, which may be manifested as either nasality or voicing depending on its headship status. In a non-Element Theory analysis, the choice of the prime element {N} over ‘voicing’ suggests that voicing is predictable from the nasal feature specification of the segment.

The purpose of the present paper is to provide an account of the physical factors responsible for cross-linguistically common patterns of co-occurrence between nasality and voicing and to argue that if the co-occurrence between the two features arises due to the physics, physiology and perception of speech and can therefore be explained functionally, then it does not need to be explained by redundancy rules or constraints in universal grammar. In the following sections we first review the acoustic-auditory factors which underlie the increased perceptibility of nasalization with a periodic sound source and which are at the origin of nasal and nasalized sounds being largely voiced. Second, we examine the aerodynamic factors responsible for nasality favoring glottal vibration and the phonological patterns that can be derived from this principle. We propose that nasal leakage is a maneuver to facilitate voicing in the stop and, ultimately, to preserve the voicing contrast. Finally, we argue that formal phonological notations which represent the nasal valve and the larynx at different nodes fail to capture the interaction between nasality and voicing.

2. Voicing favors nasality

Acoustic-auditory reasons are at the origin of nasal and nasalized segments being predominantly voiced. It is known that the sound source (glottal pulsing or turbulence) mostly excites the cavities anterior to the constriction where the sound
is generated, which contribute to resonances and antiresonances, whereas the back cavities do not contribute much acoustically (Fant 1960; Stevens 1998). The different nature and location of the sound source for voiced and voiceless nasals accounts for their different acoustic result. As illustrated in Figure 1, in voiced nasals, glottal pulses excite the oral and nasal cavities, which contribute the low frequency resonances and the characteristic nasal zeros.

For voiceless nasals (as for voiceless vowels) there is no glottal excitation of the oral and nasal cavities and, hence, no low frequency resonances; there is some turbulence generated at the vocal folds which is selectively amplified in the oropharyngeal and nasal cavities, but by virtue of this turbulence not having low frequency periodic components, the low frequency nasal murmur and resonances characteristic of nasals are missing. The large volume of air flowing through the open glottis during the articulation of the voiceless nasal gives rise to turbulence generated primarily at the nostrils (the point of maximum constriction) (see Figure 1, right), whatever the place of articulation of the nasal. Since the oral and nasal cavities posterior to the nostrils do not contribute resonances, the different nasals [m n n] do not differ much spectrally (except in the transitions in adjacent vowels), and because there is no downstream cavity to amplify the frication at the nostrils, voiceless nasals have a weak intensity (Ohala 1975; Ohala & Ohala 1993). As a result, voiceless nasals do not have the auditory cues associated with nasal coupling and form a distinct sound from voiced nasals.

1. The back cavities, however, may be excited under certain circumstances. For example, in sounds involving a large glottal abduction, as in breathy voice or in voiceless fricatives, the subglottal cavity is excited and subglottal resonances interact with the supraglottal resonances and create by this interaction antiresonances which exert a detectable influence on the speech spectrum (Stevens 1998:196–8).
This is illustrated in Figure 2, which shows distinctive voiceless nasals, [m n ṇ ŋ ŋ], and voiced nasals, [m n ṇ ŋ], in Burmese. Voiceless nasals lack the low frequency resonances characteristic of nasal coupling, and the resulting frication has a low intensity and indistinct spectral characteristics, which makes them nonoptimal sounds auditorily and thus rarely used in languages.\(^2\) Note that the voiceless nasals in Figure 2 are, in fact, phonetic sequences of voiceless nasal + voiced nasal or voiceless nasals with a voiced offglide – e.g.[ŋn] (Dantsuji 1984; Ladefoged & Maddieson 1996:113). As noted, place distinctions in voiceless nasals are obscure, but many languages that have phonemic voiceless nasals have them at more than one place of articulation. Thus, they are frequently produced with a brief voiced portion. In this way different places of articulation can be differentiated – by both the distinctive resonances of the voiced nasal and the transitions in adjacent vowels.

![Spectrogram](image)

**Figure 2.** Spectrogram (0–5kHz) for Burmese /m aʔ/ “from”, /ŋə/ “nose”, /ŋə/ “considerate”, /ŋə/ “borrow” (top), and /maʔ/ “lift up”, /nə/ “pain”, /ŋə / “right”, /ŋə/ “fish” (bottom)

There is not a great deal of perceptual data on voiceless nasals, but the available data indicate that nasality and place are more difficult to detect in a voiceless than a voiced nasal/nasalized vowel. For example, Arai (2006) found that when

---

2. The same is true of voiceless nasal vowels (with low acoustic intensity and altered spectral properties; Blankenship 1997) which have not been reported to be contrastive in languages (Ladefoged & Maddieson 1996:315).
a higher amplitude of aspiration (hiss noise) was added to nasalized vowels, perceived vowel nasality decreased. Upon examining voiceless nasals in Burmese, Dantsuji (1986) found no significant cues for place of articulation in the voiceless portion of the nasal3 – while the voiced murmur portion contains cues to convey place distinctions – which would suggest that place distinctions are harder to detect in a voiceless than a voiced nasal, hence the common voiced portion of phonologically voiceless nasals.

We have seen that the acoustic-auditory consequences of a lowered velum are not uniform for voiced and voiceless segments; in other words, the perceptibility of nasalization is dependent on the voicing specification of the segment. Nasalized fricatives provide another example showing that glottal vibration favors the perception of nasalization. The aerodynamic difficulty involved in concurrent nasalization and frication has been addressed in a number of studies (Ohala 1975; Ohala & Ohala 1993; Ohala, Solé & Ying 1998; Shosted 2006; Solé 2007). A lowered velum for nasality vents the air through the nasal cavity, lowers the oral pressure ($P_o$) and thus reduces the high rate of flow necessary to generate audible frication at the oral constriction. Nasalized fricatives, however, have been reported to occur in languages and it has been observed that voiced nasalized fricatives tend to retain nasalization and lose their friction. For example, voiced nasalized fricatives are phonetically nasalized frictionless continuants (e.g., Waffa /内科 [内科], Stringer & Hotz 1973; Umbundu /内科/ [内科], Schadeberg 1982), and voiced fricatives tend to lose their friction due to spreading nasalization and become nasalized approximants (e.g. [内科] ~ [内科内科] in Guaraní, Gregores & Suarez 1967). In contrast, voiceless nasalized fricatives tend to retain frication and lose nasality (Cohn 1993; Ladefoged & Maddieson 1996:132). Thus two different processes need to be explained: loss of friction in voiced but not voiceless nasalized fricatives, and loss of nasalization in voiceless but not voiced fricatives. Aerodynamic factors may account for the former process while acoustic factors account for the latter.

Due to aerodynamic factors, for the same degree of velopharyngeal opening, frication is more severely impaired in voiced than in voiceless fricatives. This is so because voiced nasalized fricatives have two additional mechanisms, other than nasal venting, that impair strong frication (a function of air speed or particle velocity past a constriction which is dependent on oral pressure): increased glottal resistance – which results in a lower oral pressure and inhibits the air vented through the nasal passage from being resupplied from the lungs (as is the case for

---

3. Maddieson (1983), however, reports that spectral differences may be found during the voiceless portion, e.g., greater relative amplitude in the lower frequencies for labials than for nasals at other places of articulation.
voiceless fricatives with an open glottis) – and the need to keep oral pressure low for voicing. Thus, Ohala, Solé & Ying (1998) report that when voiced and voiceless fricatives are vented with a pseudo-velopharyngeal valve – a tube inserted at the sides of the mouth via the buccal sulcus and the gap behind the molars – simulating different degrees of nasalization, when the valve has a similar impedance to that at the oral constriction (and as a result air flows out through both the nose and the mouth) voiced fricatives become frictionless continuants\(^4\) while voiceless fricatives retain their frication (though the intensity of friction is attenuated). In sum, the combination of high resistance at the glottis and lower oral pressure for voicing, and hence a lower rate of flow and particle velocity through the oral constriction (monotonically related to the intensity of turbulence), accounts for why voicing impairs audible frication.

Why nasality is retained in voiced as opposed to voiceless fricatives may have an acoustic explanation. For the reasons stated earlier, nasalization contributes more acoustically to voiced than to voiceless fricatives. Glottal vibration for voiced fricatives resonates in the nasal cavity – as well as in the oral cavity – thus adding perceptible acoustic properties of nasal coupling (intense low frequency murmur, spectral zeros, and increased F1 bandwidth in neighboring vowels) to the weak (or nonexistent) frication. In contrast, in voiceless fricatives the sound source is forward of the velopharyngeal opening (except for glottal and pharyngeal fricatives)\(^5\) and the sound generated mostly excites the anterior cavity with little acoustic coupling with the posterior nasal cavity. As a consequence, nasalized voiceless fricatives with audible frication do not differ much auditorily from non-nasalized fricatives, that is, the acoustic cues for nasalization are hardly detectable (Cohn 1993; Ladefoged & Maddieson 1996). In sum, for acoustic-auditory reasons glottal vibration favors the percept of nasality. Such acoustic reasons account for a number of cross-linguistic sound patterns,\(^6\) such as nasal consonants, nasalized

\(^4\) The aerodynamic explanation for voiced fricatives with a nasal leak becoming frictionless continuants is in line with historical data showing that voiced fricatives coarticulated with following nasals tend to weaken to approximants, taps or nasals, [zn] > [jn], [rn], [nn], or are lost altogether (Solé 2007).

\(^5\) Glottal and pharyngeal fricatives and stops, for which the build-up of pressure takes place further upstream than the velic valve, with the result that a lowered velum would not affect the pressure build-up, can be nasalized. Nasalized glottal fricatives, i.e. /h/, have been widely reported in languages (Ladefoged & Maddieson 1996), and they occur phonetically in American English, e.g., *home* [həʊm].

\(^6\) Rhinoglottophilia (Matisoff 1975:265) is a marginally related phenomenon. Matisoff noted the dependency relations between nasality and laryngeal action and coined the term ‘rhinoglottophilia’ to describe the behavior of vowels becoming nasalized in the context of a
sonorants and nasalized vowels being predominantly voiced, and the different result of voiced and voiceless nasalized fricatives.

3. Nasality facilitates voicing

Due to aerodynamic reasons nasality favors voicing in neighboring obstruents. It is well known that when two segments are in contact their articulations necessarily overlap. Coarticulation with neighboring segments may cause modifications of the aerodynamic conditions in the vocal tract which can, in turn, affect the acoustic and auditory result. The interaction between nasality and voicing in consecutive segments is illustrated in postnasal voicing. When a voiceless stop is preceded by a nasal, voicing into the stop closure is prolonged, vis-à-vis postvocalic stops, by nasal leakage before full velic closure is achieved and continued velic raising even after velic closure has occurred, thus expanding the volume of the oral cavity. Nasal leakage and oral cavity expansion lower the oral pressure which accumulates in the oral cavity and thus prolong transglottal flow for voicing (Rothenberg 1968; Westbury 1983; Ohala & Ohala 1991; Bell-Berti 1993; Hayes & Stivers 2000). These factors lead to postnasal voiceless obstruents being phonetically partially voiced and with a weaker stop burst (which may in turn lead to their being reinterpreted as voiced). Such phonetic effects have phonological significance (i) in languages with a phonological postnasal voicing rule (Japanese, Zoque, Kikuyu, and Modern Greek dialects amongst others; see Hayes & Stivers (2000) and references therein), by which voiceless obstruents become voiced after nasals, as illustrated in (1) for Japanese; (ii) in phonological alternations between voiceless stops and prenasalized voiced stops (e.g. Terena, where nasalization is affixed at the beginning of the word and spreads until an obstruent blocks it, and the obstruent becomes voiced in the process, see (2)); (iii) in progressive voicing assimilation in stops following laryngeal consonant, such as /h/, in Austronesian languages, Indo-Aryan languages, and Thai, amongst others. Such spontaneous nasalization of vowels adjacent to segments requiring wide laryngeal abduction, such as /h/, voiceless fricatives and aspirated stops, has been accounted for by Ohala & Ohala (1993) in the following terms. The open glottis required for high airflow segments extends to the margins of adjacent vowels. The coupling of the oral and subglottal cavities during glottal abdution adds spectral effects to the vowels that resemble nasalization (i.e. the coupling of the oral and nasal cavities) – increased F1 bandwidth, decreased intensity in the higher frequencies – and may be interpreted as intended nasalization. Though this process results from misinterpretation of the acoustic effects of an added resonator (the subglottal or nasal cavity), rather than laryngeal modulation facilitating the perception of nasality as in the cases of interest, it illustrates the role of acoustic-auditory factors in phonological patterns.
nasals, see (3); and (iv) in historical sound change, for example, in the development from Classical Armenian to the Armenian language of New Julfa, exemplified in (4). Postnasal voicing is also reported in infant phonological acquisition of English where the voiced realization of initial voiceless stops is always preceded by a nasal, as illustrated in (5).

(1) Japanese postnasal voicing (Itô, Mester & Padgett 1995)

root + root
fumu + kiru funçiru “give up”
word internally
unzari “disgusted” *unsari
tombo “dragonfly” *tampo

(2) Terena alternations between voiceless stops–prenasalized voiced stops
paho “I went” mbiho “he went”
iso “I hoed” iñzo “he hoed”
owoku “my house” ōwō̂̀gu “his house”

(3) Progressive voicing assimilation in nasal+stop clusters (Rohlfs 1949:88–89; Rohlfs 1970)

Southern Italian santo [ˈsanda] “saint”
pampano [ˈpambano] “hopscotch”
bianco [ˈbijano] “white”
Gascon [kanˈda] “to buy” from Lat. cantare

(4) Historical change (Vaux 1998:506)

Classical Armenian New Julfa
ənkanel ənganiel “fall”
ajnteļ əndie̞ɛ “there”
[tantʃ] tjandо̞ “fly”

(5) Child phonology (Seth between ages 1;10 and 2;2; Kager, van der Feest, Fikkert, Kerkhoff & Zamuner 2007)

[ɔnɗa̞nə] for “sun tan”
[n ɗaˈaf] for “N turn off” (where the nasal is apparently a realization of ‘want’)

The interaction between nasality and voicing, however, is not restricted to postnasal voicing. More generally, there is a dependency between contiguous nasals and voiced stops. In other words, there is a bias toward nasals occurring, being preserved or emerging next to voiced but not voiceless stops; symmetrically, voiced but not voiceless stops tend to be preserved in a nasal context. The former is
illustrated in the following sound patterns. Nasals occur before voiced but not before voiceless stops (nasal deletion) in the Kelantan dialect of Malay (Teoh 1988), and in a number of African languages, such as Venda, Swahili and Maore (cited in Pater 1999:319). Nasals are preserved before voiced but not before voiceless stops (denasalization), as illustrated in (6a) for Mandar, an Austronesian language spoken in parts of Indonesia. A similar process is found in American English whereby nasals are lost before tautosyllabic voiceless stops but not before voiced stops, see (6b). In Hindi, nasals emerge between a nasalized vowel and a voiced but not a voiceless segment, see (7).

(6) Preservation of nasals before voiced but not voiceless stops
   a. Mandar (Mills 1975)
      /maN+tunu/ mattunu “to burn”
      /maN+dundu/ mandundu “to drink”
   b. American English nasal loss⁸ (Malécot 1960)
      tent /tent/ [tʰɛ̃t] can’t /kʰɛ̃t] camp /kamp/ [kʰɛp]
      tend /tend/ [tʰɛ̃d] canned /kænd/ [kʰɛ̃d]

(7) Nasal epenthesis, e.g., Hindi (Ohala & Ohala 1991)
   Sanskrit   Old Hindi   Modern Hindi
   čandra     čhādra     [ʧʌnd] “moon”
   danta.     dâ:ta       [dât] “tooth”

7. A related case may be nasal deletion and vowel lengthening before fricatives (e.g., Proto-Greek *pans > Ancient Gk. pa:s). The loss of nasals with concomitant vowel lengthening, especially before fricatives, documented in a variety of languages (see Ohala & Busà 1995), has been explained in the following terms. Due to the aerodynamic requirements for frication, nasal consonants are shorter and the preceding vowel more extensively nasalized in VNFricative than in VNStop sequences, as shown by phonetic data (e.g., Busà 2007). A shorter nasal may be more difficult to detect, resulting in nasal loss, and the reported perceptual association between vowel length and nasalization (Whalen & Beddor 1989) may explain vowel lengthening. The sound change may involve an additional perceptual component. Perceptual data shows that nasals are harder to detect before voiceless fricatives than before other consonants, most likely because these segments are produced with a large glottal opening, spreading through coarticulation to adjacent vowels, which creates acoustic effects on the vowels similar to nasalization. Such acoustic effects resembling nasalization are attributed to the large glottal abduction and thus factored out. Listeners may factor out nasals actually occurring in this context and sound change may result (Ohala & Busà 1995; Busà 2007).

8. One reviewer suggests that rather than the nasal being lost before a voiceless stop in these examples, the voiceless obstruent shortens the preceding sonorant interval such that the nasal consonant portion is truncated, while lack of shortening of that interval before voiced obstruents preserves the nasal consonant. In other words, it is not voicing per se that determines how much of a nasal is realized, but rather the effect of voicing on the
Examples of the latter – i.e. the preservation of voiced stops but the loss or replacement of voiceless stops in a nasal context – are presented below. In Indonesian (Halle & Clements 1983) and in OshiKwanyama (Steinbergs 1985; spoken in Angola and Namibia), root-initial voiced stops are preserved after a nasal, but voiceless stops become a homorganic nasal (nasal substitution), illustrated in (8a) for Indonesian. This process is replicated in American English in the assimilation of nasality in /nt/ – but not /nd/ – clusters when they occur between a stressed and an unstressed vowel, resulting in “winter” and “winner” being pronounced the same, see (8b). In German and many dialects of English, a /t/ is realized as a glottal stop or irregular glottal pulsing when followed by a nasal, as exemplified in (9), whereas /d/ is preserved in the same context. In such contexts, the voiceless stop would be nasally released and would lack the strong fricative release burst, which is a perceptual cue for voiceless stops (Ali, Daniloff & Hammarberg 1979), whereas a glottal stop (with a constriction and build-up of pressure further upstream than the velic opening) allows velic lowering while showing a discontinuity in amplitude and a release burst characteristic of a stop (Kohler 2001). Along the same lines, a tendency for /t/ to be more prone to deletion than /d/ before a nasal (e.g., *sweeten* vs. *Sweden*) in American English, due to the lack of a release burst in this environment, is reported by Zue & Laferriere (1979) (see also Raymond, Dautricourt & Hume (2006) on /t, d/ deletion in spontaneous American English).

(8) Assimilation of nasality
a. Indonesian (Halle & Clements 1983)
   /məN+boli/   [məmboli] “to buy”
   /məN+dapat/   [mandapat] “to get, to receive”
   /məN+ganti/   [məŋganti] “to change”
   BUT:
   /məN+pilih/   [məmilih] “to choose, to vote”
   /məN+tulis/   [mənulis] “to write”
   /məN+kasih/   [məŋasih] “to give”

b. American English
   *center* [nn] vs *sender* [nd]
   *international* [nn] vs *indicational* [nd]
The reviewed sound patterns exhibit a bi-directional dependency between contiguous nasals and voiced stops. The dependency is bi-directional (or symmetrical) in the sense that, in some cases, the conditions for nasal preservation are created by the voicing of the stop, and in other cases, a nasal context determines the preservation or loss of the stop depending on its voicing specification. Such symmetrical relations in adjacent sounds are not common and indicate the intricate nature of the interaction.

3.1 Explanations for the sound patterns

The basis for the interaction between nasalization and voicing may be found in phonetic factors. Phonetically, voiced obstruents exhibit more nasal leakage preceding and following nasalized vowels and nasal consonants than voiceless obstruents. Cohn’s (1990:108,199) nasal flow data for French and Sundanese ṼC and VNC sequences shows that nasal flow is present during most of the duration of voiced stops – the soft palate is raised just before the stop is released, to ensure that the release is oral – whereas nasal flow drops abruptly at the onset of voiceless stops. Huffman’s (1990:61, 65) data shows velic leakage (i.e. nasal airflow) during the closure of voiced but not voiceless stops in C̃V and ṼC sequences in Yoruba. Ohala & Ohala’s (1991) nasal pressure data for Hindi and French ̃VC sequences exhibits velic lowering during the oral closure of voiced stops (essentially prenasalized stops) but much shorter or non-occurring nasal leakage for voiceless stops. Basset, Amelot, Vaissière & Roubeau’s (2001) nasal airflow data for French shows that voiced stops and fricatives preceding and following contrastive nasal vowels (C̃V, ̃VC) showed significantly more cases of anticipatory and carryover nasalization (78% of the cases) than voiceless obstruents (34%), and a longer temporal extent of velum lowering (throughout the duration of the voiced obstruent vs. half of sonorant is preserved. Third, Beddor (2007) notes that extensively nasalized vowels covary with short or absent nasals in VNC sequences in American English, and suggests perceptual equivalence between Ṽ and N. Taken together, these data suggest that the case of nasal loss exemplified in (6b) is not the effect of the obstruent voicing on the preceding sonorant’s duration. Rather, the early onset and offset of the nasal gesture in VNC voiceless sequences (Beddor 2007) is probably an effect of the low tolerance of voiceless, but not voiced, stops to coarticulatory nasalization that might threaten their voiceless percept (Ohala & Ohala 1991).
the duration of the voiceless obstruent). Further evidence of voiced stops, but not voiceless stops, showing nasal leakage is reviewed in Ohala & Ohala (1991:213).

These data show that the velum may be lowered during the first part of the voiced stop but may close before the release so as to produce an oral burst, or the velum may be lowered throughout the stop, which is likely to result in a relatively weak burst.9 As noted, voiceless stops tend to inhibit coarticulatory velic lowering during the stop constriction. Partial or incomplete velopharyngeal closure may prevent the build-up of intraoral pressure necessary for a noisy release of a stop. Such nasal leakage would have a larger perceptual effect on voiceless than on voiced stops, as high intensity noise is a perceptual cue for voiceless stops (Ali, Daniloff & Hammarberg 1979).

Ohala & Ohala (1991) provide an acoustic-auditory explanation for voiceless stops having less tolerance for nasalization than voiced stops in terms of coarticulatory nasalization undermining the stop or voiceless character (i.e. the spectral and amplitude discontinuity, and noisy release burst) of voiceless stops, while voiced stops can meet their auditory requirements with a partially lowered velum. On the basis of the lower tolerance of voiceless stops to coarticulatory nasalization, Pater (1999) analyzes some of the cases in (1)–(9) as resulting from a phonetically motivated constraint against NC clusters. Thus, the *NC constraint bans nasals from occurring before voiceless segments, and postnasal voicing (Examples 1–5 above), nasal deletion, denasalization (Example 6), and nasal substitution (Example 8) are ‘repairs’ to eliminate disallowed NC clusters.

3.2 Further interactions between nasalization and voicing

Ohala & Ohala’s (1991) explanation in acoustic-auditory terms (and the *NC constraint) may in part account for why voiceless obstruents, in order to preserve the segment’s integrity, tend to inhibit coarticulatory nasalization more than voiced obstruents, and hence for the different fate of nasals in a voiced or a voiceless context (Examples 6, 7 above), and the loss of buccal voiceless, but not voiced, stops in a nasal context (Examples 8, 9). However, it does not account for prenasalization of voiced stops in an oral context, the emergence of non-etymological nasals adjacent to voiced but not voiceless consonants, and maintenance of the voicing contrast only in a nasal context. Indeed, languages with distinctive voiceless stops, [p t k], and prenasalized voiced stops [ⁿb, ⁿd, ⁿg], but no simple voiced stops – such as languages in Austronesia, Papua and South America (Maddieson & Ladefoged 1993:256), for example, Waris, illustrated in (10) – suggest that such prenasalized

9. The same velopharyngeal timing patterns have been observed in contrastive prenasalized voiced stops (Henton, Ladefoged & Maddieson 1992; Ladefoged & Maddieson 1996).
stops form the voiced stop series (Maddieson 1984:67). In fact, prenasalized stops have been analysed as plain voiced stops in some varieties of Mixtec (Piggot 1992; Iverson & Salmons 1996). These languages, as well as languages with contrastive voiced and voiceless stops but with the voiced series being optionally phonetically prenasalized (e.g., Bola, exemplified in (11); Tok Pisin, Smith 2002) or being realized as nasals in certain contexts (e.g., Rotokas, Hyman, to appear) suggest that nasal leakage is utilized to facilitate voicing in the obstruent. Note that, as opposed to Examples (1) to (9), these cases neither contained a nasal etymologically nor occur in a nasal context, and the prenasalized stop, therefore, cannot simply be the result of the ‘preservation’ of historical traces or coarticulatory nasalization.

(10) Contrastive voiceless stops and prenasalized stops, Waris, Papua New Guinea (Brown 2001)

[p] panda “pitpit type”    [m] banda “snake”
  nopo “eye”             tombol “dry”
[t] tata “meat”          [d] damba “tree sp.”
  lot “banana type”       wand “pitpit grass”
[k] kao “tree sp.”       [g] gao “go!”
  okala “distant”         engala “hand”

(11) Optional prenasalized stops, Bola, Malayo-Polynesian (Wiebe 1997)

[b] ~ [m] bahele “crocodile”
  bebe “butterfly”
[d] ~ [d] dagi “dig”
  made “sit”
[g] ~ [g] ge “3rd pers. FEM”
  aga “canoe”

The suggestion that nasal leakage is used to favor voicing in the stop is consistent with the finding that in the 19 languages in the UPSID database with prenasalized segments, they are all voiced obstruents and are overwhelmingly stops (Maddieson 1984:67). Instrumental data of nasal flow during a voiced stop is provided in Figure 3 for Karitiana, a Tupi language spoken in Brazil, which lacks phonemic voiced stops but features phonetic voiced stops as allophones of nasals word-initially (Demolin 2007). Figure 3 shows that the stop is voiced throughout – though evidence of voicing during the stop is hardly observable in the audio signal, note the glottal pulses in the oral pressure trace (channel 2) – and that there is nasal flow preceding the stop constriction (beginning halfway through the segment when the oral pressure starts to rise) and to a lesser degree, during the stop closure.

Prenasalization (and postnasalization) of voiced stops, but not voiceless stops, in the absence of contextual nasals (‘spontaneous’ nasalization) is also found in data from first and second language acquisition. Although child phonology generally shows simplification of consonant clusters, various authors report cases where English-learning infants add a nasal before voiced stops (Clark & Bowerman
Acoustic and aerodynamic factors in nasality and voicing

1986) or after voiced stops (Labov & Labov 1978), as shown in the realizations in (12), in order to facilitate voicing in coda stops. Insertion of epenthetic nasals before initial voiced stops has been reported for French infants (Allen 1985). Along similar lines, Kong, Beckman & Edwards (2007) attribute the early mastery of voiced stops in Greek-learning infants to prenasalization of initial voiced stops (illustrated in (12)), though in this case prenasalization is not ‘spontaneous’ but rather a variant pronunciation of voiced stops in Standard and dialectal Greek. Adult American English learners of Spanish (a language with a voicing contrast between prevoiced and unaspirated stops) may show prenasalization of initial voiced stops in Spanish, exemplified in (13), to ensure prevoicing during the consonant constriction. Lewis (in press) reports that speakers using this strategy showed longer prevoicing in production of Spanish /b/ and their tokens were the least likely to be misperceived as /p/ by native Spanish listeners.

Evidence of nasalization during voiced stops has been observed by a variety of investigators. For example, Pape, Mooshammer, Hoole & Fuchs (2003) report that German speakers may use prenasalization to avoid devoicing of stops. Velopharyngeal opening has been found during utterance-initial and intervocalic breathy voiced stops for Hindi and Telugu (Rothenberg 1968:7.4) and for Sindhi voiced stops (Nihalani 1975). Taken together, these data suggest that speakers make use of nasal leakage as a strategy to achieve voicing during the stop consonant.

(12) Child phonology

<table>
<thead>
<tr>
<th>Language</th>
<th>Word</th>
<th>Realization</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>bent</td>
<td>[bent]</td>
</tr>
<tr>
<td>English</td>
<td>pig</td>
<td>[pɪŋk]</td>
</tr>
<tr>
<td>English</td>
<td>dad</td>
<td>[dæd]</td>
</tr>
<tr>
<td>Greek</td>
<td>gol</td>
<td>[ɡoːl]</td>
</tr>
</tbody>
</table>

Audio (channel 1), oral pressure (channel 2) and nasal flow (channel 3) for the Karitiana word [’bap’]. The prenasalized voiced stop appears between vertical lines. (From Storto & Demolin in press)
Cases of postnasalization of stops have also been reported. For example, Jones (2001) examines the spontaneous postnasalization of voiced stops in Lancashire dialects of English, documented in Wright (1952) and Orton & Halliday (1963). Postnasalization involves the emergence of a homorganic nasal at the end of voiced stops and voiced affricates phrase-finally, as illustrated in (14), and also before phrase-final plurals (e.g., [lɛŋz] for “legs”). The fact that the nasal offglide is homorganic with the stop reflects that the velum is lowered during the latter part of the stop closure, resulting in a nasal release. Jones rightly suggests that postnasalization results from a strategy to prolong voicing during the stop constriction, by venting the air through the nose, in order to maintain the voicing contrast phrase-finally.\footnote{Jones notes that this interpretation does not account for voiceless affricates also being postnasalized in phrase-final position, e.g., in March [tʃm]. \textbf{He suggests that postnasalization in voiceless affricates may reflect an attempt to keep them distinct from heavily aspirated voiceless stops phrase-finally. Indeed, the fact that, differently from nasals following voiced stops and voiced affricates, the nasal segment following the voiceless affricate [tʃ] is not homorganic suggests that velum lowering is not present during the affricate constriction and is therefore triggered by a different mechanism.}} That is, as the voicing contrast in stops is endangered in phrase-final position due to passive devoicing, velic lowering during the voiced stop helps to prolong voicing and preserves the voicing contrast. Thus in some Lancashire dialects voicing in final stops covaries with (and is cued by) postnasalization, possibly in addition to preceding vowel duration as in most dialects of English. Similar nasal releases to final voiced (but not voiceless) stops occur in the South American languages Kaingang and Içua Tupí (Herbert 1986:206, cited in Laver 1994:232), in Vietnamese dialects, and Senegalese Wolof (Ward 1939, cited in Jones 2001).

(14) Postnasalization of phrase-final voiced stops in Lancashire dialects of English (Jones 2001)
[ˈkɔːv ə ɔt ˈlɛŋ] calf of thy leg
[ˈuz ˈwedn] she’s wed (married)
[ˈspt ə ˈɡbm] spit a gob (phlegm)
[ˈkabdʒn] cabbage

It is of interest to note the different patterning of postnasalization and prenasalization in the languages reviewed. Postnasalization, [b̪m d̪n g̪ŋ], is associated with phrase-final position, where voicing is difficult to maintain due to the increase
in oral pressure during the stop closure while subglottal pressure (Ps) decreases phrase-finally (Westbury & Keating 1986; Slifka 2000). Conversely, prenasalization, \([\text{m} \text{b} \text{n} \text{d} \text{g}]\), tends to occur in word-initial and therefore utterance-initial position, where phonation is more difficult to initiate due to the lowered subglottal pressure utterance-initially and the larger pressure difference required to set the vocal folds into vibration \(\text{P}_{\text{subglottal}} - \text{P}_{\text{oral}} > 3\text{–}4\text{cmH}_2\text{O}\) vs. \(1\text{–}2\text{cmH}_2\text{O}\), respectively, Baer 1975). These patterns of post- and pre-nasalization are consistent with the interpretation that nasal venting is used to facilitate the required transglottal pressure difference in order to prolong or initiate vocal fold vibration during the stop closure.

Crucially, crosslinguistic data suggests that coarticulatory (or spontaneous) velum leakage facilitates voicing in obstruents, thus helping to preserve the voicing contrast. A case of preservation of the voicing contrast exclusively in a nasal environment is found in Basaa, a Bantu language spoken in South West Cameroon. In Basaa there is a contrast between voiceless stops \([\text{p} \text{t} \text{k} \text{kw}]\), voiced stops \([\text{b} \text{d} \text{g} \text{gw}]\), and prenasalized stops \([\text{mb} \text{nd} \text{ng}]\). The contrast between voiceless and voiced stops, however, is only present after a nasal, as illustrated in (15a). That is, voiceless stops can occur in all positions, initially, medially – following oral and nasal segments – and finally, but voiced stops only occur after the nasal (nominal) prefix \(\text{N-}\). As a result, the voicing contrast in stops is only present postnasally (Teil-Dautrey 1991). It is instructive to examine Teil-Dautrey’s (1991) historical data on the origin of voiced stops. Present-day voiced stops in Basaa are reflexes of Proto-Bantu (PB) \(*\text{b}, *\text{d}, *\text{g}\) and \(*\text{p} *\text{t} *\text{k}\) in nouns with a nominal class 9/10 prefix \(\text{N-}\). That is to say, Proto-Bantu \(*\text{p} *\text{t} *\text{k}\) were voiced and \(*\text{b} *\text{d} *\text{g}\) were preserved as voiced stops in Basaa only when they followed the nasal nominal prefix \(\text{N-}\) (but not other nasals), as exemplified in Examples 1–3 in (15b). In all other contexts, preservation of voicing in Basaa involves de-stopping or implosivization (Examples 4–8; this will be addressed further in the next section). Although morphological factors are clearly at play, the historical data illustrate that coarticulatory nasalization facilitates maintenance of voicing in following stops.

(15)  Basaa, Bantu language

a. distribution of voiceless and voiced stops (Teil-Dautrey 1991)

\[
\begin{array}{llll}
\text{pén} \text{“color”} & \text{li-pén} \text{“honor”} & \text{m-pén} \text{“prong of a fork”} & \text{lép} \text{“water”} \\
\text{ter} \text{“to crush”} & \text{li-tén} \text{“stain”} & \text{n-teî} \text{“length”} & \\
\text{kap} \text{“share”} & \text{li-kay} \text{“science”} & \eta-kal \text{“prisoner”} & \text{lek} \text{“burn”} \\
\end{array}
\]
b. diachronic data (Teil-Dautrey 1991)

1. PB *tope n-dzβɔ 9/10 “mud”
2. PB *ding n-deŋ 9/10 “love”
3. PB *teṭe n-ter 3/4 “basket”
4. PB *beγ bek “break”
5. PB *gubu ɲ-guβi 9/10 “hippopotamus”
6. PB *diŋ lek “burn”
7. PB *diŋ yej “search”
8. PB *pud pɔr (also pɔt) “speak”

(16) Majorcan Catalan, stop voicing contrast preserved postnasally
(Dols & Wheeler 1995)

/b/
  dobl [pl] “I double”
  sembl [bl] “I think”

/p/
  acopl [pl] “I fit together”
  umpl [pl] “I fill”

Similarly, data from Majorcan Catalan suggests that a preceding nasal facilitates voicing in a following obstruent, thus preventing final obstruent devoicing. In Standard Catalan, underlying voiced stops become voiceless before a pause (Final Obstruent Devoicing, FOD). In Majorcan Catalan, however, the underlying stop voicing distinction is maintained post-nasally in word-final stop + non-syllabic /l, r/ clusters (Dols & Wheeler 1995; Llach 1999; Recasens, Espinosa & Solanas 2004). Thus postvocally, the voicing contrast is neutralized (/pl/ and /bl/ both have voiceless stops due to FOD, as shown in the first Example in (16)), whereas postnasally the two clusters differ in stop voicing (second Example in (16)), with the voiced stop showing more voicing during the closure, a shorter stop closure duration and a longer preceding nasal than the voiceless stop (Recasens et al. 2004; Llach 1999). Thus the presence of a nasal helps to maintain the underlying stop voicing distinction in word-final stop + liquid clusters in Majorcan.¹¹

An example of spontaneous nasalization developing in order to maintain the voicing contrast in final stops was suggested by Jones (2001) for Lancashire dialects and was reviewed above (illustrated in (14)).

3.3 Nasalization as a maneuver to prolong/initiate voicing

The data in 10–14 illustrates the emergence of phonetic nasalization next to voiced but not voiceless stops. Such patterns cannot be explained in terms of the lower tolerance of voiceless stops for nasalization, postnasal voicing or a constraint against

¹¹. The voicing contrast in Majorcan stop + nonsyllabic liquid clusters is only preserved in the first and third person singular (if no final epenthetic vowel is introduced) of the present tense in the second and third conjugations.
nasals preceding voiceless stops simply because these cases do not contain a nasal etymologically or occur in a nasal context. Similarly, the data in 15–16 illustrating the maintenance of voicing in stops, and hence of the stop voicing contrast, only in a nasal context cannot be explained by the principles above. However, these patterns may be explained in terms of nasalization facilitating voicing in adjacent stops. As noted, due to aerodynamic factors, a slightly lowered velum reduces the oral pressure and thus increases the rate of airflow through the glottis, which favors voicing.

The specific physical factors are the following. Voicing is maintained if the vocal folds have the appropriate degree of adduction and tension, and if there is sufficient airflow through the glottis \( U_g \) to sustain vocal fold vibration. For sufficient transglottal airflow, the pressure difference across the glottis must be about 1–2cm\( \text{H}_2\text{O} \), that is, the subglottal pressure must be at least 1–2cm\( \text{H}_2\text{O} \) higher than the oral pressure (Figure 4, time 1). Since transglottal flow for voicing is pressure \( (P) \) dependent, as shown in equation (17) below, and a high oral and nasal resistance for obstruents increases oral pressure as air continues to flow from the lungs and accumulates in the oral cavity, over a few tens of milliseconds the pressure differential drops below the required threshold, airflow through the glottis decreases, and voicing is extinguished (Figure 4, time 2). Note that voicing ceases due exclusively to aerodynamic factors (‘passive devoicing’) and not to active laryngeal adjustments. As illustrated in Figure 4 (time 2’), decreasing the nasal resistance by lowering the velum reduces oral pressure, and hence increases the rate of flow through the glottis, with the result that voicing is prolonged.

\[
P_{\text{subglottal}} - P_{\text{oral}} > 1-2\text{cmH}_2\text{O}
\]

\[
P_{\text{subglottal}} - P_{\text{oral}} \leq 1-2\text{cmH}_2\text{O}
\]

\[
P_{\text{subglottal}} - P_{\text{oral}} > 1-2\text{cmH}_2\text{O}
\]

Figure 4. Diagrammatic representation of the pressure difference required for continued transglottal flow

\[
U_g = A_g (P_{\text{subglottal}} - P_{\text{oral}})^a c
\]

where \( U_g \) is the transglottal airflow; \( A_g \) is the glottal area; \( a \) varies between 0.5 and 1; \( c \) is a constant
Similarly, voicing initiation – which requires a larger pressure difference \( (P_{\text{subglottal}} - P_{\text{oral}} > 3–4\text{cmH}_2\text{O}; \text{Baer 1975}) \) than that required to sustain it, due to the need to overcome inertial effects – can be facilitated by reducing the oral pressure through the nose and maximizing transglottal flow. Nasal venting might be particularly helpful to initiate voicing phrase-initially and sustain voicing phrase-finally due to difficulty involved in achieving the pressure differential with a lower subglottal pressure in these environments (Westbury & Keating 1986; Slifka 2000).

Nasal leakage is one of a number of articulatory maneuvers, used singly or in combination, which may be used to reduce the oral pressure and thus facilitate transglottal flow for voicing. These maneuvers are directed at reducing the oral pressure by (1) diminishing the air flowing into the oral cavity (i.e. decreasing the area of glottal opening and/or increasing the adductive tension of the vocal folds); (2) releasing air from the oral cavity, by diminishing the oral resistance – i.e. allowing air to escape through the oral constriction – or diminishing the nasal resistance – i.e. nasal leakage; or (3) actively enlarging the volume of the oral cavity, by lowering the larynx, fronting the articulatory constriction, raising the velum or relaxing the walls of the supraglottal cavity (Rothenberg 1968; Bell-Berti 1993; Westbury 1983). Such active articulatory gestures aimed at prolonging or initiating voicing have been referred to as ‘active voicing’.

A number of such maneuvers to sustain voicing are illustrated in the historical data for Basaa. According to Teil-Dautrey (1991), Proto-Bantu *b, *d, *g became voiceless in Basaa except when following the nasal nominal prefix N-, most likely due to coarticulatory nasalization (as argued for Examples 1–3 in (15b) above). Voicing in the stop was also preserved in the case of labial *b as an implosive [ɓ] word-initially or a fricative [β] intervocalically (Examples 4–5 in (15b)), both gestures involving maneuvers to prolong transglottal flow (larynx lowering and oral cavity expansion in the case of implosives, and reduction of the oral resistance in the case of the fricative). Similarly, voicing was preserved in the case of *d when it developed into a /l, j, r/, that is, when the oral resistance for the stop was decreased in magnitude (resulting in a lateral or glide, Examples 6–7 in (15b)) or in time (resulting in a flapped /r/, Example 8 in (15b)), thus preventing the build-up of pressure over time and favoring transglottal flow. Elsewhere Proto-Bantu voiced stops became mostly voiceless stops or were lost. These data reveal that in the development from Proto-Bantu to Basaa maintenance of vocal fold vibration in stops was linked in all cases to active gestures to lower oral pressure in order to facilitate transglottal flow for voicing. The data also illustrate that, as expected, more anterior places of articulation are more likely to preserve voicing. Similar maneuvers and phonetic results (i.e. prenasalization, fricativization, lateralization, gliding, flapping and implosivization) have been reported for a variety of languages (e.g., Rotokas, Firchow & Firchow 1969:274, cited in Hyman, to appear; Sindhi, Turner 1924; Palenquero, Piñeros 2003).
The fact that diminished nasal resistance is a way to moderate oral pressure and maintain vocal fold vibration has been observed by previous investigators (e.g., Rothenberg 1968; Lisker & Abramson 1971; Bell-Berti 1993) and the fact that nasal venting may be exploited in languages to facilitate voicing has been noted for some isolated patterns, such as prenasalization of voiced stops (Kaiser 1934 for Sumbanese (cited in Rothenberg 1968); Henton, Ladefoged & Maddieson 1992:71; Piñeros 2003 for Palenquero); the analysis of prenasalized stops in Mixtec as simple voiced stops (Iverson & Salmons 1996), and postnasalization of voiced stops to maintain the voicing contrast (Jones 2001). However, the overarching principle relating the seemingly disparate patterns pertaining to the preservation of voicing in stops the, preservation of the [stop] feature only postnasally, the emergence or preservation of nasals, prenasalization or postnasalization of stops, and the maintenance of voicing contrasts was not noted previously.

3.4 Distribution of prenasalization and aerodynamic constraints on voicing maintenance/initiation

If prenasalization was indeed an articulatory maneuver to facilitate voicing in stops, one would expect it to apply most often to stops in which voicing is more severely endangered (i.e., velars, with a smaller back cavity, smaller area of compliant tissue, and less capacity to expand actively), followed by coronals and labials, with a comparatively larger cavity, greater area of compliant tissue, and a greater capacity for active expansion (Westbury 1983; Ohala & Riordan 1979; Ohala 1983). This is precisely what the distribution of prenasalization in Japanese dialects (Yamane-Tanaka 2005) in Table 1 shows:

Table 1. Phonetic manifestations of voiced stops in Japanese dialects

<table>
<thead>
<tr>
<th>Dialect group</th>
<th>Phonetic manifestation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>[mb nd η]</td>
</tr>
<tr>
<td>B</td>
<td>[b nd η]</td>
</tr>
<tr>
<td>C</td>
<td>[b d η]</td>
</tr>
<tr>
<td>D</td>
<td>[b d g]</td>
</tr>
</tbody>
</table>

Of those dialects which prenasalize voiced stops intervocally, some prenasalize the three voiced stops [b d g], others prenasalize only [d g], and still others prenasalize only [g], such that there is an implicational relationship [mb] ⊃ [nd] ⊃ [η]. This means that the velar stop is prenasalized in virtually all Japanese dialects where this process occurs (33/34), [d] is the next most commonly prenasalized stop (15/34), and prenasalized [b] has the most restricted distribution (11/34). Interestingly, Yamane-Tanaka (2005) shows that the historical process was in fact
one of loss of prenasalization, since all voiced stops were prenasalized in Old and Early Middle Japanese. Historical records show that prenasalization in the central dialects of Japan was first lost in labials in Middle Japanese, \([mb] > [b]\), later in coronals in Modern Japanese, \([nd] > [d]\), and only recently in velars. Thus, the distribution of prenasalized stops reflects historical stages in the language. The progression in the loss of prenasalization correlates with well-known aerodynamic constraints on the maintenance of voicing related to back cavity size and compliance, and thus provides support for the hypothesis that nasal leakage is an adjustment to favor transglottal flow for voicing, utilized most often where voicing is more severely compromised.

Iverson & Salmons (1996) note a similar distribution of prenasalization in voiced stops in Mixtec dialects. As illustrated in Table 2, labials are not prenasalized in most dialects (with only optional prenasalization word-initially in Chalcatongo Mixtec), coronals always appear with prenasalization, whereas voiced velars are rare, do not occur word-initially, and are always prenasalized word-medially.

### Table 2. Distribution of prenasalization in word-initial and medial position in Mixtec varieties (Iverson & Salmons 1996)

<table>
<thead>
<tr>
<th>Word-initial</th>
<th>Word-medial</th>
</tr>
</thead>
<tbody>
<tr>
<td>([mb])</td>
<td>([b])</td>
</tr>
<tr>
<td>([nd])</td>
<td>([nd])</td>
</tr>
<tr>
<td>(–)</td>
<td>(–)</td>
</tr>
</tbody>
</table>

The pattern of prenasalization is in line with the difficulty involved in initiating and maintaining voicing (i) in velar stops vis-à-vis stops at places of articulation farther forward, and (ii) word-initially vis-à-vis intervocally (Westbury & Keating 1986). The maintenance of voicing is easier for bilabial stops vis-à-vis stops at more posterior places of articulation. Hence active articulatory maneuvers intended to facilitate voicing – such as reducing the nasal resistance or prenasalization – are not commonly found in bilabial stops while they are found in coronal and velar stops.\(^\text{12}\) Similarly, voicing is more difficult to initiate phrase-initially than medially, and more prenasalization is found word- and phrase-initially, at least in the bilabial series. In velar stops, voicing is only

\(^{12}\) In some dialects, such as AlacatlaZala and Ayutla, the labial stop is realized as fricative \([β]\), suggesting that the oral resistance is reduced in order to facilitate voicing (Iverson & Salmons 1996).
present when the aerodynamic conditions conducive to voicing are maximized: word-medially and prenasalized.

3.5 Discussion

The data reviewed suggest that Ohala & Ohala’s (1991) acoustic-auditory explanation for voiceless stops being more resistant to nasalization than voiced stops (in terms of nasalization undercutting the stop or voiceless character of voiceless stops) may be complemented by aerodynamic factors. Voiced stops favor nasalization because nasal leakage in the initial portion, or during the whole closure, of the stop contributes to keeping a low oral pressure which favors transglottal flow for voicing. Thus nasal leakage (pre- or postnasalization) may be considered a way to fine-tune the conflicting requirements of low oral pressure for voicing and high oral pressure for obstruency. Indeed, nasal leakage and a lower oral pressure (which is detrimental for obstruency) can be tolerated precisely by voiced stops which do not rely as heavily on high intensity noise cues as voiceless stops do.

This interpretation suggests that nasal leakage may be an independently controlled gesture accompanying glottal adduction that is aimed at preserving vocal fold vibration. Note that Ohala & Ohala’s account, on the other hand, implies that speakers actively inhibit coarticulatory velic lowering during the oral closure for voiceless, but not voiced stops, to prevent venting the oral airflow required for an intense stop burst. Thus, both explanations rely on active velopharyngeal adjustments linked to stop voicing. Phonetic data available in the literature provide support for the interpretation that the velum is actively controlled during voiced and voiceless sounds. For example, measurements of velopharyngeal closure force using a pressure sensing bulb and EMG activity (Kuehn & Moon 1998) indicate that the velum closure force tends to be greater for voiceless than for voiced consonants. Rothenberg (1968) notes that for certain voiced stops two constrictions are utilized to control the airflow – one at the glottis and the other at the velopharyngeal opening: the nasal passage is adjusted to be quite narrow in order to permit venting sufficient air to maintain vocal fold vibration.

The claim that nasal leakage may be used in some languages as an active adjustment to ensure voicing is compatible with the view of a speech regulating system in which respiratory and articulatory movements are aimed at regulating speech pressures for sound production (Warren 1986). Evidence of compensatory responses to changing aerodynamic conditions (for example, loss or addition of resistance within the vocal tract) and to structural defects (for example, velopharyngeal inadequacy) suggests that speakers attempt to maintain adequate intraoral pressures for consonant production (Kim, Zajac, Warren, Mayo & Essick 1997). Sensory information
may allow adjustment of nasal resistance in response to increased oral pressure and its effect on voicing. Thus nasal leakage during the production of voiced stops may be considered an adjustment to moderate the high resistance to exiting airflow at the oral constriction, and thus keep oral pressure low for voicing.

It must be stressed that phonetic prenasalization or postnasalization of voiced stops is viewed as a maneuver to initiate or prolong voicing in the stop and is not a deliberate nasal segment. Such phonetic nasalization, however, may be reinterpreted by listeners as an intended nasal and thus get encoded in the phonology (Ohala 1983). Such perceptual reinterpretation may be assisted by the perceptual bias towards hearing nasals before voiced vis-à-vis voiceless segments found by Ohala & Busà (1995), precisely the context where nasal leakage is found. Thus, the phonological nasals we find in the patterns reviewed are most likely the result of the reinterpretation of nasal leakage to facilitate voicing.

4. Conclusions

The reviewed typological, dialectal and instrumental data indicate that the likelihood of nasalization to combine with other features, in particular voicing, depends to a large extent on physical and perceptual factors. Acoustic-auditory reasons – specifically, laryngeal vibration (but not turbulence generated downstream from the velopharyngeal opening) resonating in the nasal cavities and adding the characteristic perceptual cues to nasalization – have been shown to be at the origin of nasal segments being largely voiced. On the other hand, manipulating the nasal seal to facilitate transglottal flow and vocal fold vibration during the stop constriction accounts for prenasalization and postnasalization of stops and a variety of typological patterns relating nasalization and voicing. The analysis of nasal leakage being used in languages to facilitate and preserve voicing in stops is supported by data indicating that nasals occur predominantly more often in contexts where vocal fold vibration is more difficult to maintain/initiate, for example, in velars vis-à-vis stops at more anterior places, and in phrase-initial vis-à-vis medial position. In addition, while prenasalization is found word- and phrase-initially, postnasalization occurs phrase-finally where voicing is more difficult to sustain due to oral pressure build-up over time and decreasing subglottal pressure.

The review of the data presented here suggests that the propensity of features to combine depends on their articulatory-aerodynamic and acoustic requirements. As stated by Ohala (2005), dependency relations between features due to speech aerodynamics, acoustics or perception cannot be captured by models such as Feature Geometry. For example, the aerodynamic interaction between nasalization and voicing illustrates that what happens at the velum can influence the
continuation or extinction of voicing (e.g., in postnasal voicing, preservation of voicing after a nasal) and that what happens at the glottis may be associated to velopharyngeal adjustments (e.g., prenasalization of voiced but not voiceless stops, preservation or emergence of nasals adjacent to voiced stops). Such dependency relations cannot be accounted for in a model where the nasal feature is at a different branch from the laryngeal features, with the result that neither one can specify the value of the other. Similarly, aerodynamic and acoustic factors are at the origin of nasalized voiced fricatives losing frication earlier vis-à-vis voiceless fricatives. Current phonological models, however, do not allow laryngeal features, which are at a different branch from supralaryngeal features, to dictate frication.

For acoustic reasons, the nature and location of the sound source (anterior or posterior to the velopharyngeal opening) determines the acoustic coupling to the nasal cavity and the perceptibility of nasalization (in the case of voiceless vs. voiced nasals and nasalized fricatives). Phonological models which represent the place of the supraglottal constriction, the nasal valve and the larynx at different nodes fail to capture the acoustic interactions between the glottal state and perceptibility of nasalization. In addition, finer quantitative detail is needed than what available phonological notations may allow us to represent. For example, in nasalized fricatives, where air flows out of the nose and the mouth, the size of the velopharyngeal opening relative to the area of the oral constriction is crucial because, due to the quantal nature of speech, small variations in the size of either opening may involve an abrupt acoustic change, such as the loss of audible frication and the percept of an approximant.

If, in effect, dependency relations between features or constraints are functional and can be explained by phonetic theory, then there is no need to encode them as constraints in universal grammar as proposed within Optimality Theory. In other words, the common patterning in languages can be accounted for by physical, physiological and auditory factors rather than by ‘formal’ constraints, and variation across languages would result from the way in which different languages (or dialects) deal with such physical constraints. For example, in the case of voiced stops, languages may either (i) yield to the ‘aerodynamic voicing constraint’ and devoice stops, or (ii) avert or resist the constraint and preserve voicing with a variety of articulatory maneuvers – such as nasal leakage, oral cavity expansion or decreasing the oral resistance – thus giving rise to the observed cross-linguistic variation. Similarly, in the case of the lower tolerance of voiceless obstruents to nasalization, some languages may yield to the effects of coarticulatory nasalization, and the characteristic burst at stop release may be weakened or absent, leading to the perceptual loss of buccal voiceless stops, other languages may inhibit nasal coarticulation and show an early raising of the velum to the occasional detriment of the nasal, thus preserving the cues for the stop, while still other languages may
make the precise alignment of the oral and velopharyngeal gestures so that the nasal and the voiceless stop are preserved. These patterns will certainly be encoded in the grammars of particular languages, given that languages differ in the patterns they exhibit, most likely in the form of allowable ranges of phonetic values for the sounds and sound sequences in the language.

In sum, one should be cautious about positing formal constraints to account for the dependency relations between features – and taking these constraints as being explanatory – when phonetic factors have not been discarded. As we advance in our understanding of the physical, physiological and auditory-perceptual aspects of speech, the need for formal statements of constraints is likely to dwindle significantly.

References


Hayes, Bruce & Tanya Stivers. 2000. “Postnasal Voicing”. Ms., Dept. of Linguistics, UCLA.


Fixed and variable properties of the palatalization of dental stops in Brazilian Portuguese

In an Italian immigrant community

Elisa Battisti¹ & Ben Hermans²
¹University of Caxias do Sul/²Meertens Institute

Assuming that unranked constraints generate variation and that features can reoccur at various levels in the segmental tree, the variable palatalization of dental stops in a speech variety of Brazilian Portuguese is analyzed as a process which is applied in order to link C(Aperture) of high vowels to a higher consonantal position, explaining the cross-linguistic tendency of high vowels to spread to preceding segments. A mixed approach which includes the representation of segments and the set of constraints referring to vowels and the metrical grid is adopted to explain the different rates of palatalization by underlying high vowels and raised vowels.

1. Introduction

For some time it has been claimed that quantitative data of phonological variation can help the linguist evaluate hypotheses or choose among different theories (Hinskens, Van Hout & Wetzels 1997; Anttila 1997, 2007). Naturally, models or theories can also shed light on language in use. The variable palatalization of dental stops in Brazilian Portuguese in its typological and quantitative generalizations is an instance of this two-way street.

In the Brazilian Portuguese variety spoken in Antônio Prado (Rio Grande do Sul, Brazil), a small city founded by Italian immigrants at the end of the 19th century, high front vowels can palatalize preceding dental stops. Battisti, Dornelles Filho, Lucas and Bovo (2007) observed that the total frequency of variable palatalization in that community is 30%, a modest rate when compared to other Brazilian speech varieties. These authors claim that, rather than being a change that is in progress, palatalization shows signs of stabilizing in the community due to strong social
conditioning. In this paper, we refer to a number of different theoretical approaches to linguistic analysis in an attempt to show that palatalization can be explained by a set of unranked constraints and by the representation of high vowels.

2. The palatalization of dental stops in Antônio Prado

In Brazilian Portuguese, both an underlying high vowel /i/ in stressed or unstressed positions and a phonetic [i] raised from an underlying /e/ in unstressed positions may palatalize the preceding dental stop:

a. Stressed /i/: medida~medida “measurement”, ativo~ativo “active”, dica~dica “hint”, tipo~tipo “type”.

b. Unstressed /i/: difícil~díficil “difficult”, tirar~tirar (to take), médico~médico “doctor”, ótimo~ótimo “great, very good”, Zatti~Zatti (surname), ti~ti (2nd PSm pronoun).

c. Unstressed /e/ (raised to [i]): vinte~vinte “twenty”, onde~ondi “where”, teatro~teatro “theatre”, desconhecido~desconhecido “unknown”, de manhã~demi manhã “in the morning”, te~te (2nd PSm pronoun).

Rates of palatalization differ significantly among speech varieties in Brazil – around 90% in Salvador, Bahia (Abaurre & Pagotto 2002) and in Porto Alegre, Rio Grande do Sul (Bisol 1991; Kamianeccky 2003); 62% in Alagoinhas, Bahia (Hora 1990), 47% in Flores da Cunha, Rio Grande do Sul (Almeida 2000), 8% in Florianópolis, Santa Catarina (Kamianeccky 2003) – as well as unstressed mid front vowel raising, which feeds palatalization (Hora 1990; Bisol 1991; Almeida 2000; Pagotto 2001).

---

1. Social networks play a role in the process in Antônio Prado. The authors found that dense networks spread palatalization, but at the same time the quality of the ties of the networks blocks the diffusion of the rule, mainly to the rural area. The tension between these two social patterns of the network is one of the reasons for the tendency towards the stabilization of the rule observed in the community.

2. In Brazilian Portuguese, palatalization applies only to dental stops. It does not affect labial (hábil “skillful”; piada “joke”) or velar stops (guitarra “electric guitar”; aqui “here”), nor other kinds of coronal segments (sino “bell”; azia “heartburn”; rio “river”; perigo “danger”; liso “smooth”; menino “boy”).
The palatalization of dental stops in Brazilian Portuguese

3. The palatalization of dental stops in Antônio Prado (Rio Grande do Sul, Brazil)

3.1 Antônio Prado

Antônio Prado is located about 180 km from Porto Alegre, the capital city of the Brazilian state of Rio Grande do Sul, and about 61 km from Caxias do Sul, the largest city in the state after Porto Alegre (see Figure 1). The city has an area of 347.6 square kilometers and a population of 14,344 inhabitants,\(^3\) 65.2 % of whom live in the urban area.

![Figure 1. Location of Antônio Prado, Caxias do Sul and Porto Alegre in the state of Rio Grande do Sul, Brazil](image)

The social history of Antônio Prado is exogenous: its population was implanted in an isolated region, where it stayed due to the precarious conditions of the roads connecting the city to the rest of the state and the country. The situation started changing only 30 years ago and contributed towards the development of a localism similar to the one described by Milroy (1980).

As the sixth and last Italian settlement in the northeast of Rio Grande do Sul, Antônio Prado was established by Brazil’s emperor Dom Pedro II in 1886 and

---

was called after the Minister of Agriculture who introduced Italian immigrants to Brazil, most of them from the north of Italy. Antônio Prado owes most of its characteristics, such as its catholic religious practices, to the Italian immigrants.

3.2 The palatalization of dental stops in Antônio Prado: Quantitative aspects

The variable rule analysis (Labov 1972, 1994, 2001) of 26,600 tokens collected from the interviews of 48 BDSer informants made by Battisti, Dornelles Filho, Lucas and Bovo (2007) showed that palatalization applies to 30% of these tokens, i.e., 7,971 out of 26,600 tokens were palatalized. Half of the informants live in the urban area of the municipality and half in its rural surroundings. Informants are of both sexes and are divided into four age groups (15 to 29 years old, 30 to 49 years old, 50 to 69 years old, and 70 years old and over).

Three social variables – Gender, Age, Area – and five linguistic variables – Preceding Phonological Context, Following Phonological Context, Vowel Status, Syllable Position and Syllable Stress were controlled. Of these eight variables, Age, Area and Vowel Status were statistically significant in all runs. Of interest here are the results of the last variable, Vowel Status (Table 1).

Table 1. Results of variable Vowel Status (from Battisti, Dornelles Filho, Lucas and Bovo, 2007)

<table>
<thead>
<tr>
<th>Factors</th>
<th>Applic. / Total</th>
<th>%</th>
<th>Relative weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underlying high vowel /i/ (mentira)</td>
<td>5 661 / 9 533</td>
<td>59</td>
<td>0.89</td>
</tr>
<tr>
<td>Raised vowel [i] (gente, de manhã)</td>
<td>2 310 / 17 067</td>
<td>13</td>
<td>0.23</td>
</tr>
<tr>
<td>TOTAL</td>
<td>7 971 / 26 600</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

Input 0.198  
Significance 0.005

Most of the data (17,067 tokens) consist of unstressed mid vowels which can be variably raised, although the amount of unstressed mid vowels that are raised is only 13%. The environment itself is unfavorable to palatalization (relative weight of 0.23). In other Brazilian varieties the raising of mid vowels is very frequent and feeds palatalization. This is not the case in Antônio Prado, perhaps because of the contact of Portuguese with Italian dialects that are still spoken in the city.

---

4. Banco de Dados de Fala da Serra Gaúcha (Gaucha Sierra Speech Data Bank), UCS.
5. At the level of 0.05.
6. This is the hypothesis Roveda (1998) followed in her investigation of the raising of final unstressed /e/ in Flores da Cunha, a neighboring town near Antônio Prado. The low rate of
In contexts with the underlying high vowel /i/, the rule applied at a 59% frequency rate. The environment favors palatalization (relative weight of 0.89). The underlying high vowel may be stressed or unstressed, and the rule applies slightly more frequently in unstressed positions, as we can see in Table 2.

Table 2. Distribution of the palatalizing /i/ in different environments

<table>
<thead>
<tr>
<th>Environment</th>
<th>Applic./Contexts</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stressed /i/ (tipo)</td>
<td>3 464 / 6 098</td>
<td>57</td>
</tr>
<tr>
<td>Pre-tonic and non-final post-tonic /i/ (tirar, ótimo)</td>
<td>2 015 / 3 171</td>
<td>64</td>
</tr>
<tr>
<td>Final post-tonic /i/ (Zatti)</td>
<td>165 / 245</td>
<td>67</td>
</tr>
<tr>
<td>Clitic (ti, obj.pron.2nd pp)</td>
<td>19 / 19</td>
<td>100</td>
</tr>
<tr>
<td>TOTAL</td>
<td>5 661 / 9 533</td>
<td>59</td>
</tr>
</tbody>
</table>

Clitics with an underlying high vowel are very small in number and therefore not meaningful for the present study. The frequency rates of palatalization by the underlying high vowel are around 60% in every position in the word, whether stressed or unstressed. These similar frequency rates do not allow us to conclude that stress (or its absence) in different word positions has effects on palatalization by the underlying high vowel.

The raised vowel [i], however, is always unstressed. It palatalizes the dental stops more frequently when it is in the final post-tonic position, as we can see in Table 3. But even in this position, palatalization by the raised vowel [i] is not frequent in Antônio Prado.

Table 3. Distribution of the palatalizing [i] in different environments

<table>
<thead>
<tr>
<th>Environment</th>
<th>Apl/Contexts</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-tonic and non-final post-tonic [i] (teatro, síntese)</td>
<td>186 / 2 150</td>
<td>9</td>
</tr>
<tr>
<td>Final post-tonic [i] (vinte, onde)</td>
<td>1 487 / 9 129</td>
<td>16</td>
</tr>
<tr>
<td>Clitic (de, te)</td>
<td>637 / 5 788</td>
<td>11</td>
</tr>
<tr>
<td>TOTAL</td>
<td>2 310 / 17 067</td>
<td>13</td>
</tr>
</tbody>
</table>

Vowel raising in Flores da Cunha is explained by the author as the result of the influence of the Italian morphological system (in which vowel quality is distinctive) on Portuguese. This may also be the case in Antônio Prado.
It could be said that palatalization in Antônio Prado not only tends to apply with the underlying high vowel, but it is also almost entirely restricted to it. Palatalization by the raised vowel, which is much less frequent, seems to be subject to the effects of word stress. In the next section we will adopt a mixed approach to attempt to explain why this is so, or in other words, as Anttila (2007) posits about variation and phonological theory, why the process occurs more in one environment than in others. We will also attempt to answer the question about what determines the quantitative preferences among the variants, what is universal about the pattern, and particularly, why high vowels tend to spread to a neighboring consonant. The Parallel Structures Model (PSM) devised by Morén (2003, 2005) will enable us to represent the consonantal nature of high vowels and explain their tendency to spread to preceding segments. We will follow the proposal put forward by Hyde (2001) on metrical stress to analyze the interaction between vowel raising and palatalization. Anttila’s Multiple Grammars Theory (1997) will be needed to model the variable palatalization in the speech variety under analysis in its quantitative patterns.

4. The palatalization of dental stops

4.1 Palatalization by the high underlying vowel

In the spirit of Morén’s (2003, 2005) Parallel Structures Model (PSM), we assume that phonological features can recur at various positions in the segmental tree. The PSM is a model of feature geometry which combines insights from different feature theories and models of segment-internal representations:

a. the feature geometry designed by Clements (1985), with the unification of consonant and vowel place features (Clements, 1991a) and the set of vowel height features connected to degrees of vocal tract constriction (Clements, 1991b);

b. Steriade’s (1993, 1994) idea that consonant manners are differentiated via root nodes which correspond to different degrees of vocal tract constriction;

c. the idea adopted in Particle Phonology (Schane, 1984), Dependency Phonology (Anderson & Ewen, 1987; Hulst 1989, 1999) and Element Theory (Harris & Lindsey, 1995) that vowels and consonants make use of the same set of features/models.

An assumption made by the PSM is that the grammar is structurally economical and that more complex structures are built from less complex ones. The ensuing prediction of such an assumption is that every language has simple segments that are featurally minimal.
In the PSM line of research, phonological segments are composed of a limited set of identical structures or features. We assume features ‘C’ and ‘V’. These features occur at the root node, where they define major classes; they occur at an intermediate level, at the vocalic or consonantal node where they define constriction.

(1) dental stop

\[
\begin{array}{c}
C \quad \text{ROOT} \\
| \\
C \quad \text{CONSONANTAL} \\
\end{array}
\]

In vocoids, they also occur at the level of the aperture features, where C characterizes relatively closed vowels (high and mid vowels) and V characterizes relatively open vowels (mid and low vowels).

(2) high (front) vowel \hspace{1cm} mid (front) vowel \hspace{1cm} low (front) vowel

\[
\begin{array}{ccc}
V & | & V \\
| & | & | \\
V & | & V \\
\mid C & \mid C & \mid C \\
\mid \text{Place} & \mid \text{Place} & \mid \text{Place} \\
\mid \text{Cor} & \mid \text{Cor} & \mid \text{Cor} \\
\end{array}
\]

When combined, features C and V enter into a head-dependency relation, in which head constituents are more prominent.

We claim that palatalization is the requirement whereby C(Aperture) is linked to a higher C-node. This means that high vowels tend to spread to a neighboring consonant, palatalizing it (3).

(3) The Structure of a palatalized consonant

\[
\begin{array}{c}
C \quad V \\
| \\
C \quad V \\
\mid \text{Cor} \\
\end{array}
\]

There are three assumptions here: (i) in order to create a link between C(Root) and C(Aperture), V(Vocalic) spreads to the consonantal root node taking C(Aperture) with it; (ii) as a secondary node on C(Root), V(Vocalic) can only allow Cor (the unmarked place feature); (iii) as a secondary node on C(Root), V(Vocalic) cannot have its own independent place feature; it must therefore share its place feature.
with the onset, C(Root), palatalizing the consonant. Taken together, the second and third assumptions account for the fact that only front vowels can create a secondary place of articulation on the consonant (there is no labialization) and only coronal consonants are palatalized (there are no palatalized labials).

In an optimality-theoretical framework, the constraints involved are:

(4) Palatalization (Pal)
    Head-C(Aperture) must be linked to a higher C-position

(5) NoComplexRoot (NoCR)
    A root node may not branch

The constraint Pal requires that C(Aperture) spreads to a consonantal position, provided it is the head feature. The headedness requirement accounts for the fact that C(Aperture) only spreads if it is in the domain of a high vowel. We assume that in mid vowels V(Aperture) is the head.

Pal favors palatalization, NoCR is unfavourable to palatalization. On the assumption that they are unranked (Anttila, 1997), we obtain a 50–50 distribution of palatalized/non-palatalized consonants.

(6) Constraints: Pal, NoCR; rankings: none
    Pal >> NoCR  tá  (palatalization applies)
    NoCR >> Pal  tí  (palatalization does not apply)

The interaction of these constraints explains the attested frequency of palatalization by the high underlying vowel in Antônio Prado. We will next look at what happens regarding palatalization by the raised vowel.

4.2 The raising of mid, unstressed vowels and palatalization

As seen above (Table 1), there are 17,067 tokens of unstressed /e/ in the corpus. Only 2,310 of those tokens (13%) are raised and palatalize the dental stops, and most of the palatalizing tokens (Table 3) occur in final post-tonic position. This raises the question of how these quantitative regularities relate to grammar and why it is that the frequency of palatalization by a raised vowel in unstressed positions is low. It seems that vowel raising and palatalization interact with stress. Such interaction is explored in our analysis, following Hyde’s proposal (2001) on metrical stress.

---

7. Antilla (2007) calls this competing grammars model The Multiple Grammars Theory. According to this, any combination of possible total rankings of constraints is a possible individual’s collection of grammars. In other words, an individual’s grammar may contain several different grammars. In mathematical terms, this results from a factorial combination: the number of possible grammars, with n constraints, is n! = n x (n-1) x (n-2) x ... x 3 x 2 x 1.
Hyde (2001) focuses on the discrepancy between the binary patterns predicted by standard accounts and the patterns actually attested cross-linguistically. His proposal restricts the generation of stress patterns by means of (a) strict succession at all levels of the prosodic hierarchy, especially the exhaustive parsing of syllables into feet; (b) improper bracketing, allowing categories on the same level of the prosodic hierarchy to intersect, i.e., to share entries on the metrical grid; and (c) violable foot-stress relationship, allowing feet to share stress. The basic stress typologies are established by both Alignment and NonFinality, but the central role is played by the last symmetrical constraint.

We assume that unstressed mid vowels which are raised are devoiced because, in the line of Hyde (2001), they do not have a corresponding gridmark, i.e., they are not represented on the metrical grid, as in (7).

\[
\begin{array}{c|c}
\text{Voiced [e]} & \text{Voiceless [e]} \\
\hline
* & V \\
V & V \\
| & |
\end{array}
\]

Vowels tend to have a gridmark: the greater the sonority of a vowel, the greater the pressure to assign it a gridmark. Unstressed vowels, however, shun gridmarks. We express all this with the following set of constraints:

\[
\text{(8) Constraints regulating the interface between vowels and the grid}
\begin{align*}
\text{a.}& \quad V(\text{Root}) \rightarrow * \\
& \quad \text{A vowel must have a gridmark.} \\
\text{b.}& \quad V(\text{Aperture}) \rightarrow * \\
& \quad \text{An open vowel must have a gridmark.} \\
\text{c.}& \quad V_{\text{DEP}} \rightarrow \text{No}^* \\
& \quad \text{A vowel in a dependent (unstressed) position may not have a gridmark.}
\end{align*}
\]

Two constraints are in favor of assigning an unstressed mid vowel a gridmark, one constraint disfavors the assignment of gridmarks to unstressed mid vowels. Assuming that these three constraints are unranked, we obtain the following distribution: 1/3 of the unstressed mid vowels are devoiced; 2/3 of the unstressed mid vowels are voiced, as we may see in (9).

\[
\begin{align*}
\text{(9) } & \quad V(\text{Root}) \rightarrow * \rightarrow V(\text{Aperture}) \rightarrow * \rightarrow V_{\text{DEP}} \rightarrow \text{No}^* \rightarrow \text{voiced [e]} \\
& \quad V(\text{Root}) \rightarrow * \rightarrow V_{\text{DEP}} \rightarrow \text{No}^* \rightarrow V(\text{Aperture}) \rightarrow * \rightarrow \text{voiced [e]} \\
& \quad V(\text{Aperture}) \rightarrow * \rightarrow V(\text{Root}) \rightarrow * \rightarrow V_{\text{DEP}} \rightarrow \text{No}^* \rightarrow V(\text{Root}) \rightarrow * \rightarrow \text{voiced [e]} \\
& \quad V_{\text{DEP}} \rightarrow \text{No}^* \rightarrow V(\text{Root}) \rightarrow * \rightarrow V(\text{Aperture}) \rightarrow * \rightarrow \text{devoiced [e]} \\
& \quad V_{\text{DEP}} \rightarrow \text{No}^* \rightarrow V(\text{Aperture}) \rightarrow * \rightarrow V(\text{Root}) \rightarrow * \rightarrow \text{devoiced [e]}
\end{align*}
\]
Mid vowels are not subject to the Pal constraint, i.e., they do not palatalize the preceding dental stops because in mid vowels the feature C(Aperture) is not the head. However, they may be devoiced and raised in unstressed positions when V(Aperture) is deleted and C(Aperture) becomes the head. They are then subject to the Pal constraint and palatalization applies.

The frequency of palatalization in such an environment is very low in Antônio Prado, only 13%. This suggests that at least two other constraints may interact with the ones in (7), a constraint requiring C(Aperture) to be the head in vowels in dependent positions and a faithfulness constraint. We will not go further into this topic in the present study. For its purposes, it is enough to have an idea of the reason why in Antônio Prado, unstressed mid front vowels trigger palatalization much less frequently than high vowels; it is due to their representation and to the interaction of constraints referring to vowels and the grid.

5. Conclusion

As a variable linguistic process, palatalization is naturally correlated to social factors, but it is linguistically conditioned as well. The quantitative patterns of palatalization in Portuguese in the Brazilian city of Antônio Prado (Rio Grande do Sul, Brazil) refer to the grammar and so can be theoretically interpreted and predicted. We have attempted to show that the distribution of palatalized and non-palatalized forms is explained by the representation of (high) vowels and the constraints operating on representations.

Based on the approaches quoted above, in particular the idea that unranked constraints generate variation and that in vowel representation, features (‘C’ and ‘V’) can recur at various levels in the segmental tree, we analyzed palatalization as a process which is applied in order to link C(Aperture) of high vowels to a higher consonantal position, explaining their cross-linguistic tendency to spread to preceding segments.

Unlike other varieties of Brazilian Portuguese, unstressed mid front vowels do not tend to be raised in Antônio Prado. We have shown that mid vowels do not palatalize the preceding dental stops because of their representation and the constraints referring to the metrical grid. However, when devoiced and raised, they may palatalize the dental stops.

The study of a variable process in a speech variety of Brazilian Portuguese allowed us to show that quantitative aspects observed in large amounts of usage data can shed light on representational issues, and vice versa, representational issues can shed light on phonological variation.
References


Post-tonic vowel harmony in some dialects of Central Italy
The role of prosodic structure, contrast and consonants*

Stefano Canalis
Università di Padova

In Central Italy several dialects display post-tonic regressive vowel harmony, by which post-tonic vowels copy all the features of the word-final vowel. On the basis of phonetic and phonological arguments I argue that the penultimate vowel of proparoxytones, the typical target of this process, is a prosodically weak position, which makes it a good target for assimilation. In some dialects harmony is active only if a liquid consonant intervenes between the trigger and target vowels; since in these dialects liquids do not contrast for place, underspecification can explain this asymmetry. Since place specification of non-liquid consonants is required in other varieties, which nevertheless display harmony across any intervening consonant, following Clements (2001) I argue that in this case some nodes of feature geometry are not active.

1. Introduction

Among the fairly numerous types of vowel harmony processes that can be observed in the dialects of Italy,1 several varieties of Central Italy display a regressive post-tonic harmony process which raises several problems, notably with regard to the role of prosodic structure and consonants.

* I would like to thank the audience at PaPI 2007 in Braga, Laura Vanelli and three anonymous reviewers for their very helpful comments, which improved several aspects of my paper. Obviously, any remaining errors and inconsistencies are my own.

1. It may be useful to point out a potential source of misunderstanding in terminology: the Italian word dialetto, which is usually translated as ‘dialect’ in English, refers to minor Romance languages, not to local varieties of Italian (which are usually called italiani regionali ‘regional Italians’). This means that Italian ‘dialects’ can differ fairly significantly from Italian (although the dialects of Central Italy, from which my examples are drawn, are definitely closer to it than any other dialect).
This harmony is attested – without a uniform distribution – within an area including South-Eastern Tuscany, Umbria, the Marche and Northern Lazio, plus a small isolated zone in North-Western Tuscany (similar phenomena are attested in some areas of Southern Italy as well – Salento, Southern Calabria and Eastern Sicily – but they will not be discussed here).

Synthetic general descriptions of the dialects of Central Italy can be found in several of Maiden & Parry’s (1997) chapters, especially in Vignuzzi (1997); the data presented here come from Maiden (1988, 1991, 1995) (covering most of the varieties examined here), Camilli (1929) (an account of the dialect of Servigliano, in the Marches), Elwert (1958) (on the dialect of Sant’Oreste, Northern Lazio), and Venturelli (1979) (on varieties spoken in Garfagnana, a region of North-Western Tuscany). Moreover, two atlases of Italian dialects, AIS (Jaberg & Jud 1928–1940) and ALI, include many relevant examples. As for the dialect of Umbertide (North-Western Umbria), data also come from my own fieldwork.

In the vowel harmony process examined in this paper, the post-tonic vowels copy all the features of the word-final vowel. On the basis of several phonetic and phonological arguments it will be argued that, given the metrical structure of these dialects, the penultimate vowel of proparoxytones – the most common target of this harmony – is a metrically weak position, which makes it a good target for assimilation. Some hypotheses about directionality in vowel harmony recently put forward by Hyman (to appear) will be adopted to account for the right-to-left direction of harmony. Additionally, in some dialects harmony is active only if a liquid consonant intervenes between the trigger and target vowels: since in these dialects liquids do not contrast for place, underspecification can explain this asymmetry. As place specification of nonliquid consonants is also needed in the varieties whose harmony is never blocked by consonants, following Clements (2001), it will be argued that in this case some nodes of feature geometry are not active.

2. Data

Like Italian, all these dialects have a vowel system /i, e, a, ɔ, o, u/ in stressed position, while in unstressed vowels the contrast between low-mid /ɛ, ɔ/ and high-mid /e, o/ segments is neutralized: all mid vowels are realized as [e, o], giving rise to alternations like [ˈvɛngo] ‘I come’ [veˈnete] ‘you (pl) come’, [ˈmɔro] ‘I die’ [moˈrete] ‘you (pl) die’ (examples from the Servigliano dialect (Camilli 1929)).

The harmony process consists of a regressive complete assimilation of all the post-tonic vowels: in other words, all the post-tonic vowels are a copy of the word-final vowel, thus creating post-tonic sequences of identical vowels.
Post-tonic vowel harmony in some dialects of Central Italy

(1) a. Servigliano (Camilli 1929)

[ʼpredoko] ‘(I) preach’
[ʼprediki][2] ‘(you (sg)) preach’
[ʼpredaka] ‘(s/he) preaches’
[doʼmenaka] ‘Sunday’
[doʼmenaceke] ‘Sundays’
[ʼstomuku] ‘stomach’
[ʼstomiiki] ‘stomachs’
[ʼfrate] ‘brother’
[ʼfratutu] ‘your brother’

The pattern exemplified in (1) is the most widespread. Yet, there are two important parameters of variation.

In dialects like Servigliano, the word-final vowel features spread freely across any intervening consonant in the post-tonic domain. However, in a fairly high number of dialects vowel copy occurs only if the intervocalic consonant is [l] or [r].

In a few dialects (in the Garfagnana area (Venturelli 1979) and in Southern Italy (Maiden 1988:133)) there is no complete copy, but only assimilation in height: if the word-final vowel is [i] or [u], preceding post-tonic /e/ and /o/ raise to [i] and [u] respectively.

The two parameters can intersect: for example in some localities of Garfagnana, mid vowels are raised by word-final [i] ([u] is never present word-finally in this variety) only if the intervening consonant is a liquid.

b. [ʼmitulu] /ʼmetti=lu/ ‘put it-M’
[ʼmitili] /ʼmetti=li/ ‘put them-M’
[ʼmetala] /ʼmetti=la/ ‘put it-F’
[ʼmettele] /ʼmetti=le/ ‘put them-F’
[ʼmitutulu] /ʼmetti=lu/ ‘put it-M there’
[ʼmetatala] /ʼmetti=la/ ‘put it-F there’

(2) Garfagnana (Venturelli 1979:104)

[ʼkavolo] ‘cabbage’
[ʼkavuli] ‘cabbages’
[ʼandjelo] ‘angel’
[ʼandqlii] ‘angels’
[ʼalbero] ‘tree’
[ʼalbiri] ‘trees’

2. The height alternations in some of the stressed vowels in (1a, 1b) are due to metaphony: stressed high-mid and low-mid vowels are raised to high and high-mid respectively when the word-final vowel is /i/ or /u/ (compare e.g. [ʼprediki] to [ʼpredoko], [ʼmitulu] to [ʼmetala]).
Moreover, it is important to note that even when non-liquid consonants block harmony the quality of the post-tonic vowel – although not dependent on harmony – is not free. It is not a copy of the final vowel, yet it is not drawn from the full set of vowels which can occur in unstressed position ([i, e, a, o, u]) either; on the contrary, it is limited to [i] or [e] depending on the variety. The dialect of Umbertide is (or rather was) one of the most regular examples of this type of harmony: in (3a) post-tonic vowels harmonize when followed by a liquid, but in (3b) only [i] is found.

(3) a. Umbertide

[koˈkomːoro]  ‘watermelon’
[ˈsigoro]  ‘cigar’
[ˈfodara]  ‘lining’
[ˈskatala]  ‘box’
[ˈskatele]  ‘boxes’
[ˈfragwala]  ‘strawberry’
[ˈfragwele]  ‘strawberries’

b. [ˈsabito]  ‘Saturday’
[doˈmenːika]  ‘Sunday’
[ˈstomːiko]  ‘stomach’
[ˈkaliˈe]  ‘goblet’
[ˈmonika]  ‘nun’
[ˈpetine]  ‘comb’
[ˈʤovino]  ‘young man’
[kreˈdetiːje]  ‘believe him’

The same restriction holds for other varieties as well: in Sant’Oreste (Lazio) the penultimate vowel of proparoxytones is always [i] (4a), except when the following consonant is a liquid, which allows total harmony (4b).

(4) a. Sant’Oreste (Elwert 1958)

[ˈsab:iːtu]  ‘Saturday’
[ˈʃkomːida]  ‘uncomfortable-F’
[ˈʃtomːiku]  ‘stomach’
[ˈʃiːflu]  ‘bug’
[ˈtrapinu]  ‘drill’
[ˈʃɛfine]  ‘Stephen’
[ˈʤakimu]  ‘Jacob’

b. [ˈarbulu]  ‘tree’
[ˈarbili]  ‘trees’

3. Or [i, e, a, o] in the dialects which do not have any word-final [u].

4. Some varieties display a limited number of post-tonic [o] in proparoxytones.
As for the domain of harmony, in these dialects – as in most Romance languages – stress is restricted to a three-syllable window at the right edge of the word. But – again as in other Romance languages – enclitic particles attached to a word can cause a relaxation of this constraint, allowing stress to come before the antepenultimate syllable: thus stress on the fourth – as the last two examples in (1b) – or even fifth last syllable is possible. Since harmony assimilates all the post-tonic vowels to the last one, the restriction on stress position implies that up to three vowels can undergo assimilation, if enclitics are present (although most cases of harmony are in proparoxytones, where only one vowel, the penultimate, is assimilated).

It must be pointed out that the regularity of this harmony process shows a large degree of variation from variety to variety; while in some localities it is regular, in other areas there are several exceptions, and in still other areas it seems to be a relic of a past synchronic state, with only few traces left. In general, the influence of Italian and other dialects seems to have reduced this phenomenon. Furthermore, back vowels seem to be slightly more resistant to assimilation, and in varieties in which harmony is influenced by consonants there can be sporadic traces of consonants other than liquids allowing assimilation. Finally, in several varieties vowel copy across an intervening liquid consonant is often less stable than neutralization to [i] (or [e]) before the other consonants; for instance [ˈfragola] instead of [ˈfragwala], [ˈandʒelo] instead of [ˈandʒolo] can be heard in Umbertide today, although forms like [ˈstɔmiko], [ˈmonika] still resist.

3. **Theoretical problems**

These harmony processes raise various problems, which are summed up below:

1. cross-linguistically vowel harmony goes from the root to affixes much more frequently than the other way round (cf. Hyman (to appear) for a general discussion on directionality in vowel harmony); here the triggering vowel is in most cases an inflectional ending
2. The trigger of harmony is usually a ‘strong’ vowel: even if it is not within the root, it is usually a stressed vowel\(^5\) (cf. Majors’ (2006) cross-linguistic study on the relation between phonetic vowel-to-vowel coarticulation, stress and vowel harmony). On the contrary, here the triggering vowel is unstressed

3. A relation with metrical structure seems to exist: harmony is limited to post-tonic vowels, thus its domain depends on word stress position. Metrically-conditioned vowel harmony domains are all but infrequent (for instance the scope of [+ATR] harmony in Catalan loanwords is, in most cases, one trochee (Cabré 2009, this volume)), but what is the exact nature of the relation in the dialects under discussion here?

4. The target of vowel harmony requires an explanation as well: as the trigger is usually ‘strong’, so the target is usually ‘weak’ (unstressed, in an affix, and so on). Why should the penultimate vowel of proparoxytones be weaker (up to undergoing total assimilation) than the final vowel, rather than the opposite? Put another way, why does harmony go from right to left, rather than from left to right?

5. Why is harmony blocked by the intervocalic consonant in some varieties, and why is it that only [r] and [l] never block harmony?

In the following paragraphs it will be argued that the motivation for this harmony process, including its direction and domain, lies in the particular metrical status of proparoxytones in many Italian dialects, and that the penultimate vowel of proparoxytones is prosodically weaker than the final one. Vowel copy undergone by this vowel is due to its reduction up to complete neutralization, something that cognate varieties without vowel harmony show as well. These dialects thus provide an interesting example of metrically-dependent case of vowel harmony: they differ from the most usual instances, in which the vowel carrying word stress is the trigger of harmony. Nevertheless, an explanation with reference to metrical structure is still required.

---

5. Although metaphony, present in some of the dialects examined here – cf. Footnote 2 – is a counterexample to this generalization, and calls for an explanation. Vowel harmony systems I am aware of in which the stressed vowel assimilates (or assimilated) to an unstressed vowel – among others Catalan vowel harmony in loanwords (Cabré 2009, this volume), metaphony in Romance, Umlaut in Germanic, similar phenomena in Albanian and Dravidian – are all anticipatory: the stressed vowel always precedes the unstressed harmony trigger. In such cases the tendency for assimilation to be anticipatory rather than perseverative might prevail over the tendency which favors ‘strong’ vowels as harmony triggers and ‘weak’ vowels as targets (cf. Hyman (to appear) and §6 below). Alternatively, such processes could arise from the need for features in weak positions to be licensed by a ‘strong’, stressed vowel in order to make them more perceptible (Walker 2005).
As for the varieties which display harmony, only when a liquid consonant intervenes, this restriction may be explained as due to the absence of contrastive place features in liquids.

3.1 Previous analyses

The phenomena under discussion here are relatively well known, and some of them have been used as the empirical basis to support several theories.

On the basis of several assimilation processes in pretonic and post-tonic vowels, including the one discussed here, Maiden (1991, 1995) argues for a model of suprasegmental structure in Italian dialects: as the syllable is divided into a pre-nuclear, nuclear and postnuclear domain (onset, nucleus and coda respectively), so the word would be analizable in a pretonic, tonic and post-tonic constituent, each of them being a potential domain in the phonological processes. The type of harmony discussed here would be limited to the post-tonic domain.

Clements & Hume (1995) cite data from the Servigliano post-tonic harmony to support their model of feature geometry. To justify the assumption of a vocalic node grouping all vowel features, they write:

By grouping all place and aperture features of vocoids under the vocalic node, we predict that all these features should be able to spread freely across intervening consonants, even if they are specified for place features of their own. This is because consonants (at least those with no secondary articulations [...] ) have no vocalic node that would block them.

There is considerable evidence that this prediction is correct. An example can be cited from the Servigliano dialect of Italian.

[...] Crucial to the point at issue, [in the Servigliano dialect] all consonants are transparent, whatever their places and manners of articulation. (Clements & Hume 1995:283–284)

They represent vowel and consonant place features as in (5) (all higher and some lower nodes omitted):

\[
\begin{array}{ll}
\text{Consonants} & \text{Vowels} \\
\text{C-place} & \text{C-place} \\
\text{[labial]} & \text{vocalic} \\
\text{[coronal]} & \text{aperture} \\
\text{[dorsal]} & \text{[open]} \\
\end{array}
\]
According to Clements & Hume (1995), Servigliano vowel harmony can be represented as in (6). Since consonants do not have a vocalic node, it is expected that they are transparent to harmony.

\[
\begin{array}{c|c|c|c|c|c|c}
& V & C & C & V & \# \\
\hline
C-place & C-place & C-place & & & \\
\hline
\end{array}
\]

vocalic node

Obviously, this model is valid not only for Servigliano, but it is assumed to be universal: consonants (at least those without a secondary articulation) always lack a vocalic node, so the prediction is that in no language can they block total harmony which according to this representation, is precisely the spreading of the vocalic node.

Nibert (1998) is interested mainly in Servigliano metaphony (cf. (1) and Footnote 2), but she also provides a formal rule for vowel copying (Nibert 1998:77).

Walker (2006) discusses Servigliano vowel harmonies (which include three other processes as well, i.e. metaphony, pretonic vowel raising and vowel copy in proclitics). She argues that post-tonic vowel copy targets the prosodically weakest vowels, and is motivated by minimization of contrast in weak syllables (which are perceptually difficult), improvement of constancy of duration in sequences of unstressed syllables, and maximization of V-to-V coarticulation. Leftward directionality is due to the constraint \texttt{idcnt-10(v-feature)-Right} (“the rightmost association of a [vowel] feature is faithful” (Walker 2006:8)), which preserves the features of the rightmost input vowel in the output.

3.2 Discussion of the previous proposals

A detailed discussion of all these proposals would lead us too far afield, since most of them also base their arguments on other varieties that are not discussed here, their goals differing from mine. But such goals in part intersect the questions raised in Section 3, so at least a brief comment is in order.

As for Clements & Hume (1995), harmonies in which consonants interact pose a problem for feature geometry, or at least for models of feature geometry, assuming that all vowel features are dominated by a vocalic node which is absent in consonants. They predict that consonantal features should never interfere with processes spreading the vocalic node (that is, processes creating total vowel harmony, since they spread every vocalic feature) because this node is not present in consonants. This is true for the Servigliano dialect, which Clements & Hume (1995) mention, but dialects such as Umbertide (3) or Sant’Oreste (4) – which are otherwise similar to Servigliano – are a problem in this respect: all consonants but liquids can block harmony, contrary to what is predicted by their feature geometry model.
Maiden’s (1991, 1995) proposal is interesting, but his definition of pre-tonic and post-tonic vowels as prosodic domains leaves their relation with other well-established prosodic constituents unanswered. For example the first syllable of a foot can carry word stress, hence it would be in Maiden’s nuclear domain, while the second syllable would be part of the post-tonic domain. A foot would thus be split into two superordinate prosodic domains, contrary to the ‘Strict Layer Hypothesis’ (Nespor & Vogel 1986) he himself assumes: constituents are exhaustively dominated by the immediately superordinate constituent, and constituents dominate only whole subordinate constituents.

As for Walker (2006), her hypothesis that one of the motivations for vowel copy is the improvement of durational uniformity has not been confirmed by the data reported in §4.1: in proparoxytones the penultimate vowel is significantly shorter than the last. On the other hand, her proposal that vowel copy is motivated by the weak prosodic status of the vowels involved, which favors a reduction of contrast in weak syllables and coarticulation, is broadly similar to the hypotheses I present in the following sections.

4. Metrical structure

A question has to be answered precisely with regard to the prosodic status of the vowels within the harmony domain: why is the penultimate vowel of proparoxytones such a good target for assimilation, to the extent that all its content is a copy of another vowel?

Maiden (1988) suggests that several vowel harmonies in Italian dialects could be due to different levels of stress intensity, reflected in different degrees of neutralization (although Maiden (1991) partially rejects this idea). I want to argue that in many Italian dialects the penultimate vowel of proparoxytones has a metrically weak status, which has caused an extreme degree of reduction and neutralization, making this context extremely sensitive to the influence of adjacent vowels.

Processes of reduction in the penultimate vowel of proparoxytones had to be at work already in Late Latin, causing syncope of this vowel (compare for example Classical Latin viridis ‘green’ with verde in Spanish, Portuguese, Italian). In the so-called Appendix Probi (a 3rd or 4th century A.D. prescriptive text containing a list of

---

6. If disyllabic feet are adopted. As one reviewer observes, Maiden does not explicitly claim that feet in the forms in question are disyllabic. Thus, my objection applies if a usual assumption about metrical structure is made, but not necessarily to Maiden’s analysis.
'erroneous' forms, actually in most cases those undergoing diachronic change) we find *calda* listed as the 'incorrect' pronunciation of *călĭda* 'hot-F', *veclus* for *vĕtĭlus* 'old-M', and so on.

In the dialects of Italy there are cases of syncope and reduction other than those common to all Romance languages. In the dialects of North-Western Italy syncope is very widespread, giving rise to consonant clusters otherwise rare in Romance: e.g. Lat. *fēmina* > ['fumna] 'woman' in Piedmontese, *cŭbitus* > ['gumde] 'elbow' in Emilian (Rohlfs 1966 [1949]:171–173). In central and southern dialects, the degree of syncope is similar to that of languages such as Italian or Spanish, but, interestingly, at least in some of the dialects under discussion here, reduction sporadically reached more words than in Italian: e.g., ['skelto] 'skeleton' (cf. It. ['skeletro]) in Umbertide.

Furthermore, even if syncope was not a regular and pervasive process in North-Eastern, Central and Southern Italy, in many varieties a similar phenomenon took place involving the neutralization of all vowel contrasts (fairly obviously, neutralization and syncope can be seen as two stages of the same reduction process, syncope simply being the most radical form of the reduction of a weak vowel).

In Friulian, the penultimate /a/ and /e/ of Latin proparoxytones merged with /i/: *orphănus* > [’warfin] 'orphan', *sabbăta* > [’sabide] 'Saturday', *stomăchus* > [’stomi] 'stomach', *iuvĕnis* [’ʤovin] ‘young’ (Vanelli 2007:54–55). In Veroli (Southern Lazio), only [ə] is possible as the penultimate vowel of proparoxytones: [’maɲːe] ‘eat yourself!’ vs. [maɲːa’telːu] ‘eat it yourself!’ (Vignoli 1925:18, 23). Tuscan too, which has a non-etymological [a] in several proparoxytones (as [’ʤovane] < *iǔvĕnis*, [’indako] < *indĭcus* ‘indigo dye’), probably had a stage in its history when all post-tonic vowels of proparoxytones were reduced to this vowel (cf. Tuttle (1974) for this proposal; vowels other than [a] would have been re-introduced by later processes).

In a fairly high number of dialects in Central Italy, only one vowel can be found in this position still today, but it is either [i] or [e] (in some dialects only one of the two vowels is always found, whereas in others both can be found in proparoxytones). Significantly, it happens in varieties without vowel harmony, although in areas neighbouring those under discussion here: for example in Cortona, only a few kilometres from Umbertide, we find [’sabeto], [’skatela] (Italian [’sabato], [’skatola]). It is also significant that this neutralization is very common in varieties in which harmony interacts with consonants: either [i] or [e] is the vowel present in all the proparoxytones without a liquid as their last consonant (cf. (3b) and (4a) above, where the vowel is always [i]).

What is relevant here is the motivation for reduction in the penultimate vowel of proparoxytones. The fact that more prominent metrical positions allow a wider array of contrasts than less prominent ones is uncontroversial; in many
(if not most) languages, contrasts possible in stressed vowels are neutralized in unstressed vowels.

As for the dialects of Italy, it was reported in §2 that many of them contrast /e, o, ɛ, ɔ/ in stressed vowels, but only [e, o] are possible in an unstressed position. If weaker metrical positions allow fewer contrasts, the extreme degree of neutralization in the penultimate vowel of proparoxytones should be associated with a weak position in the metrical grid. But whether all unstressed vowels are equal, or different degrees of stress and metrical prominence exist among the vowels which do not carry word stress, is a controversial issue in the analysis of Italian dialects. In other words, is there a difference between vowels bearing secondary stress and unstressed vowels? Is there a hierarchy of strength, other than stressed vowel > unstressed vowels? This is one of Maiden’s (1988, 1995) conclusions, mainly on the basis of vowel harmony and vowel neutralization processes also discussed here; moreover, Maiden (1997:10) observes that in Italian dialects “[t]he extent of neutralization is never greater to the left of the stressed vowel than to its right [...] [and] [t]he effect of harmony is also often greater to the right of the stressed vowel”.

This also seems to be consistent with diachronic phenomena. If we look at the diachrony of Italian dialects, Tuscan for instance, we can observe cases of syncope in pretonic vowels, too. The vowel immediately preceding word stress frequently underwent syncope (7), especially if it was not [a].

(7) Tuscan (Rohlfs 1966 [1949]:178, Tuttle 1974:452–453)

\[
\begin{align*}
*\text{longitânu} & > \text{lontano ‘far away’} \\
\text{computâare} & > \text{contare ‘to count’} \\
\text{ululâre} & > \text{urlare\textsuperscript{7} ‘to scream’} \\
\text{septimâna} & > \text{semmana ‘week’ (Old Tuscan)}
\end{align*}
\]

An unstressed vowel was lost or reduced to [a] in intertonic position, between a secondary stress and the word stress. In terms of metrical structure, a weak syllable between the word stress and a secondary stress was deleted or neutralized (8). The weak becomes weaker (in some cases, so weak that it disappears), in a natural and typologically widespread pattern.

(8) \[
\begin{array}{cccc}
\times & \times & \times \\
\times & \times & \times & \times
\end{array}
\]

lon gi ta nu lon ta no

7. The presence of [r] is due to an intermediate stage *urulare, with consonant dissimilation; indeed, urulare ‘to scream’ is attested in Sardinian (Tuttle 1974:453).
It is tempting to relate pretonic syncope (or reduction) to the post-tonic syncope (or reduction) in the penultimate vowel of proparoxytones: as the former process deleted a weak vowel immediately before word stress, so the latter deleted a vowel immediately after word stress (or neutralized contrasts). Indeed, in several dialects the latter process appears to be more regular and widespread than the former. But syncope took place in pretonic position mostly when a secondary pretonic stress was present; for post-tonic syncope in proparoxytones to become a mirror reflection of pretonic syncope, there should be a secondary stress on their last syllable.

4.1 Post-tonic vowel duration in proparoxytones

Given a tendency towards trochaic strong/weak rhythm in the dialects under discussion here, we would indeed expect a secondary stress two syllables after the word stress in proparoxytones. This is what Camilli (1965) and Lepschy & Lepschy (1981) argue for Italian, while other authors (e.g., Bertinetto (1981)) are skeptical about the very existence of (phonological) secondary stress in Italian.

Similar to Italian, in the dialects investigated here, vowel length is never contrastive but vowels in stressed open non-final syllables have increased duration. Obviously vowel duration is not the only physical correlate of stress in these dialects, but it can be reasonably assumed to be the most significant (cf. the results in Bertinetto (1981) with regard to Italian). For this reason, vowel duration was the parameter measured. Data from one of the above-mentioned varieties (Umbertide) show that in proparoxytones, the penultimate vowel is shorter than the last one. Data from two speakers were recorded: each produced a set of thirteen words, with three repetitions for each item, within the carrier sentence [ˈdiko ˈpja nin pjaˈnino] ‘I say <word> slowly’. Twelve words had antepenultimate, and one had pre-antepenultimate stress (this word included two enclitic particles). All vowels, both penultimate and final, were in open syllables (it was not a choice, rather it was imposed by the kind of data examined: in these dialects virtually all lexical words end in a vowel, and proparoxytones having a penultimate closed syllable are extremely rare).

The recorded utterances were analysed with Praat (Boersma & Weenink 2007); I measured vowel duration from onset to offset. In measuring it, the following criteria were used: for vowel onset, I measured from the second vowel cycle where formant

---

8. One reviewer correctly points out that we might not expect this secondary stress if extrametricality or nonfinality were assumed for antepenultimate stress, and asks whether this analysis appeals to either of these notions. Although there is no space here to fully expound the metrical principles behind my analysis, I believe neither extrametricality nor nonfinality are necessary (or even useful) assumptions to represent stress in these dialects.
structure was clearly visible, and for vowel offset, I measured to the last peak where formants were clearly visible.

The mean duration of the penultimate vowel was 78 milliseconds for speaker SC (standard deviation 11 ms) and 73 ms (sd 13 ms) for speaker GB, while the mean duration of the final vowel was 110 ms (sd 12) and 103 ms (sd 17) respectively. Therefore, in proparoxytones the final vowel is on average 40% longer than the penultimate (more precisely, the ratio is 1.406:1 for speaker SC and 1.4:1 for speaker GB).

5. Metrical representation and harmony

Syncope, neutralization and vowel duration all suggest a representation of proparoxytones as below:

(9)  *
    * *
    ... * * *
    ... σ σ σ

Since it can be reasonably supposed that the basic foot in these dialects is the syllabic trochee (× .), in proparoxytones it is predicted (if extrametricality is not adopted, cf. Footnote 8) that the last syllable would form a degenerate foot: (× .) (×). It is useful to point out that, since monosyllables are licit words in these dialects, there is independent motivation for degenerate feet. The metrical structure assigned to proparoxytones is shown in (10).

(10)  ×
     (×  ×)
     ... (×  ×) (×)
     ... σ σ σ

If this representation is correct, syncope, reduction to schwa (as in Veroli, see above), and neutralization of vowel contrasts would all be related to the prosodically weak status of the penultimate vowel of proparoxytones: being weaker than the final vowel, it would be subject to more radical processes of reduction than the latter. In this process of reduction the step before syncope would have been a complete loss of contrast, the short and weak vowel being unable to preserve phonemic oppositions. Given this situation, there are two logically possible developments: in being completely unspecified, the vowel either has a default realization or it receives its content from another vowel (since it is weak, it is expected to be prone to assimilation). This is precisely the outcome found in the dialects examined: in some of them the penultimate vowel of proparoxytones is always [i] or [e], in the others, it is a copy of the following vowel. In varieties displaying vowel copy, a word like [ˈsigoro] ‘cigar’ has
a completely unspecified penultimate vowel which receives its content from the
place node of the following vowel.

(11) /s i g Ø r o/ ['sigoro]

CV CV CVºººººº root node

Directionality of vowel harmony: The interaction of phonetic and phonological tendencies

The representation of proparoxytones in (9) and (10) can explain why their penultimate vowel is the target of harmony. But why is the final vowel the trigger? After all, the preceding vowel carries the word stress, and is thus more prominent and stronger than the secondary stress of the final vowel.

Hyman (to appear) has recently proposed an explanation for directional asymmetries in vowel harmony systems as being the result of interaction and conflict between two tendencies. The first has a phonological motivation, since it is caused by the different prosodic status of roots and affixes: in affixes there is usually more vowel reduction than in roots. The second is closer to phonetics: all other things being equal, segmental assimilatory processes and VCV coarticulation effects appear to be more robustly anticipatory than perseverative (Hyman, to appear:16). They have the following consequences for directionality in vowel harmony:

i. vowel reduction in affixes favors root-controlled harmony on them (hence right-to-left root-controlled harmony on prefixes, and left-to-right on suffixes)

ii. the anticipatory nature of segmental assimilation favors right-to-left suffix-controlled harmony

Together, the two tendencies correctly predict that prefixes rarely trigger harmony on the root, whereas the opposite is frequent. As for suffixes, the first tendency seems to be stronger: “[s]uffix controlled V[owel]H[armony] is less frequent than root-control, presumably because roots do not as readily undergo reduction as do affixes” (Hyman, to appear:24). But if a root vowel undergoes more reduction than an affix vowel, the tendency towards anticipatory assimilation would no longer be overridden, and could operate. This model precisely predicts the actual data found
in the dialects discussed here: anticipatory assimilation from the final vowel, which in most cases is a suffix but has not undergone any significant reduction, to a weak root vowel.

The solution proposed has so far taken into account proparoxytones only. When, due to the presence of one or more clitics, stress is before the antepenultimate syllable, two (cf. the last two Examples in (1b)) or even three vowels are subject to harmony. It has already been observed (by Canepari (1986:29–30) among others, with regard to Italian) that in such words the final vowel carries a secondary stress. In this case there are two intertonic vowels, both weaker than the word-final vowel, instead of only one: the expected outcome is the final vowel’s features spreading to all the preceding weak segments.

6. Liquid consonants

In several dialects the possibility of harmony depends on the presence of a liquid consonant. In this respect, it must be pointed out that the dialects examined here (unlike Italian and other Italian dialects) lack the consonant /ʎ/ in their inventory; the north-western border of the area where harmony is present is also the border between varieties with and without /ʎ/. In a vast area south-east of this border, cognate words have /j/ instead of /ʎ/ (compare for instance, Florentine [fiʎːo] ‘son’ with Serviglianese [fijːo]).

Moreover, the Garfagnana area in Northern Tuscany includes a small isolated outcrop of the harmony process under discussion, and whereas in most of Tuscany /ʎ/ contrasts with /l/, in some areas of Garfagnana /l/ is the only lateral. Interestingly, a palatal stop can be found here instead of /ʎ/ (for instance, ‘son’ is [fiɟːo]), and some villages in the same area display vowel harmony which takes place only between liquid consonants.

The absence of /ʎ/ implies that in these dialects there are only two liquid consonants, /r/ and /l/, which are predictably alveolar. Phonologically, their place of articulation need not be specified, since it does not distinguish them from other liquids; once they are specified for manner, this is enough to contrast them with any other consonant. The latter, on the contrary, require phonological specification of place: there are three nasals /n, m, ɲ/, two semivowels /j, w/, several fricatives /s, f, ŋ .../, stops /t, k, p .../ and affricates /ts, tj .../.

9. Note that the inventory of word-final unstressed vowels is either identical to that of unstressed vowels in any other position, or only lacks [u] (cf. Footnote 3).
Therefore liquid consonants, the only segments which in several dialects do not block vowel harmony, are also phonologically placeless. Since vowel copying spreads all the place features of a vowel, the hypothesis of contrastive underspecification of place features in liquids offers an explanation for their transparency: vocalic place features are not blocked by liquids because the latter do not have place features (12a), whereas all other consonants, which must be specified for place, possess a place node that impedes spreading (12b). In a feature geometric model, a straightforward way of expressing this hypothesis would be to represent the place nodes of vowels and consonants on the same plane.

\[
\begin{align*}
(12) & \quad \text{a.} \quad \begin{array}{ccc}
V & r & V \\
\ldots & \ldots & \ldots
\end{array} \\
& \quad \begin{array}{c}
\text{place}
\end{array}
\end{align*}
\]

\[
\begin{align*}
& \quad \begin{array}{ccc}
V & b & V \\
\ldots & \ldots & \ldots
\end{array} \\
& \quad \begin{array}{c}
\text{place place}
\end{array}
\end{align*}
\]

Feature underspecification has been subject to much debate. Objections against underspecification in general have been made on the basis that the principles for establishing whatever features have to be left unspecified are not clearly determined, thus allowing for arbitrary choices and many equally valid alternative solutions; also, many redundancy rules are required to fill in predictable values; and it raises technical problems such as the possibility of ternary values (plus, minus and zero), etc. Space does not permit me to address these points, so I will only mention the possibility that several problems might be solved by adopting a contrastive feature hierarchy to determine contrasts (e.g., Dresher 2003a, 2003b), while other problems could be avoided if unary features were used (Steriade 1995:147–157).

A more specific potential counterargument to my analysis precisely involves the transparency of liquids. Steriade (1995) objects that the transparency of liquids may depend on their being sonorants, rather than on the lack of a distinctive place of articulation. She observes that there are many languages in which the only segments transparent to vowel-to-vowel assimilations are sonorants, whereas, although /s, z/ are frequently the only fricatives occurring in a language, they do not behave as transparent with respect to assimilation rules: “the syndrome of liquid placelessness [...] has no connection to issues of distinctiveness: the liquid is transparent not because its place features are predictable from its stricture, but for different reasons, which remain still unclear” (Steriade 1995:146).
If /s, z/ do not contrast with other fricatives, they seem to contradict the prediction of contrastive underspecification: their place of articulation is predictable from their stricture, so they should be transparent, yet no known vowel harmony is transparent only across /s, z/. Nonetheless, we may wonder whether this situation depends on the groundlessness of contrastive underspecification, or rather on the special status of /s, z/. In many respects /s, z/ are exceptional: before stops (in complex onsets) as well as after them (in complex codas), they cause the most frequent type of sonority scale violation; they are frequently the only consonants which can occur as the first segment of complex onsets or as the last of complex codas; /t, d/ apparently share the same place of articulation, but when they are changed to fricatives by lenition processes, in numerous languages they become [θ, ð] rather than [s, z]. If /s, z/ have an unclear phonological status, it could be asked whether they are reliable counterexamples to contrastive underspecification.

Moreover, alternative explanations for transparency of liquids in the dialects discussed here are not obvious. The other sonorants – that is, nasals and semivowels – block harmony (e.g., [ˈpetine] in (3b), [ˈtrapinu] in (4a)), thus Steriade’s hypothesis is not viable here.  

6.1 Feature activation

The varieties in which harmony takes place across any consonant do not have /ʎ/ in their consonant inventory either, thus assuming underspecification their liquids would be underspecified for place as well. Nonetheless, their harmony is not influenced by consonantal place of articulation at all. More generally, dialects with and without restrictions on the type of intervocalic consonant in the harmony domain are in

---

10. One reviewer suggests that the transparency of liquids could be independent of their place (under)specification. Liquid consonants not impeding spreading of vowel features occur frequently cross-linguistically (as laryngeals: “it is noteworthy that it is often h or a glottal stop that intervenes between the conditioning vowel and the [reduplicated] vowel in total harmony” (Aoki 1968:142)). Their transparency could derive from a hierarchy of ‘consonantality’, liquids (and laryngeals) being low in this hierarchy, thus closer to vowels and therefore less prone to block vowel-to-vowel assimilation. More generally, the reviewer suggests the same hierarchy could be the reason for other phenomena as well (e.g., raising of low-mid vowels when they are unstressed (cf. §2), which could be analyzed as a movement towards increased ‘consonantality’, as vowels become less sonorous and more closed). While interesting, this idea would not explain why semivowels – which are intuitively more similar to vowels than any other segment – block harmony (e.g., the last Example in (3b)). Besides, a decrease in sonority and ‘vocalicity’ is not necessarily associated with vowel reduction, since in some languages mid vowels which lose stress are lowered rather than raised (in Belorussian, for example – cf. Crosswhite (2004); to remain within the domain of Italian dialects, in some varieties of Ticinese Lombard (Salvioni 1885–86:206–207) all unstressed front vowels have merged into [a]).
other respects very similar, and sometimes their areas are contiguous. But if in both groups of dialects place features are on the same tier, we should always have harmony blocked by non-liquid consonants; and if place features were not on the same tier, we should never have any blocking. Since the difference cannot reside in the phonological inventories of the two groups of dialects, some sort of parametric variation must be admitted to explain the ambivalent behavior of liquids.

Clements (2001) offers a proposal along this line: to account for problematic cases of transparency, he assumes that, while the geometrical organization of features is universal – with only one hierarchy of features and feature nodes available to all languages – a specific language is free to activate only one subset of this hierarchy. All constituents in the feature representations of a given language must be constituents of the universal feature hierarchy (Clements 2001:98), but not all constituents of the universal feature hierarchy must be active in a given language. As a result, only some features will be autosegmentalized, whereas the other features are simply bound on the root node.

Adopting this view, we can still assume that feature geometric organization is universal, but only some of the dialects discussed above activate their place nodes in both consonants and vowels: in this case, represented in (12) above, place features of consonants – if present – block harmony. In the other dialects only vowels activate their place node, while place features in consonants remain attached to the root node (13). Even if consonantal features are present, as in (13a), the spreading of a vocalic place node is not blocked: consonantal features are not projected on the autosegmental tier of the vowel, and thus cannot interfere with it.

(13) a. V b  V #
   ...  ...  ...
   root root
   ...  ...
   [labial] [labial]
   place place

b. V r  V #
   ...  ...  ...
   root root
   ...  ...
   place place

7. Conclusion

Apparently the vowel harmony process discussed here displays some typologically uncommon traits. Firstly, the domain of harmony is metrically conditioned (it is restricted to the post-tonic vowels), but harmony is triggered by an unstressed vowel
rather than a stressed one; secondly, in most cases the harmony trigger is part of a suffix or clitic, while its target is frequently in the root. Furthermore, the leftward directionality of harmony and the limitation of its domain to the post-tonic vowels of proparoxytones need an explanation, rather than simply being stipulated.

Once several phonetic and phonological properties of proparoxytones are taken into account, data converge to suggest that their penultimate vowel has a weak prosodic status. This weakness makes it a good target for assimilation, and a universal phonetic tendency towards anticipatory assimilation accounts for the leftward direction of harmony.

Blocking of harmony by non-liquid consonants in several dialects is another problem. But since their liquids are the only two consonants without contrastive place features, resorting to underspecification can explain their transparency. On the other hand, dialects without consonantal interference in harmony require place specification in all the non-liquid consonants as well; assuming a relativized notion of universal feature geometry, their consonant place specifications do not have to interact with vowel features.

References


Vowel reduction and vowel harmony in Eastern Catalan loanword phonology*

Teresa Cabré
Universitat Autònoma de Barcelona

The aim of this paper is to account for the phonological adaptation of loanwords in Eastern Catalan. As the phonology of these new words deviates from that of the native Catalan vocabulary set (with a certain amount of variation among speakers), the new phonetic features would seem to be borrowed from Spanish. We suggest that a new phonology has emerged whose purpose is to identify loans among the lexicon, the most striking element of this phonology being a harmony effect on stressed mid vowels in the presence of post-tonic [+ATR] mid vowels. The existence of unstressed [+ATR] mid vowels [e, o] in Eastern Catalan has been previously interpreted as lexical exceptions to vowel reduction (Fabra 1912 and Mascaró 2002, among others). However, the phonetic variation in the new lexicon is analyzed here as being fully consistent with Catalan phonology within the theory of lexical strata (Itô & Mester 1999).

1. Introduction

This study examines the strategies that Catalan has developed in the incorporation of loans and offers an optimality-theoretical account of the adaptation process of these new words in Eastern Catalan. As the result of being a minority language, Catalan has historically borrowed a certain number of words from neighboring languages, particularly Spanish. This borrowing process has been accelerated on a massive scale over the last 100 years, with increasing numbers of loans coming

*This work was partially presented at the XIVth International Colloquium on Catalan Language and Literature (Budapest, September 2006) and at PaPI 2007 (Phonetics and Phonology in Iberia, Braga). I would like to thank the audience at these gatherings as well as Maria Ohannesian, Pilar Prieto, Donka Steriade, Jaume Solà and especially Michael Kenstowicz, for their useful observations and comments. Thanks are also due to the informants who patiently answered my questions about a long list of loans. I am grateful to the reviewers for their suggestions and advice, which has allowed me to improve certain aspects of this paper. This research is funded by grants 2005 SGR 00753 from the Generalitat de Catalunya and HUM2006-01758/ FILO and HUM2006-13295/FILO from the Ministerio de Educación y Ciencia-FEDER.
from languages other than Spanish, especially English. The consequence has been
the emergence of a new phonology.

The more recent loanwords show striking phonetic differences relative to
the native vocabulary in terms of the vowel system of Eastern Catalan. Most
notably, speakers of this Catalan variety identify new words through the pro-
nunciation of unstressed mid vowels. Unlike what happens with native words, in
loans, unstressed mid vowels never reduce to schwa (from [–post]) or [u] (from
[+lab]), instead surfacing as [+ATR], e.g., C[o]p[e]nhagu[e], N[e]pal, V[e]rsall[e]s,
In addition, a vowel harmony effect arises: mid vowels in stressed position are pro-
nounced as close mid vowels when in the presence of post-tonic close mid vowels, e.g.,
pesto [ˈpesto], Boston [ˈboston], gueto [ˈgeto] ‘ghetto’, Oslo [ˈoslo], Volvo [ˈbolʃo],
Tere [ˈteɾe] ‘Teresa’, secre [ˈsekre] ‘secretary’, cole [ˈkole] ‘school’, etc. This paper ana-
lyzes why reduction fails to occur in such loanwords and examines the scope of
and variation in the vowel harmony phenomenon. Our main goal is to provide a
unified account of the different stages in the adaptation process from within the
theory of lexical strata (Itô & Mester 1999). If we accept a hierarchy of foreignness,
we can account for the phonology of Catalan vocabulary as a consistent whole,
notwithstanding the phonetic variation observed.

Interest in loanword phonology has increased in the last fifty years mainly
because of its relation to linguistic change and language acquisition. All languages
borrow the vocabulary they need from neighbors and they develop adaptation
strategies in different ways. Nevertheless, the new vocabulary frequently pre-
serves some phonological (and morphological) aspects that distinguish it from
the native vocabulary.

When one language (L1) borrows words from another (L2), a conflict occurs
between trying to preserve the phonological information of L2 on the one hand
and trying to satisfy the phonological conditions of L1 on the other. Speakers of
L1 tend to keep the more salient characteristics of L2 words while modifying less
prominent segments to bring them closer to L1 (Steriade 2001; Kenstowicz 2001).
In other words, speakers distinguish loans by pronouncing them differently from
how they would pronounce them if they were native words, and speakers do this
more o less consciously.

When linguistic borrowing occurs on a very large scale (as tends to happen
with minority languages such as Catalan), the L1 may adopt elements of the L2’s
phonology. In this process of incorporation the intrinsic system of the L1 undergoes
some extrinsic intrusion and change. These modifications are consistent with nat-
ural, attested processes and conditions in many languages. It is therefore possible
to attribute these conditions to Universal Grammar (UG), with the implication
that adult speakers can always resort to them (Kenstowicz 2001).
The particular situation of Catalan under the influence of Spanish has meant that it adapts any sounds in the foreign vocabulary to its own phonology in the same way that it has done for Spanish, at least apparently. For example, modern Catalan has assimilated the Spanish segments [x] and [θ], and uses them in all borrowings of any language which contains similar segments. This is in contrast with the situation before the 20th century, when, as far as we know, Spanish loans were completely assimilated into Catalan phonology, with more or less the same being true of all other foreign vocabulary (Bruguera 1985).

The assimilation of L2 phonological segments raises a question: if loanwords exhibit a different phonology from the native vocabulary, how many phonologies (i.e. grammars) do we have and how many lexicons correspond to each language? Variation among speakers adds another level of difficulty to this already complex map. However, it may be that the adaptation of loans sometimes allows us to discover certain regularities or default elements that are hidden in native vocabulary because of faithfulness to the input.

Our task in this paper is to explain this kind of phonological variation within a language's lexicon and, if possible, describe the factors or processes that can phonologically differentiate loanwords from native vocabulary. Specifically, the aim of this paper is to account for the vowel phonology of loanwords in Eastern Catalan, including the stressed and unstressed vowel systems. The unstressed system differs from the native system in its resolution of mid vowels: since unstressed mid vowels in loans never reduce to schwa or [u], as they would in native vocabulary. Instead, they surface as [+ATR], as in the loans C[o]p[e]nhagu[e], Ibs[e]n, básqu[e]t 'basketball', t[e]mpura 'Japanese dish', c[o]ns[o]mé 'consommé', B[o]mbai 'Bombay', Xil[e] 'Chile', etc. The stressed system in loans follows the general tendency of the language to open mid vowels, except for those words with a close mid vowel in post-tonic position. In these cases the [+ATR] value of the post-tonic vowel spreads left onto the stressed mid vowel, thus triggering a vowel harmony process which is completely new in Catalan phonology, as in Penèlope [peˈnelope], UNESCO [uˈnesko], Beethoven [beˈtoβen], Volvo [ˈbolβo], Irene [iˈrene], polièster [poˈljeʃtet] 'polyester', profe [ˈpɾoфе] 'teacher', etc. These phenomena are subject to a certain amount of inter-speaker variation, not only in blocking vowel reduction but also in spreading [+ATR] to the stressed vowel within the scope of this vowel harmony process. Finally, we adapt Ito & Mester’s (1999) model according to which lexical items are organized in a core-periphery structure. A lexicon divided in

---

1. Formerly, these two segments from Spanish words were always rendered in Catalan as [k] and [s] respectively, and vowels were also adapted: ojo! Sp. [ˈəko], Cat. [ˈsku] ’eye’ (meaning ’Be careful!’), Rodriguez Sp. [ɾoˈðɾiɣəθ], Cat. [ɾuˈɾiɣəs] ’Spanish surname.’
strata provides us with a feasible explanation as to the different subphonologies and variation found in Catalan loans, as well as the different stages of loan adaptation.

This paper is organized as follows. Section 2 presents the extended vowel system of Eastern Catalan that includes native and non-native vocabulary. Section 3 focuses on the earliest attested words showing blockage of vowel reduction in unstressed position and the subsequent expansion of the phenomenon. Section 4 shows the preference for [–ATR] mid vowels in stressed position in loans. Section 5 describes the vowel harmony process and accounts for it by means of constraint ranking. Finally, section 6 links the dialectal variation found to stages in the nativization process and adapts the lexicon strata theory to the three main varieties of non-native words found in Eastern Catalan.

2. Catalan vowel system

The phonology of Eastern Catalan loanwords shows striking differences from the phonology of native vocabulary. All these differences have been traditionally attributed to the influence of Spanish (Fabra 1912; Recasens 1993, etc.) due to the fact that vowel reduction to schwa or labial high vowel is blocked when mid vowels appear, yielding outcomes with close mid vowels, as in the Spanish unstressed system: sin[e] di[e], v[e]det ‘starlet’, Pinotx[o] ‘Pinocchio’, Nix[o]n. Mascaró (2002:110) points out that exceptions to vowel reduction may appear in a great variety of lexical items without any specific phonological context. In addition, some stressed mid vowels do not follow the distribution of Catalan native vocabulary, but instead are pronounced as close mid vowels as well, which once again appears to be consistent with the Spanish system: gu[e]t[o], c[ó][e], Ir[é]n[e], N[é]st[ó]r, B[é]th[ó]v[é], B[ó]st[ó]n (Cabré 2002).

Concerning the stressed vowels, studies have been carried out in this area but none of them make any reference to the harmonic process. Badia (1968, 1970) and Pi-Mallarach (1997) merely emphasize the generalized tendency of Catalan to open all stressed mid vowels in loanword adaptation. Fabra (1906:22) refers only to front mid vowels: “Les noms empruntés possédant un e tonique, sont prononcés par règle général avec un e ouvert”. In fact, many borrowings – whether from Spanish or any other languages – involving mid vowels in either stressed or unstressed position have been adapted to the Catalan native system, as is shown in (1).

(1)  robo [ˈɾɔbɔ] ‘theft’  sòtano [ˈsɔtɔnu] ‘basement’
bulto [ˈbuʎtu] ‘lump’  enterro [ənˈtuɾu] ‘burial’
regalo [ɾəˈɣalʊ] ‘gift’  pago [ˈpaɣu] ‘payment’
desaiguè [daˈzaʝwɔ] ‘drain’  laringe [laˈɾinʒə] ‘larynx’
merci [ˈmersi] ‘thanks’  futbol [ˈfuβɔl] ‘football’
In addition, certain Spanish proper names such as Francisco, Fernando, Celestino or Jacinto have not only been completely assimilated phonologically but also used as bases for the traditional Catalan truncation process (Cisco ['sisku], Nando ['nandu], Tino ['tinu], Cinto ['sintu]) (Cabré 1993). Therefore, it is not that Eastern Catalan simply borrows and uses the Spanish vowel system for loans, but rather that Eastern Catalan has introduced certain changes to its phonological system that yield words that sound closer to Spanish. A more detailed examination of the phonology of loans shows that the Eastern Catalan extended system for loans is quite removed from the five-vowel Spanish system, as we can see in (2).

\[
\begin{array}{c|c|c|c|c|c|c|c}
\hline
& \text{Stressed system} & \text{Native unstressed system} \\
& i & u & e & o & e & o & a \\
\hline
\text{Unstressed system for loans} & i & u & e & o & e & o & a \\
\hline
\text{Spanish (stressed and unstressed system)} & i & u & e & o & a \\
\hline
\end{array}
\]

The stressed vowel system is common to both native vocabulary and loanwords, because we can find both [+ATR] and [–ATR] mid vowels, but their distribution within each sublexicon is not exactly the same. The tendency to have [–ATR] stressed mid vowels in adapted loans, as the authors referred to above (Fabra, Badia and Pi-Mallarach) state, follows the general pattern of Catalan. If we look at the stress position in native words, we see a clear predominance of [–ATR] among front vowels in non-derived stems. The preference for [–ATR] stressed mid vowels is also evident in suffixed forms: more than twenty stressed suffixes contain a [–ATR] mid vowel (Mascaró 1985). In addition, all pre-stressed suffixes (all of them with a high vowel) and all the so-called suffixed forms (such as /ləg/ or /fɔn/ in fonòleg ['fuɲɔlak] ‘phonologist’ or gramòfon [ɡɾəˈmɔfun] ‘gramophone’) show the same pattern. In (3) we offer some examples that illustrate this generalization.

\[
\begin{align*}
\text{(3)} & \quad \text{bleda ['bledə] ‘chard’} & \quad \text{hora ['ɔɾə] ‘hour’} \\
& \quad \text{metro ['metru] ‘subway’} & \quad \text{lloro ['ʎɔɾu] ‘parrot’} \\
& \quad \text{col·legi ['kuləʒi] ‘school’} & \quad \text{petroli [peˈtrɔli] ‘oil’}
\end{align*}
\]

2. The most common suffixes with [+ATR] mid vowel are /eɾ/ > ['e], /ment/ > ['men], /on/ > ['ɔ], /ɔt/ > ['o], /ɔz/ > ['os]: mentider ‘liar’ (from mentida ‘lie’), sofriment ‘suffering’ (from sofrir ‘to suffer’), petitó ‘small, dim.’ (from petit ‘small’), negror ‘blackness’ (from negre ‘black’), plujós ‘rainy’ (from pluja ‘rain’).
The emergence of [-ATR] mid vowels in loans depends on the context: partial reduction of mid vowels after stress (that is [+ATR] mid vowels) generally blocks [-ATR] mid vowels in stressed position, whereas stressed mid vowels surface as [-ATR] when a post-stressed high vowel or a schwa occurs, as in the preceding examples of native vocabulary (3).

Catalan native vocabulary exhibits what Crosswithe (2004) calls a sonority-driven unstressed system, which is presented in (4). Seven stressed vowels are reduced to three in unstressed position following the vocalic prominence scale (Crosswhite 2001), according to which [i, u, ə] are the lesser prominent vowels. Consequently, they are the most suitable to appear in unstressed position in a prominent-driven neutralization system, which is what Catalan has.

(4)

<table>
<thead>
<tr>
<th></th>
<th>- back</th>
<th>+ back</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ high</td>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td>- high</td>
<td>e</td>
<td>ə</td>
</tr>
<tr>
<td>- low</td>
<td>ɛ</td>
<td>o</td>
</tr>
<tr>
<td>+ low</td>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>

In loanword phonology, the unstressed system is extended relative to that of native words. Schwa always comes from a, but we also find instances where this segment is not reduced. Vowel reduction involves the ATR feature, with the rest of the feature values remaining unchanged. The extended system is shown in (5):

(5)

<table>
<thead>
<tr>
<th></th>
<th>- back</th>
<th>+ back</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ high</td>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td>- high</td>
<td>e</td>
<td>ə</td>
</tr>
<tr>
<td>- low</td>
<td>ɛ</td>
<td>o</td>
</tr>
<tr>
<td>+ low</td>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>

We may regard the extended system as following a sonority-driven reduction (Crosswhite 2004) since low and low mid vowels change to [+ATR], with only the lesser prominent vowels arising in unstressed position, according to the vocalic prominence scale shown in (6).

(6) Vocalic prominence scale (after Crosswhite 2001): a > ɛ, ə > e, o > i, u > ə.
3. Vowel reduction blockage

We have some evidence of an extended unstressed system from the beginning of the 20th century. At that time, Catalan grammarians noted the non-reduced pronunciation in post-tonic position in some learned words in the Barcelona area, as in bas[e] ‘base’, catàstraf[e] ‘catastrophe’, class[e] ‘class’, fas[e] ‘stage’, fras[e] ‘sentence’, pest[e] ‘plague’, superfici[e] ‘surface’, etc. (Fabra 1912). This phenomenon is consistent with the behavior of another group of learned words which is made up of compounds built with two (or more) Greek or Latin elements or learned variants and created as scientific or technical terminology, even though some of them may later move into colloquial vocabulary (Mascaró 1985). They differ from other compounds because of the inflectional element –o that appears suffixed at the end of each non-final element in many compounds. Speakers are aware of this fact and often apply the extended unstressed vowel system to at least the first component. The result is a phonetic realization with unstressed [+ATR] mid vowels. It is important to stress that this pronunciation is not extended to the isolated colloquial elements: compare, for example, politico-social [polïtikοsuʃjal] ‘sociopolitical’ or sociocultural [soʃokultuɾal] ‘sociocultural’ with politic [puˈlitik] ‘politician’ or social [suʃjal] ‘social’. In (7) we show some examples of this lexical class:

(7) antropomòfic [əntrɔpɔˈmɔrfik] ‘anthropomorphic’
    monolingue [moˈnɔlingwe] ‘monolingual’
    oto-rino-laringòleg [oˈtɔrinɔlɾiŋˈɡɔlɛk] ‘ear, nose and throat specialist’
    politico-social [polïtikοsuʃjal] ‘sociopolitical’
    sociocultural [soʃokultuɾal] ‘sociocultural’
    morfo-sintàctic [morfoʃiɾtaktik] ‘morphosyntactic’


3. We can add to this group some recent words with the monosyllabic unstressed prefix co- ‘with’ such as codirector [koˈdɪɾakˈto] ‘codirector’ and coeditar [koˈðiɾaˈta] ‘copublish’. In all items in the native vocabulary with the same prefix, vowel reduction occurs, e.g., col·laborar [kuɾɔˈðʊɾaɾ] ‘collaborate’, confiar [kuɾʃiˈaɾ] ‘trust’.
B[o]d[e]l[aire], etc. Some of these words also exhibit close mid vowels in stressed position, as we will see below: c[ó]sm[ó]s, UN[é]SC[ó], pr[ó]f[e], d[é]m[ó], B[e]th[ó]v[ó]n. We must bear in mind that a large degree of variation among speakers can be observed in both processes, but in general the more recent an item, the more patent the process.

Leaving aside the idiolectal and dialectal variation we find in all these types of words, the important thing about them is that they all have something in common in that they somehow stand out in the way speakers treat them. That is, speakers do not regard these words as belonging to the native vocabulary because they resort to a phonological process that never occurs in native words.

In contrast with older borrowings, new words are not generally common words introduced through oral transmission but instead pertain to more formal registers and tend to be introduced through the written form. Speakers have to read them and incorporate them into scientific or technical speech. So the phonology of the language that lends a word is overridden by the graphic form in which this word is transmitted.

In Eastern Catalan orthographical a is usually pronounced as [ə] (closer to [a] in the area of Barcelona) both in pre-tonic and in post-tonic position, e.g., Chanel, canapé, in fraganti, Jaguar, Karamazov, Berna, Armani, or Philadelphia, and even in words with unstressed mid vowels that are not reduced like alel·luia [əleˈluja], or Antígona [əˈniˈtwayne]. And at the present time, Eastern Catalan speakers apparently perceive schwa to be closer to [a] than [e, ε]. This correspondence illustrates the psychological (and phonetic) distance between [ə] and the low vowel [a] on the one hand or the mid vowels [e, ε] on the other. We have evidence for this from the early stages of writing, in which children tend to systematically represent all schwas by the letter a. In addition, many adult speakers pronounce [e] or [ə] depending on the written form, associating e with [e] and a with [ə]. Often in Catalan, the same proper name can have different spellings, and this can determine whether the pronunciation shows vowel reduction or not: Queralt [keˈral] / Caralt [kaˈral], and Torres [ˈtores] / Torras [ˈtorəs].

We also find other instances of schwa in neologisms. Initial epenthesis in foreign words with a specific initial cluster is obligatory in Catalan. Thus, words such as Fnac ‘a chain store’, slogan, sketch, speaker, Slazenger or slalom are unfailingly pronounced with an initial schwa. Note here that schwa occurs even though no vowel is transcribed.

When mid vowels appear in written form, [e o] are the general pronunciations. (8) provides some examples of words in which unstressed mid vowels are not reduced as they would be in native vocabulary but surface instead as [+ATR] mid vowels. To facilitate understanding, only the mid vowels are transcribed:
Regarding the realization of unstressed mid vowels, it must be noted that [e o] (+ATR) are unmarked relative to [ɛ ɔ] (–ATR). As we saw in the table in (5) above, [ɛ ɔ] are blocked in unstressed position, thus changing the ATR value. The phonetic distinction of five vowels based on the written forms a, e, o, u, i has the goal of emphasizing loans by improving the equidistant positions of vowels in the vowel trapezoid, thereby making them clearer, due to the fact that there is an unambiguous correspondence between the written and phonetic forms. This can be interpreted as a faithfulness constraint that arises in loanword phonology. This constraint must be indexed to the specific stratum because the native vocabulary remains unaffected. The following hierarchy in (9) accounts for the extended vowel system in unstressed positions for a subset of the lexicon.

\[(9) \quad * \varepsilon \hat{\delta} >> \text{L–Faith (e o)} >> * \varepsilon \hat{o} \]

The constraints \( * \varepsilon \hat{\delta} \) and \( * \varepsilon \hat{o} \) force vowel reduction of mid vowels and they are active when native vocabulary is used. When one of the so-called non-native words occurs, the indexed faithfulness constraint applies, thus blocking complete reduction of mid vowels. We can also rewrite this ranking to show how the ATR feature values are crucial in this process. In (10) the new ranking is shown and the tableau in (11) exemplifies this ranking. The written form acts as the underlying form.

\[(10) \quad *[\text{–Stress, –ATR}] >> \text{L–Faith (e o)} >> *[\text{–Stress, +ATR}] \]

\[(11) \quad \begin{array}{|c|c|c|}
\hline
\text{síndrome} & *[\text{–Stress, –ATR}] & \text{L–Faith (e o)} & *[\text{–Stress, +ATR}] \\
\hline
\text{síndr}[\varepsilon]\text{m}[\varepsilon] & *!* & & \\
\hline
\text{síndr}[\varepsilon]\text{m}[\varepsilon] & & *!* & \\
\hline
\text{⇒ síndr}[\varepsilon]\text{m}[\varepsilon] & & & ** \\
\hline
\end{array} \]

The ranking presented in (10) can be considered an example of what Pater (2004) calls exceptional blocking by faithfulness, that is, when a general process of the
language is blocked in specific words. In this case, the general process is that mid vowels reduce to [ə] or [u] and the blocking yields the exceptional pattern seen in loans.4

4. **The emergence of the unmarked: The preference for [–ATR] stressed vowels**

In contrast to what was claimed by the grammarians referred to above, our data show that the tendency for Eastern Catalan to open mid vowels in stressed position in foreign adaptations has been reduced to certain specific contexts (Cabré 2006). We find open mid vowels in borrowed oxytone words ending in a consonant or glide, regardless of the feature values of the segment. Some examples of this are provided in (12).

(12) minu[ε]t  ved[ε]t  Budap[ε]st
    sk[ε]tch  estr[ε]s  Queb[ε]c
    Nova Y[ə]rk  cowb[ə]y  ad h[ə]c

If we compare these outcomes with those of the most productive stressed suffixes that yield oxytone words, we see that the adaptation of these loans follows the general tendency of the language because mid vowels generally surface [–ATR] in native words too, as in verd[ε]t ‘mold’, ungl[ə]t ‘hoof’, puj[ə]l ‘mountain’, portugu[ε]s ‘Portuguese’, nadal[ε]nc ‘relative to Christmas’, cavaller[ε]sc ‘chivalric’. This pronunciation is generalized independently of whether preceding syllables undergo vowel reduction or not. Though there is some variation, pre-tonic mid vowels generally surface as closed [+ATR] and stressed mid vowels surface as open [–ATR]: Quebec [keˈbek], Flaubert [floˈbɛrt] or [floˈbɛr], Repsol [reˈpsəl], Tolstoj [tolsˈtoj] (at least this is the most common pronunciation in the area of Barcelona).

The preference for open mid vowels in stressed position is also evident in penult and antepenult stress words, but only when a, i or u appear in post-tonic position, as shown in (13).

---

4. Catalan exhibits other examples of this typology. For example, final deletion of /n/ after a stressed vowel in nominal forms is active in native words but again, not in loans: compare camí, [kəˈmi], camins [kəˈmins] ‘path, sg/pl’ with caiman [kəˈmæn], caimans [kəˈmans] ‘caiman, sg/pl’. 
The vowel pattern of the language clearly asserts itself in this case, above all when final high vowels are involved. All stressed mid vowels from native words ending in i or u in the written form are pronounced as [–ATR] (e.g., inn[ɔ]cu 'harmless', ing[ɛ]nu 'candid', micr[ɔ]bi 'microbe', sil[ɛ]nci 'silence', petr[ɔ]li 'petrol', pr[ɛ]mi 'prize', dim[ɔ]ni 'devil', mist[ɛ]ri 'mystery', etc).

We can interpret this [+Stress, –ATR] preference as the emergence of the unmarked (McCarthy & Prince 1994) since open vowels are phonetically longer and acoustically more noticeable (Maddieson 1997) than closed ones because of their greater sonority. This is also favored by the unstressed mid vowels so that they surface as [+ATR], that is, [–Stress, +ATR] is phonetically shorter because of its lower sonority. On the basis of this generalization, we propose the constraint * [+Stress, +ATR] in order to guarantee the preference for open mid vowels in stressed position in either loanwords or native vocabulary.

(14) * [+Stress, +ATR]: Closed mid vowels are not allowed in stressed position.

The examples in (12) and (13) show that the ATR value of stressed mid vowels is independent of the presence of partial reduced vowels. The faithfulness constraint that blocks native vowel reduction is ranked above * [+Stress, +ATR] because it is specifically indexed for loans (L–Faith (e o) >> * [+Stress, +ATR]).

Nevertheless, when stressed mid vowels occur in final position – mostly in gallicisms – the tendency is to surface as [+ATR], as is shown in (15).

(15)  beb[ɛ]  consom[ɛ]  neglig[ɛ]  
      jaqu[ɛ]  pur[ɛ]  cup[ɛ]  
      frivolit[ɛ]  clix[ɛ]  pat[ɛ]  
      ximpanz[ɛ]  quinqu[ɛ]  moar[ɛ]  

These examples contrast with the following adapted loans: t[ɛ] ‘tea’, caf[ɛ] ‘coffee’, obo[ɛ] ‘oboé’, comit[ɛ] ‘committee’. Relative to the native vocabulary, such words must be regarded as morphological exceptions because the lack of final /n/ in the

5. It is important to point out that words ending in unstressed and orthographic ‘-u’ are practically non-existent in the native Catalan vocabulary. On the other hand, they are abundant in borrowings from Latin.
underlying form prevents them from undergoing any suffixation process. If we compare their morphological productivity with that of final-stressed native words, the contrast is even more evident: morè [muˈɾe], morens [muˈrenʃ] ‘brown, sg. and pl’, morenet [muɾəˈnet] ‘little brown’, morenor [muɾəˈno] ‘brownness’, etc. versus clixé [kliˈɾε] and clixé<s>és [kliˈɾes] ‘cliqué sg. and pl’, and nothing else). We must also bear in mind that the native process of final /ɾ/ deletion in stressed syllable is not active (with a few exceptions) in loanword phonology: plumi<er> ‘pencil case’, necess<er> ‘toiletry bag’ and premi<er> ‘prime minister’, while Molière and Baude-laire are pronounced with final [er]. Even though the context calls for a [–ATR] mid vowel, the stressed vowel in native words is [+ATR]. One would expect the pronunciation to follow the pattern we find in the stressed suffix /–er/ with final /ɾ/ deletion (one of the most productive suffixes in the language: fuster [fusˈte] ‘wood-worker’, limoner [liˈmoɾe] ‘lemon tree’, sabater [səbəˈte] ‘shoemaker’). Nevertheless, loans never delete final /ɾ/.

We therefore propose the markedness constraint \( L^*[–\text{ATR}]# \) for loans ending in a mid vowel:

\[
(16) \quad L^*[–\text{ATR}]#: \text{Open mid vowels in stressed final position are not allowed in loans.}
\]

This markedness constraint indexed to loans accounts for what Pater (2004) denotes an exceptional triggering process, that is, a process that only applies to exceptional forms. Because of its exceptionality this constraint is ranked above and dominates all other conditions, as is shown in (17).

\[
(17) \quad L^*[–\text{ATR}]# \gg *[–\text{Stress, –ATR}] \gg L–\text{Faith (e o)} \gg *[+\text{Stress, +ATR}] \gg *[–\text{Stress, +ATR}]
\]

We illustrate in (18) how this ranking accounts for the surface form of loans such as consomé or Dostoievsky:

<table>
<thead>
<tr>
<th>Dostoievsky</th>
<th>L*[–ATR]#</th>
<th>L–Faith (e o)</th>
<th>*[+Stress,+ATR]</th>
<th>*[–Stress,+ATR]</th>
</tr>
</thead>
<tbody>
<tr>
<td>d[ɔ]st[o]i[ˈɛ]vsk[i]</td>
<td>*!</td>
<td>*</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>d[u]st[u]i[ˈɛ]vsk[i]</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>c[o]ns[o]m[ˈɛ]</td>
<td>*!</td>
<td>*</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>c[o]ns[o]m[ˈɛ]</td>
<td>*!</td>
<td>*</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>c[u]ns[u]m[ˈɛ]</td>
<td>*!</td>
<td>*</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>c[u]ns[u]m[ˈɛ]</td>
<td>*!</td>
<td>*</td>
<td>*!</td>
<td>*</td>
</tr>
</tbody>
</table>

We illustrate in (18) how this ranking accounts for the surface form of loans such as consomé or Dostoievsky:
5. Vowel harmony

When post-tonic mid vowels surface as [+ATR], stressed mid vowels undergo a process of vowel harmony: the otherwise expected outcome with [–ATR] is blocked and open mid vowels are raised to close when followed by a close mid vowel (Cabré 2006, Bonet et al. 2007). Some examples provided in (19) show the harmonic process.

(19) pesto [ˈpesto] ‘pesto sauce’ Boston [‘boston]
gueto [ˈgeto] ‘ghetto’ Oslo [‘oslo]
Volvo [ˈbolβo] Tere [ˈtere]
secre [ˈsekre] ‘secretary’ grosso modo [ˈgrɔsomɔdo]
cole [ˈkole] ‘school’ jefe [ˈxeфе] ‘boss’
Irene [iˈrene] Penèlope [peˈnelpe]
UNESCO [uˈnesko], Beethoven [ˈbeʃoßen]
Hölderlin [ˈxolderlin] Essex [ˈeseks]
polièster [pɔlˈjestɛr] ‘polyester’ Euterpe [ˈewterpe]
Montse [ˈmonse] forense [ˈfoɾense] ‘forensic’
pòquer [ˈpɔker] ‘poker’ cosmos [ˈkosmos]
Pinotxo [piˈnɔtʃo] ‘Pinocchio’ Nèstor [ˈnestor]
Tòquio [ˈtɔkjɔ] ‘Tokyo’ Mendelsohn [ˈmendelson]
Opel [ˈopel] folklore [ˈfolˈklɔre]

In the examples above, we note that the [+ATR] value has spread from the post-tonic vowel to the stressed mid vowel. The scope of this harmonic spread is one trochee, the basic foot of Catalan (Cabré 1993). If we focus on proparoxytones, the harmonic process arises when a stressed vowel is followed by a close mid vowel. When the medial post-tonic vowel is a schwa or high, the general outcome is an absence of harmony (e.g., Sòcrates [ˈsɔkɾates], Hèrcules [ˈerkules]). Nevertheless, it is important to bear in mind that this process presents a great amount of variation across speakers and dialects. For example, data gathered from Eastern Catalan speakers showed that some of them did not apply the harmonic process to words such as r[ɛ]qui[ɛ]m ‘requiem’, r[ɛ]c[o]rd ‘record’, N[o]bel, S[e]n[e]ca, C[o]rc[e]ga, or G[ɔ]lg[o]ta. Others extended the process outside the trochee and pronounced the words Sòcrates [ˈsɔkɾates] and Hèrcules [ˈerkules]. At any rate, the more distant (or foreign) a word is felt to be, the more likely it is that the harmony process will apply; by the same token, the closer (or nativized) a word is felt to be, the more likely it is that vowel reduction will occur.

The vowel harmony process presented here shows some parallels with the vowel harmony in Central Italian described by Canalis (2009, this volume). In both cases the trigger for the harmony is an unstressed vowel and the target is also to the left of the trigger. This indicates that the direction of the harmonic process goes from right to left, exactly the opposite of the best-known harmony systems.
such as in Turkish or Finnish. Conversely, there is a crucial difference between the two processes, because whereas in the Italian dialects the target is likewise a weak vowel, in Eastern Catalan the target is the stressed vowel of the word; the processes thus entail distinct analyses. Moreover, the prosodic context of the Catalan process contrasts with the general harmony cases in which the trigger is a prominent position and the target a weak position.

Indeed, the fact that the trigger for vowel harmony is the post-tonic [+ATR] vowel, an element in a weak position, seems to disagree with analyses based on the perceptual prominence of strong positions, which is certainly what the stressed syllable is. According to Walker (2005), the ATR feature of the post-tonic vowel is associated with the preceding stressed vowel in order to become more perceptible. Weak position (the post-tonic vowel) “is licensed by association to strong position” (the stressed vowel) in the scope of a trochee. In other words, the ATR feature is attracted to the stress position, yielding ATR harmony. This phenomenon has a particularly pronounced effect on loans and generates a prominent difference between them and the native vocabulary. In order to account for this fact we propose the following constraint of positional markedness:

(20) License [ATR post-tonic, σ]’: ATR value spreads to the preceding stressed vowel.

License [ATR post-tonic, σ]’ dominates *[+Stress, +ATR]. This markedness constraint is indexed to a subset of words we have labeled as loans that is in turn a subset of learned words. As L–Faith (e o) is indexed to learned words, License [ATR post-tonic, σ]’ must also dominate L–Faith (e o) and *[+Stress, +ATR]. The ranking in (21) accounts for the vowel harmony process, the most salient feature of Catalan loanword phonology.

(21) [ATR post-tonic, σ], L*[–ATR]# >> *[–Stress, –ATR] >> L–Faith (e o) >> *[+Stress, +ATR] >> *[–Stress, +ATR]

Tableau (22) illustrates an example of vowel harmony in Eastern Catalan. The constraints L*[–ATR]# and *[–Stress, –ATR] are not relevant for the process. Pinotxo ‘Pinocchio’ may be interpreted by some Catalan speakers as a native name or it has simply become completely nativized, thus surfacing as pin[ɔ]tx[u]. Nevertheless, this pronunciation never arises when the word is indexed as a loan.

(22)
Even though adjacency inside the trochee is needed for the general process, we find some loans – in greater or lesser quantity depending on the speaker – with the same effects of vowel harmony yet without the requirements of adjacency in the foot. The mere presence of an unstressed [+ATR] mid vowel inside the word is enough to trigger unbounded vowel harmony. In this case, the scope of the harmony process is the whole word. Some examples are Otegi [oˈteyi], Sòcrates [ˈsokrates], Khomeini [ xoˈmejni], and vedet [beˈðeıt]. In sum, much work is needed to clarify the differences among dialects and speakers.

6. Variation and stages of nativization

As we have noted, there is a certain amount of variation in the application of all these processes. For example, unstressed closed mid vowels in older borrowings are quite common in the northeast area (from Barcelona to Girona), but less so in the center and the south of the dialect region where Eastern Catalan is spoken. Moreover, we also have found disagreement within the same dialect, often across different generations. This variation may or may not involve the presence of vowel harmony when the segmental context allows it: euro [ˈewro]/[ˈewru], Nobel [ˈnoʃel]/[ˈnoʃal], Irene [iˈrene]/[iˈɾεnə]. Within the Barcelona area, it is also possible to hear some words with unstressed [+ATR] vowels pronounced without evidence of the vowel harmony process: Sèneca [ˈsεnekə], Còrcega [ˈkorseyə], Nobel [ˈnoʃel], réquiem [ˈrekjem], récord [ˈrekor]. As far as we know, this possibility is blocked in the rest of the linguistic domain when mid vowels occur in stressed positions.

The age of a speaker also seems to play an important role: in general, younger people are more innovative than older people and therefore apply the different processes across a broader range of words. In addition, vowel harmony applied to the whole word is quite widespread in the younger generation, where one hears vedet [beˈðeıt], Khomeini [ xoˈmejni], Toyota [toˈjota], Minnesota [mineˈsota], etc.

If we accept the notion of different strata in the lexicon, it is possible to explain these different phonologies and the variation found in their application. As Ito & Mester (1999:64) say “Lexical items do not come neatly packaged into groups labeled [+/- foreign]; rather, different degrees of nativization among foreign words are commonplace. Instead of the partitioning into parallel and disjoint [+foreign] and [−foreign] sublexica, we have a hierarchy of foreignness, with exceptions to one rule always being exceptions to another rule, but not vice versa.”

In summarizing the differences found in the vocabulary at hand that we have labeled new words, it is possible to define three groups according to the three main
processes discussed. The first group includes those new words (learned words, classical language compounds, etc.) whose main characteristic is to fail mid vowel reduction; the second group corresponds to loanwords or any new word that undergoes a harmony process inside the trochee; the third one includes the newest vocabulary (unassimilated words), and shows unbounded harmony when no reduced mid vowels are present in the word. Native vocabulary consistently exhibits none of these processes.

By taking up the proposal of Itô & Mester, we regard lexical items as “organized in terms of an overall core-periphery structure” and adapt their model to the Catalan sublexica: Lex$^0$ – native vocabulary; Lex$^1$ – learned words; Lex$^2$ – loanwords; Lex$^3$ – unassimilated words. Considering the variation found in the data, a word belongs to one or another level according to how many processes it has undergone, which will depend on the speaker and dialect. Moreover, we must bear in mind that some items historically labeled as borrowings (such as enterro [ənˈtɛɾu] ‘burial’) have a core behavior, that is, they are completely nativized, whereas other items that are historically native (such as literature [liteɾəˈtuɾə] ‘literature’) are still peripheral for some speakers because of the pronunciation of the front mid vowel. The structure of the Catalan sublexica is represented in (23):

(23)

Native words constitute the core of the lexicon because they fulfill all constraints. As Ito & Mester (1999:65) emphasize, “moving outwards from the core, we encounter items that violate more and more constraints until we encounter, at the periphery, items fulfilling only a small subset of the constraints. These constraints are truly fundamental in the sense that they define the basic syllable canons and other central aspects of the language”. Thus, our data clearly show that initial epenthesis is one of the basic, unavoidable phonological processes of Catalan; “native” vowel reduction, on the other hand, is not such a central process.

6. It is not completely clear in what set words with stressed final [+ATR] mid vowel, such as consomé, belong.
We show in (24) the relation between the different classes of lexical items discussed above and the processes active in the Catalan lexicon. The basic syllable structure is represented here by the initial vowel epenthesis but includes other phonological processes (such as sonority or articulation place assimilation) that we have not discussed.

Except for initial vowel epenthesis, processes applying to a specific sublexicon must be labeled in negative terms. Native vocabulary items must fulfill all core requirements, so it would violate them in positive terms because of their peripheral status. The table in (25) clearly shows the hierarchical relations between the different strata in the Eastern Catalan phonological lexicon.

<table>
<thead>
<tr>
<th></th>
<th>InEp</th>
<th>No-WdHarm</th>
<th>No-FtHarm</th>
<th>No-ExRed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native words</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Learned words</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>violated</td>
</tr>
<tr>
<td>Loanwords</td>
<td>√</td>
<td>violated</td>
<td>violated</td>
<td>violated</td>
</tr>
<tr>
<td>Unassimilated</td>
<td>√</td>
<td>violated</td>
<td>violated</td>
<td>violated</td>
</tr>
</tbody>
</table>

The dialectal and idiolectal variation is interpreted in terms of item place, that is, each dialect – even each speaker – can vary in the consideration of a particular word in its lexicon. As Itô & Mester (1999:70) say, “the less nativized an item is, the more it disobeys lexical constraints, i.e. the more it falls outside of various constraint domains and is located towards the periphery of the lexical space.”

The stratification of the Catalan lexicon presented above accounts for the gradual application of the constraints to different subsets of words. A grammar with indexed faithfulness constraints interspersed between markedness constraints solves not only the problem of apparent different grammars but also its learnability in an Optimality Theory account (Pater 2005).

7. Conclusion

Summarizing our study, we can say that it offers a feasible interpretation of the amalgam of data at hand, although considerable work is still needed to clarify dialectal differences. On the one hand, the exceptions to vowel reduction in learned
words are interpreted in terms of faithfulness to written form in order to make them stand out by improving the equidistance of positions in the vowel trapezoid. On the other hand, Catalan loanwords show two opposite effects with respect to stressed vowels: the emergence of the unmarked generalizes the [–ATR] value on stressed mid vowels in some specific positions, whereas the [+ATR] value harmonizes the stressed mid vowel in the presence of a post-tonic [+ATR] mid vowel. A word that undergoes this vowel harmony becomes more perceptible, so that the scope of the process is the stress foot. The lexical strata framework accounts for the apparent random facts and idiosyncrasies of non-native words in a very intuitive way. The stratification of the lexicon into domains is clearly evident in Eastern Catalan, but the variation found in the data must be interpreted as reflecting the idiolectal location in one of these domains.

References


Kenstowicz, Michael. 2001. The Role of Perception in Loanword Phonology, Ms.

——— 2005. Loanword Phonology. Lectures at LSA Institute, July. Cambridge, MIT.


Index of Subjects and Languages

A
accent placement 91, 96
accent type 62, 91–94, 96, 98
acoustic prominence 95
acoustic-auditory factors 207, 212, 228
acoustic-auditory 210, 217, 227
acquisition 73, 109, 120, 129, 156, 213, 218–220
acquisition of topic and focus marking 91–104
adult-like focus realization 91, 94
adult-like topic realization 91, 94
aerodynamic 210–212, 214, 223, 225–229
aerodynamic factors 207, 210, 212, 223, 227, 228, 229
aerodynamic constraint 206, 225–227, 229
airflow 216, 223, 227–228
alignment of F0 maximum 96, 100, 102–104
alignment of F0 minimum 96, 100, 102, 104
articulation dynamics 31
articulatory phonology 16, 193, 206
articulatory transition 8
ATR 252, 267, 268–282, 284
B
Basaa 221–222, 224
Bola 218
Burmese 209–210
breaking 17
directionality of vowel harmony 248, 251, 254, 260, 265
disfluency 109–110, 114–115, 120, 125–126, 128, 130–131
domain-edge effects 155
domain-head effects 155
Dutch 2, 37, 48, 91, 93–96, 100, 102, 104, 118, 147–148, 156

e
emergence of the unmarked 176, 178, 276–277
American English 15–16, 18, 20–22, 27–28, 32, 37, 93–95, 214–216, 219
Lancashire English 220, 222
epenthesis 15–21, 27, 29–31, 274, 282–283
epenthetic vowels 19, 182–183; see also vowel insertion, vowel intrusion
experimental phonetics 2

glide formation 159–160, 173, 175
glottal stop 215, 263

F0 peak alignment

glottalization 216
gradient 4–5, 7, 10, 64, 91, 94, 96, 104, 138, 140–141, 147
grammar 1–2, 4, 111, 117, 205, 207, 229, 240, 242, 244, 268–269

F0

Greek 2, 109–110, 121–124, 130–131, 212, 219

H

Hindi 214, 216, 219

I

Indonesian 215

information structure categories 3, 5, 6

information value 92, 94–96

informational completeness 102, 104

insertion 15–21, 27, 29–31, 129, 169, 182, 184; see also vowel insertion

intensity 35–40, 41–46, 47–48, 208, 211, 212, 216, 227

intonation grammar 4

intonation systems 5, 115

intonational phrase 111, 113, 153

intonational realization 91, 93–95

intrusion 19, 181–182, 184

intrusive vowel 19, 182–184, 196


Dialects of Central Italy 247–248, 256

Garfagnana 248, 249, 261

Northern Italian 141, 154

Neapolitan Italian 73, 140

Sant’Oreste 248, 250, 254

Servigliano 248, 249, 253–254

Sicilian Italian 137, 146–148, 150, 152, 154–156

Southern Italian 137, 140–142, 147, 154, 213

Umbertide 248, 250, 254, 256

Venetian Italian 137, 146–151, 154–156

J


K

Karitiana 218–219

L

labial postures 123

laboratory phonology 1–3

laryngeal postures 126


lengthening effect 38, 47–48, 52–53, 62, 64, 86, 149, 154–155

lenition 124, 130–131, 156

lexical strata 267–268, 284

lexical stress 85–86, 88, 139–140, 143

lexicon strata 270

lingual posture 112, 114, 126, 117

linguistic rhythm 6

liquids 183, 247–249, 253–254, 261–265

loanwords 267–271, 277, 282–284

low plateau 100, 102–104
M

Mandar 214
manner of articulation 184, 189–190, 193
metaphony 249, 252, 254
metrical position 112–113, 114, 256–257
metrics 137–142, 146, 149, 156
Mixtec 218, 225–226
morphological and prosodic factors 162
motor control 115, 119, 131
multiple grammars theory 240, 242

N

nasal airflow 216
nasal leakage 207, 212, 216–219, 224, 226–229
nasal release 220
nasality 207, 209–213
nasalized fricatives 210–212, 229
native vocabulary 268–275, 277, 280, 282–283
New Julfa 213
null prominence 165, 170–171, 173, 176–177

O

obstruent 120, 125, 206–207, 212, 214–218, 221, 222, 223, 229
Optimality Theory 160, 173, 207, 229, 283
oral cavity expansion 212, 224, 229
oral pressure 210–212, 217–218, 221, 223–225, 227–228
OT, see Optimality Theory
overall intensity 40–49

P

palatalization 235–242, 244
palatalization of dental stops 235–238, 240
palatalized consonants 141–242
paradigm uniformity 173
Parallel Structures Model 240
peak alignment 74, 77, 81–82, 96
perception 1–3, 23, 35–39, 42–48, 94, 118–119, 137, 139, 147, 156, 182, 184, 212, 228
perceptual prominence 280
phonetic description 72
phonetic factors 85, 87–88, 206, 216, 230
phonetic implementation 4
phonetic marking 91, 95, 96, 104
phonetic realization strategy 67
phonetics 1–4, 6, 16, 36, 91, 260
phonetics-phonology connections 1
phonological acquisition 120, 213
phonological features 240
phonological marking 91–92, 94, 104
phonological representation 16, 18, 194
phonology 1–4, 6, 10, 16, 48–49, 181, 197, 228, 267–272, 274–275, 278, 280
phrase accent 111, 112
physical factors 205, 207, 223
pitch accent 35, 52–53, 58, 59, 63, 73, 78–85, 86–88, 92, 95, 111, 115, 123–124
pitch accent distribution 5
see also alignment, focus accent, phrase accent, topic accent, complex pitch pattern, complex pitch gesture, simple pitch movements, tonal compression, truncation
pitch alignment 71, 73, 76–77, 82, 86–88
pitch range 58–59, 71, 73, 78–82, 83, 85–88
expanded pitch range 71, 79, 86
pitch range increase 73–74, 85
pitch range reduction 85
see also post-focal pitch range reduction, pre-focal pitch range reduction
place 185, 188–191, 193, 196, 208–210, 224, 226, 228, 229, 240–242, 247–248, 253, 254, 260–265
place features 206, 240, 253, 262, 264–265
Portuguese 2, 54, 73, 86–88, 138, 141, 155–156, 235–236, 238, 244
Antônio Prado 235, 237, 238, 244
Brazilian Portuguese 238, 141, 153–156, 235–236, 244
European Portuguese 54, 73, 86, 87, 88, 138, 141, 155
post-focal contexts 36, 37, 71, 79, 80, 83–86, 88
post-focal pitch range reduction 85
postnasal voicing 213, 222, 229
postnasalization 220, 225, 227–228
pre-focal contexts 71, 79, 83
pre-focal pitch range reduction 85
prenasalization 207, 217–221, 224–226, 227, 228–229
prominence levels 175
prosodic (π-) gestural model 181–182, 195–196, 197
prosodic and segmental factors 181–182, 197
prosodic lengthening 149–150, 154–155
prosodic prominence 124, 175–176
prosodic structure 2, 87, 109, 119–120, 122, 125, 140, 156, 247
prosodic timing 142–143, 148–149, 155–156
PSM, see Parallel Structures Model

Index of Subjects and Languages 289
Index of Subjects and Languages

R
raddoppiamento 141
reduction 16, 79–81, 84–86, 124, 141–142, 150, 156, 255–256, 259–261; see also vowel reduction
redundancy rules 207, 262
resonances 208–209
rhythm 116, 137–143, 146–147, 149–150, 154–156
rhythm classes 138
rhythmic structure 109, 117
rhythm metrics 136, 138, 140, 142, 146, 149, 156
see also syllable-timed, stress-timed
rising sonority sequences 160–161
Romance languages 73–74, 86–88, 137, 140, 156, 160, 166, 247, 251
Romanian 2, 37, 71–74, 79, 82, 84–88
S
schwa 15, 17–32, 38, 182, 259, 268–270, 274
schwa-like element 15, 19–21, 23, 25–32
second language acquisition 218, 220
segmental durations 72, 86–88
semi-spontaneous production 93, 96
simple pitch movements 4
sonority sequences 160–161, 173–174
sonority-driven unstressed system 272
sound pattern 205, 211, 214, 216
speaking rate 15, 20–22, 28, 30–31; see also rate
spectral tilt 35, 37, 40, 45–48
speech errors 110, 113–114
spontaneous nasalization 212, 218, 220, 222
stages of nativization 281
stress contrast 35, 38–40, 47–48, 147, 149, 154–155, 156
stress-timed 138, 140, 146, 154–155
stutterers 114
syllabification 159–161, 168–169, 174
syllable duration 57, 100–101
syllable fusion 124
syllable-timed 137–138, 140–141, 155
T
temporal stress contrast 139, 147, 154–155
Terena 212–213
three-syllable window 165, 183, 251
timing 67, 137, 145–149, 154–156, 194–195
tonal categories 4
tonal compression 4
tongue posture 117, 126; see also labial postures, laryngeal postures, lingual posture
topic 91–104
sentence-initial topic 91, 93–94, 96, 98, 104
topic accent 93
types of topic 95
see also adult-like topic realization
truncation 51–53, 64–67, 271
underspecification 247–248, 262–263, 265
uniform exponence 174
unstressed mid vowels 238, 243, 268–269, 274–275, 277
V
velic 212, 215–217, 220, 227
velic lowering 215–217, 220, 227
velic raising 212
verbal stress 166–167, 172
vocalic prominence scale 272
voiced stop 183, 215, 217, 219, 220
vertical stress 208–209
voiceless nasals 208–210
voiceless stop 183, 212, 213, 214–215, 217, 218, 221, 222–225, 227, 229, 230
voicing 184, 189, 192, 205–207, 210–229
voicing contrast 207, 217, 220, 221, 223, 225
vowel duration 73, 74, 83–85, 101, 141, 148–149, 154–155, 196, 258–259
vowel insertion 129, 182
vowel intrusion 19, 181–182
vowel lengthening 149, 214
vowel postures 112
vowel quality 38–39, 139
vowel type 35, 39, 44, 46, 48–49
W
Waris 217–218
word boundary 81, 181–182, 188–190, 193, 195–196, 200
π-gestural model, see prosodic (π-) gestural model
Current Issues in Linguistic Theory (CILT) is a theory-oriented series which welcomes contributions from scholars who have significant proposals to make towards the advancement of our understanding of language, its structure, functioning and development. CILT has been established in order to provide a forum for the presentation and discussion of linguistic opinions of scholars who do not necessarily accept the prevailing mode of thought in linguistic science. It offers an outlet for meaningful contributions to the current linguistic debate, and furnishes the diversity of opinion which a healthy discipline must have. A complete list of titles in this series can be found on the publishers’ website, www.benjamins.com

311 SCALISE, Sergio and Irene VOGEL (eds.): Cross-Disciplinary Issues in Compounding. Expected February 2010


307 CALABRESE, Andrea and W. Leo WETZELS (eds.): Loan Phonology. vii, 273 pp. Expected November 2009


297 CAMACHO, José, Nydia FLORES-FERRÁN, Liliana SÁNCHEZ, Viviane DÉPREZ and Maria José CABRERA (eds.): Romance Linguistics 2006. Selected papers from the 36th Symposium on Romance Languages (LSRL), New Brunswick, March-April 2006. 2007. viii, 340 pp.

298 WEIJER, Jeroen van de and Erik Jan van der TORRE (eds.): Voicing in Dutch. (De)voicing – phonology, phonetics, and psycholinguistics. 2007. x, 186 pp.


302 PRIETO, Pilar, Joan MASCARÓ and Maria-Josep SOLÉ (eds.): Segmental and prosodic issues in Romance phonology. 2007. xvi, 262 pp.

303 VERMEERBERGEN, Myriam, Lorraine LEESON and Onno CRASBORN (eds.): Simultaneity in Signed Languages. Form and function. 2007. viii, 360 pp. (incl. CD-Rom).


312 GESS, Randall S. and Edward J. RUBIN (eds.): Theoretical and Experimental Approaches to Romance Linguistics. Selected papers from the 34th Symposium on Romance Languages (LSRL), Salt Lake City, March 2004. 2005. viii, 367 pp.


238 NÚÑEZ-CEDEÑO, Rafael, Luis LÓPEZ and Richard CAMERON (eds.): A Romance Perspective on Language Knowledge and Use. Selected papers from the 31st Linguistic Symposium on Romance Languages (LSRL), Chicago, 19–22 April 2001. 2003. xvi, 386 pp.


220 SATTERFIELD, Teresa, Christina TORTORA and Diana CRESTI (eds.): Current Issues in Romance Languages. Selected papers from the 29th Linguistic Symposium on Romance Languages (LSRL), Ann Arbor, 8–11 April 1999. 2002. viii, 412 pp.


LEMA, José and Esthela TREVIÑO (eds.): Theoretical Analyses on Romance Languages. Selected papers from the 26th Linguistic Symposium on Romance Languages (LSRL XXVI), Mexico City; 28–30 March, 1996. 1998. viii, 380 pp.


Herschensohn, Julia: Case Suspension and Binary Complement Structure in French. 1996. xii, 200 pp.


