International Speech Communication Association

ExLing 2016

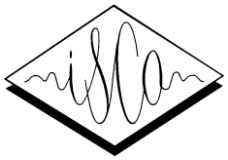
Proceedings of 7th Tutorial and Research Workshop on Experimental Linguistics

27 June - 2 July 2016, Saint Petersburg, Russia

Edited by Antonis Botinis

Saint Petersburg
State University

National and Kapodistrian
University of Athens



International Speech Communication Association

ExLing 2016

Proceedings of 7th Tutorial and Research Workshop on Experimental Linguistics

27 June – 2 July 2016, Saint Petersburg, Russia

Edited by Antonis Botinis

Saint Petersburg
State University

National and Kapodistrian
University of Athens

ExLing 2016
Proceedings of 7th Tutorial and Research Workshop on Experimental Linguistics

Ebook ISSN: 2529-1092
Ebook ISBN: 978-960-466-161-9
Copyright © 2016 Antonis Botinis

Foreword

This volume includes the proceedings of ExLing 2012, the 5th Tutorial and Research Workshop on Experimental Linguistics, in Athens, Greece, 27-29 August 2012. The first conference was organised in Athens, in 2006, under the auspices of ISCA and the University of Athens and is regularly repeated thereafter, including the last one in Paris, in 2011.

In accordance with the spirit of this ExLing 2012 conference, we were once again gathered in Athens to continue our discussion on the directions of linguistic research and the use of experimental methodologies in order to gain theoretical and interdisciplinary knowledge. We are happy to see that our initial attempt has gained ground and is becoming an established forum of a new generation of linguists.

As in our previous conferences, our colleagues are coming from a variety of different parts of the world and we wish them a rewarding exchange of scientific achievements and expertise. This is indeed the core of the ExLing events, which promote new ideas and methodologies in an international context.

We would like to thank all participants for their contributions as well as ISCA and the University of Athens. We also thank our colleagues from the International Advisory Committee and our students from the University of Athens for their assistance.

Antonis Botinis

Contents

Tutorial papers

Remanence of sentence prosody in Romance languages *1*
Philippe Martin

Rich Reduction: Sound-segment residuals and the encoding of communicative functions along the hypo-hyper scale *11*
Oliver Niebuhr

Research papers

Visual search strategies and letter position encoding in Russian *25*
Svetlana Alexeeva

Emergence of word prosody in (Seoul) Korean *29*
Angeliki Athanasopoulou, Irene Vogel 29

Voice Activity Detector (VAD) based on long-term phonetic features *33*
Andrey Barabanov, Daniil Kocharov, Sergey Salishev, Pavel Skrelin, Mikhail Moiseev

The identification of two Algerian Arabic dialects by prosodic focus *37*
Ismaël Benali

Intonation and polar questions in Greek revisited *41*
Antonis Botinis, Anthi Chaida, Olga Nikolaenkova, Elina Nirgianaki

The imprint of disposition in social interaction *45*
Mark Campana

Intonation and polar questions in Greek *51*
Anthi Chaida, Angeliki Sotiriou, Athina Kontostavlaki

Contextual predictions and syntactic analysis: the case of ambiguity resolution *55*
Daria Chernova, Veronika Prokopenya

Vocal fatigue in voice professionals: collecting data and acoustic analysis *59*
Karina Evgrafova, Vera Evdokimova, Pavel Skrelin, Tatiana Chukaeva

Creating a subcorpus of a heritage language on the example of Yiddish *63*
Valentina Fedchenko, Ilia Uchitel

Affricates in the spontaneous speech of Aromanians in Turia *67*
Anastasia V. Kharlamova

L1 transfer, definiteness and specificity of determiners in L2 English *71*
Sviatlana Karpava

Writing-based wordforms vs. spoken wordforms *75*
Vadim Kasevich, Iuliia Menshikova

On the buildup of an integrated database for the formal description of grammars for the hearers *79*
Vadim Kasevich, Iuliia Menshikova, Maria Khokhlova, Elena Shuvalova, Anna Lastochkina

How to write an oral dialect or about some problems of the Tsakonian Corpus *83*
Maxim Kisilier

Some aspects of /r/ articulation in French Vocal Speech *87*
Ulyana Kochetkova

Different acoustic cues for emphasis in teaching English word stress to Hong Kong Cantonese ESL learners of different proficiencies *91*
Wience Wing Sze Lai, Manwa Lawrence Ng

Cognitive approach to translation and interpreting teaching methods *95*
Julia Levi

Perception of reduced words: Chunking and predictability *99*
David Lorenz, David Tizón-Couto

Neurological state manifestation in infants' and children's voice features *103*
Elena Lyakso, Olga Frolova

Features of written texts of people with different profiles of Lateral Brain Organization of Functions (on the Basis of RusNeuroPsych Corpus) *107*
Tatiana Litvinova, Ekaterina Ryzhkova, Olga Litvinova

Semantic differential as a method in empirical investigation of Self-Image as father *111*
Robert Manerov, Kristina Manerova

Automatic assignment of labels in Topic Modelling for Russian Corpora *115*
Aliya Mirzagitova, Olga Mitrofanova

The time course of sociolinguistic influences on wordlikeness judgments *119*
James Myers, Tsung-Ying Chen

The function of olfactory experience in reasoning: An empirical study *123*
Katalin Nagy

Gender Features in German: Evidence for Underspecification *127*
Andreas Opitz, Thomas Pechmann

Distributional analysis of Russian lexical errors *131*
Polina Panicheva

Serbian pitch accents in tri-syllables produced by Serbian and Russian speakers *135*
Ekaterina Panova

Effect of saliency and L1-L2 similarity on the processing of English past tense by French learners: an ERP study *139*
Maud Pélissier, Jennifer Krzonowski, Emmanuel Ferragne

Phonostylistic study of Spanish-speaking politicians: Populist vs. Conservative ... *143*
Carmen Patricia Pérez

Experimental L2 text production with WinPitch LTL *147*
Darya Sandryhaila-Groth

Exploring prosodic convergence in Italian game dialogue *151*
Michelina Savino, Loredana Lapertosa, Alessandro Caffò, Mario Refice

Syllable cueing and segmental overlap effects in tip-of-the-tongue resolution *155*
Nina Jeanette Sauer

An experimental study of English accent perception *159*
Elena Shamina

Phonetic words duration simulation using Deep Neural Networks *163*
Alexander Shipilo

Transcription: what is meant by accuracy and objectivity? *167*
Pavel Skrelin, Nina Volskaya

Grammatical change and hindcast model statistics – A comparison between Medieval French and Brazilian Portuguese *171*
Eduardo Correa Soares

The Phonetics of Russian North Bylinas *175*
Svetlana Tananaiko, Marina Agafonova

Association experiment in practice of linguistic and cultural dominants research *179*
Svetlana Takhtarova, Diana Sabirova

Filled pauses and lengthenings detection using machine learning techniques *183*
Vasilisa Verkhodanova, Vladimir Shapranov, Alexey Karpov

Psycholinguistic evidence for the composite group *187*
Irene Vogel, Angeliki Athanasopoulou

Remanence of sentence prosody in Romance languages

Philippe Martin
LLF, UFR Linguistique, Université Paris Diderot Sorbonne Paris Cité

Abstract

Romance languages uses surprisingly similar melodic contours to encode the sentence prosodic structure. The fact that these contours are governed by similar prosodic grammars and that similar stress rules are also applicable to these languages (except on French deprived of lexical stress) suggests that these phonological facts are inherited from Latin without much change, despite the constant evolution occurred during twenty centuries.

Key words: intonation, prosody, Romance languages, stress, prosodic grammar

Introduction

Sentence intonation is always present in the linguistic communication, even in silent reading. We cannot process language, whether in oral or written form, without decoding the prosodic structure intended by the speaker or recover (or approximate) the intonation intended by the writer.

Indeed, due to memory limitations, it is not possible to retain long lists of objects such as words or syntagms without structuring these lists by some hierarchical grouping. Remembering large numbers or long lists of digits as found in telephone or credit card numbers requires to structure this information into small chunks, eventually organized into two or more levels, in order to form a structure. In these specific cases, where digits lack of any morphological information, only the prosody, organized into a prosodic structure, that will give to the listener enough indications to restore the intended structure of the data. In reading, this role is devoted to graphic indicators such as blanks separating groups of digits or of words.

In speech communication, although many morphological or grammatical tools are available to recover a structure from the sequence of syllables pronounced by a speaker, it is again the prosodic structure which provides the first and essential hints to decode the sentence structure.

The Romance language family

The concept of language family is well known, and can be traced back to the XIXth century or earlier. The membership of a given language to a specific family is established by comparisons of lexical, syntactic, phonological and phonetic similarities between languages candidates, leading to the virtual

reconstitution of non-attested languages that would be the mother of the family. Examples of such well documented families are Nordic (including English, German, Dutch, Norks, Swedish, Danish…), Romance (French, Italian, Spanish, Portuguese, Catalan, Romanian…) or Turkic (Turkish, Mongolian, Sakha…).

Whereas comparative linguistics deals with phonetic, phonological, lexical and syntactic data, few scholars did compare prosodic features such as stress location, not to mention properties and grammar of melodic contours indicating the sentence prosodic structure. One notable exception is Paul Garde (1968, 2013) who gives simple and convincing word stress placement rules in various language families, including Romance. Comparison of prosodic features in Romance has been also the subject of two recent books. One, edited by Frota and Prieto (2015) operates in the Autosegmental-Metrical framework to compare phonetic and phonological features of many Romance languages, actually mixing both types of phonetic and phonological properties. The other (Martin, 2015) conducts its comparisons in the incremental prosodic structure framework, aiming to better establish the similarities between a phonological description of prosodic properties, carefully separated from phonetic differences. The adopted framework allows to establish clear similarities between prosodic systems in the languages considered (Italian, Spanish, Catalan, Italian, Romanian) as well as the important differences present in the system of French.

Intonation in Romance

Investigate similarities of prosodic structures in Romance languages involves three main topics:

1. Stressed syllable location
2. Melodic contours on stressed syllables
3. Grammar of melodic contours in the prosodic structure

In the Romance languages considered, Italian, Spanish, Catalan, European Portuguese and Romanian, prosodic markers, instantiated by melodic contours, appear surprisingly similar in these three categories: similar stressed syllable placement rules are applicable, similar phonological melodic contours are revealed through acoustic analysis, and the distribution of melodic contours in the sentence are described by the same grammar (Martin, 2015). Being so comparable, would these features be also valid for classical Latin from which the Romance languages are derived, leading to the reconstitution of Latin sentence prosody?

Stress syllables in Latin

It is remarkable that the position of lexical stress in most Romance languages can be traced back to Latin despite twenty centuries of phonetic, phonological and syntactic evolution. Classical Latin had five phonological vowels, |i| |e| |a| |o| and |u| (the vowels included in today Latin computer Latin fonts). Each vowel can be short or long so that the vocalic system includes five short and five long vowels. Latin has also three diphthongs, |aj|, |oj| and |aw|, written ae, oe, au.

The stress rule is as follows (Alkire and Rosen, 2010): stress is located on the penultimate syllable if this syllable is heavy, i.e. a diphthong, a long vowel or a vowel ended by a consonant. If the penultimate is light, stress falls on the prepenultimate syllable. If the penultimate syllable is neither closed by a consonant and neither a diphthong, stress is predictable only if we know that its syllable vowel is long or light (which is a property of the lexicon). If the word has only two syllables, stress falls on the penultimate, and if it has only one syllable, this unique syllable is stressed (but only if the word is a noun, an adjective, an adverb or a verb).

The following examples illustrate theses different cases:

in.fer.no has its second syllable closed by the consonant |r| and is therefore heavy, so that the penultimate is stressed: *in**fer**no*.

The syllable *mi* in *a.mi.ca* “friend” contains a long vowel, and the stress is therefore on the penultimate: *a**mi**ca*.

The syllable *ro* in au.ro.ra “dawn” has a light vowel, then the stress of *aurora* falls on the prepenultimate: ***au**rora*.

The word *sp.ina* “plug” has only two syllables. Since there is no other possibility, its first syllable, whether light or heavy, is stressed: ***spi**na*.

Stressed syllables in Romance languages

The position of stress in Romance languages is restricted to a six-syllable window at the right edge of the word (six syllables for verbs and generally up to four syllables for nouns, adjectives, adverbs and other grammatical categories) and is determined by the same morphological rule, originally suggested by Paul Garde (1968, 2013). This rule is based on 1) the stress rules in Latin; 2) a morphological analysis of nouns, adjectives and verbs into their morphological structure and 3) the stressability of suffixes and flections:

(prefix) + stem + (suffixes) + (flections)

Suffixes and flections can be classified as stressable and unstressable, i.e. susceptible to be stressed, or not susceptible to be stressed (unstressable). As most lexical entries in Italian are derived from Latin (excluding borrowed words), the stem follows the Latin stress rule given above.

The stress rule for Romance languages (except French) is very simple: the last stressable morphological element (stem, suffix, flexion) of the word determine the position of the stressed syllable. Given the relatively large number of suffixes and flections homographs, it is important to match corresponding morphological categories (i.e. suffixes and flections for verbs, nouns and adjectives), in order to obtain a correct morphological analysis.

Things may appear occasionally more complicated with homographs either belonging to distinct grammatical categories, or worse (for a computer program) to the same category. An often quoted example in Italian is *sono cose che **ca**pitano capi**ta**no* “these are things that happen captain”, where the first ***ca**pitano* is a verb (3rd person singular of the verb *capitare*) and is stressed on the fourth syllable from the end, whereas the second *capi**ta**no* is a noun (here in its singular masculine form) and is stressed on its penultimate syllable.

Examples of homographs can belong to the same grammatical category. Examples in Italian are ***prin**cipi* “princes” and *prin**ci**pi* “principles”, or ***tur**bine* “whirlwind” (singular, *il **tur**bine*) and *tur**bi**ne* “turbines” (plural of *la turbina*).

An example of homograph in Spanish: ***cé**lebre* “famous”, *ce**le**bre* (from *cele**brar***, “to celebrate” 3rd person present subjunctive of *cele**brar***, “to celebrate”), *cele**bré*** “I celebrated”.

Some examples of various stress placement resulting from the general rule are given below.

Stress on the last syllable (oxyton)

Italian: ***tronco***: *vir**tù*** “virtue”, *ca**ffè*** “coffee”, *ame**rò*** “I will love” (marked in the orthography by a stress mark);

Spanish: ***agudas***: conver**sar** “converse”, *pas**tor*** “pastor”, *ora**ción*** “prayer ” (sometimes marked in the orthography by a stress mark);

Catalan: ***agudas***: na**ció** “nation”, des**prés** “after”, va**lor** “valor”

Portuguese: ***agudas***: *rui**rã**o* "they will collapse"

Romanian: ***tronco***: *cerceta**tor*** “researcher”, *co**bor*** “I descend”

Stress on the penultimate syllable (paroxyton)

Italian: ***piano***: *a**ma**re* “To love”, *natio**na**le* “national”

Spanish: ***llanas***: *li**bro* “book”, *di**fí**cil* “difficult”, *á**n**gel* “angel”, (sometimes marked in the orthography by a stress mark);

Catalan: ***llanas, plana***: *Barce**lo**na* “Barcelone”, ***pla**ça* “place”, *lin**güi**sta* “linguist”

Portuguese: ***plana**: du**vi**da* "he doubts", *fa**la**ram* "they spoke", ***tú**nel* "tunnel"

Romanian: ***paroxytone***: *fân**tâ**na* ‘fountain’

Stress on the antepenultimate syllable (proparoxyton)

Italian: ***sdrucciolo***: *te**le**fono* “telephone”, ***ce**lebre* “famous”, ***pren**dilo* “Take it”

Spanish: ***esdrújulas***: ***prés**tamo* “let’s loan”, *hi**pó**crita* “hypocritical”, *ag**nós**tico* “agnostic”, ***cré**dito* “credit|”, (always marked in the orthography by a stress mark);

Catalan: ***proparoxítono, esdrújulas***: ***tí**pica* **“**typical”, *po**lí**tica* “politic”, (always marked in the orthography by a stress mark);

Portuguese: ***proparoxytone***: ***dú**vida* "doubt" (noun), *di**nâ**micos* “dynamic”, ***lâm**pada* “lamp”

Romanian: ***proparoxytone***: ***mo**dele* “the fashions”, *in**ca**leca* “to mount a horse’”

Stress on the anteantepenultimate syllable (preproparoxyton)

Italian: ***bisdrucciole***: ***ca**ustico* “caustic”, ***fa**bbricalo* “fabricate it”

Spanish: ***sobreesdrújulas***: ***có**metelo* “eat it”, ***trá**emela* “bring it to me”, (always marked in the orthography by a stress mark);

Catalan: ***sobreesdrújulas***: *trans**pór**taselo* “transport it”, ***trá**gatelo* “swallow it” (always marked in the orthography by a stress mark);

Romanian ***doi**sprezece* “twelve”, ***lin**gurile* “the spoons”, ***ve**veriță* “squirrel”

Stress on the anteanteantepenultimate syllable (Prepreproparoxyton)

Italian: ***trisdrucciole***: ***fa****bbricalmelo* “fabricate it for me”

Spanish: (http://www.romaniaminor.net/ianua/Ianua11/01.pdf)

Romanian: ***sia****ptesprezece* “seventeen”

Stress on the anteanteanteantepenultimate syllable (preprepreproparoxyton)

Italian: ***quadrisdruciole***: ***fa****bbricalmecelo* “fabricate it for me to him”

Romanian: ***sia****ptesprezecelea* “seventeenth”

Stressed syllables in French

French has no lexical stress, only a group stress. Progressively from Old French, all segmental units following the accented syllables were dropped, at the exception of a single mute [ə] in certain cases. By this process, the position of stress lost its function of marking morphological boundaries as in the other Romance languages. From lexical the stress became demarcative in French, indicating boundaries not of words but of groups of words, whether content and grammatical, or even of single syllables.

The Incremental Prosodic Structure

The second step in comparing Romance languages prosodic features pertain to melodic contours located on stressed syllables, as these contours are assumed to instantiate prosodic markers indicating the prosodic structure. From the analysis of various examples with increasing syntactic complexity, it is possible to infer a grammar of intonation that would show striking similarities between Romance languages (again except French), despite possible experimental uncertainties pertaining to the assumed congruence between prosody and syntax in the data (Martin, 2015).

The prosodic structure is defined as a hierarchical grouping of minimal prosodic units, instantiated by accent phrases (aka prosodic words, stress groups…). These groupings, operated dynamically along the time axis by the listener to reconstitute the prosodic structure intended by the speaker, is indicated by prosodic markers instantiated by melodic contours located on accent phrases stressed syllable which has no emphatic function. In certain configurations, melodic contours located on accent phrases final syllables, if not stressed, are part of a complex prosodic marker together with the melodic change located on the stressed syllable. If the accent phrases last syllable is stressed, the complex contour results from a two distinct melodic movements occurring partly on the stressed syllable and partly on the final syllable of the accent phrase.

The melodic movements located on accent phrases stressed (and final) syllables are not realized at random. Their acoustical characteristics in term of rise or fall, high or low, long or short, constitute the material to implement phonological features which indicate dependency relations between accent phrases.

These dependencies operate “to the right”, i.e. towards another accent phrases carrying a specific contour planned in the immediate future by the speaker, to indicate that the given accent phrases (or all accent phrases already part of a group ended by the given accent phrases) carrying the first contour has to be merged with another accent phrases carrying the other contour placed further in the sentence from which it depends. For example, in French, the occurrence of a well-documented contour usually called *continuation majeure* (after Delattre, 1966), presupposes the occurrence of a terminal conclusive contour later in the sentence. In terms of dependency relations, the *continuation majeure* depends on the assumed occurrence of a final conclusive contour ending the sentence, even if it will appear in the future of the pronounced sentence.

The dependency relations indicate to the listener how and when to merge the prosodic syntagms (a group of accent phrases) ended by a *continuation majeure* and by a terminal conclusive contour to form the overall prosodic structure of the sentence. The dependency relations are not limited to the grouping of *continuation* majeure and terminal contour. They function as well at lower levels of the prosodic structure, where other types of contour do indicate a relation of dependency towards a *continuation majeure*, etc.

Prosodic grammar

The object of prosodic research is to determine the phonological features of the contours located on accent phrases stressed syllables, and to discover the underlying grammar which implement the dependency relations between contours. Another remarkable point pertains to the time linear properties related to the processes of encoding and decoding the prosodic structure.

Considering that prosodic events instantiated by melodic contours occur not simultaneously but one after the other on the time line instantiated by the sequence of syllables, it can be shown (Martin, 2015) that it is necessary and sufficient to evaluate dependency relation between two consecutive contours, provided a ranking between phonological contours has been established.

If C0, Cc, C1, C2, Cn designate classes of prosodic events instantiated by melodic contours located on accent phrases stressed syllable (and essentially on its vocalic nucleus) defined as follows:

C0: terminal conclusive contour (declarative case), falling and low

Cc: complex contour, flat or slightly falling on the accent phrase stressed syllable, and rising on the accent phrase final syllable

C1: rising above the glissando threshold (i.e. above a parametric rate of melodic change)

C2: falling above the glissando threshold

Cn: neutralized, falling, flat or rising below the glissando threshold

The ranking of prosodic contours in French is Cn < C2 < C1 < C0, and presents an inverted ordering C1 < C2 for the other Romance languages:

Cn < C1 < C2 < Cc < C0.

Given these differences, the prosodic grammar operates the same way in French and in the other Romance languages. By comparing two successive melodic contours, say Cx and Cy, relative to their ranking, the listener is able to assemble or not the prosodic words implied:

if Cx < Cy, the accent phrases attached to Cx and Cy are merged [Cx Cy]

else if Cx = Cy, the accent phrases attached to Cx and Cy are part of a list, to be terminated by the occurrence of a contour of higher rank [Cx Cy …

else if Cx > Cy, the accent phrases attached to Cy is not merged with the one attached to Cx [Cx [Cy…

This process is local as it involves only differences between two successive contours in the same domain. The examples of Figure 1 (Italian), Figure 2 (Portuguese) and Figure 3 (French) illustrate the mechanism of local dependency relations leading to the (re)construction of the prosodic structure, using the same set of melodic contours and the same prosodic grammar. The resulting prosodic structures presented in these figures show different cases of local non-congruence with syntax.

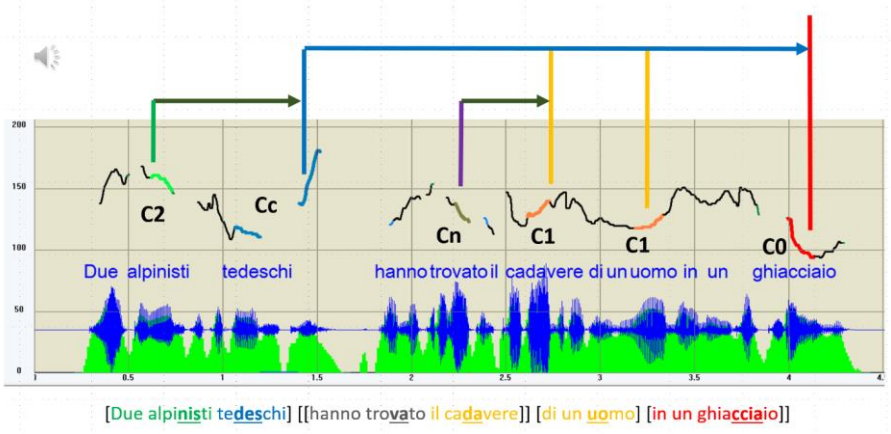


Figure 1. Italian example of prosodic structure built by increments along time axis.

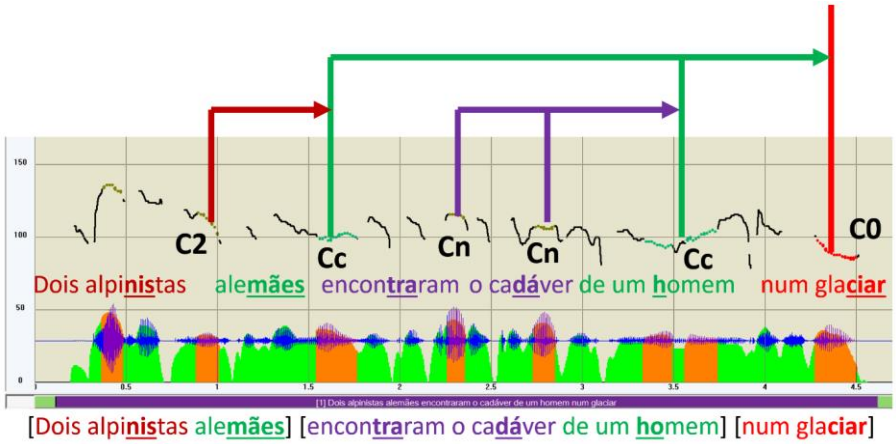


Figure 2. Portuguese example of prosodic structure built by increments along time axis.

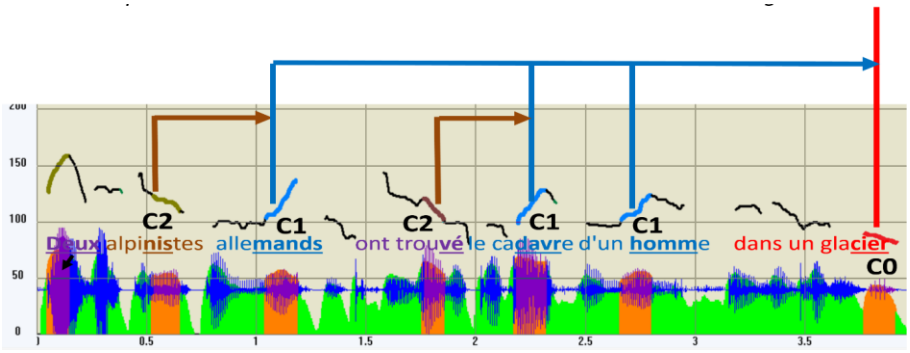


Figure 3. French example of prosodic structure built by increments along time axis.

Conclusion

No language is likely to escape the constrain of generating and decoding the sentence prosodic structure. However, it may be more surprising that Romance languages (except French) would use the same phonological melodic contours and the same grammar of intonation to encode the prosodic structure, leading to suggest that the melodic contours and the grammar that describe their use are inherited from Latin, despite the large differences in phonology, morphology and syntax existing among the languages derived from Latin.

The remanence of phonological prosodic features among Romance languages (including French when the absence of lexical stress is considered) is remarkable and pertain to the following topics:

1. The position of lexical stress
2. The classes of melodic contours
3. The grammar of melodic contours as dependency markers between accent phrases

Stress placement in the accent phrase is clearly derived from the classical Latin stress rules, with the addition of suffixes ad flexions classified as stressable or unstressable. The same simple stress rule applies to all Romance languages. Classes of melodic contours are phonologically similar, with the exception of French which has no complex contour Cc since it has no lexical stress. Finally, the principle of contrast of melodic slope also applies to all Romance languages, French deprived of the complex contour Cc using another ranking in the prosodic grammar Cn > C2 > C1 > C0, instead of Cn < C1 < C2 < Cc < C0.

References

Alkire T., C. Rosen, 2010. *Romance Languages, an Historical Introduction*, Cambridge University Press.

Delattre P. 1966. Les dix intonations de base du français, *French Review* 40, 1-14.

Garde P. 1968. *L'accent*, PUF, Paris, 172 p. / (2013) Lambert-Lucas, Paris.

Martin, Ph. 2015. The Structure of Spoken Language. Intonation in Romance, Cambridge University Press.

Frota, S., P. Prieto (eds.), 2015 Intonation in Romance. Oxford University Press.

Rich Reduction: Sound-segment residuals and the encoding of communicative functions along the hypo-hyper scale

Oliver Niebuhr
Mads Clausen Institute, IRCA, University of Southern Denmark, Alsion, Denmark

Abstract

The H&H (Hypo-Hyper) Theory of Lindblom (1990) is probably one of most prominent theories of the phonetic sciences. It was put forward at a time when research on speech reduction started to undergo a shift in focus from the description and linguistic embedding of phonological processes to questions about their phonetic details, contextual factors, perception, and cognitive processing. This shift in focus, in combination with the application of digital technologies and resources have fundamentally changed our knowledge of speech reduction. The present chapter will argue with reference to examples from different languages - and in accord with Lindblom's own expectation - that his well-known "tug-of-war" metaphor needs to be adapted in the light of these changes. The "tug-of-war" metaphor conceptualizes the realized degree of reduction as a compromise between economic and intelligible speech. However, first, growing perception evidence questions the metaphor's key assumption that more articulatory economy and hence a higher degree of reduction make speech less intelligible for listeners. Moreover, a one-dimensional hypo-hyper continuum controlled by two antagonistic forces (speaker and listener) ignores that fact that communicative functions are another separate driving force for variation in the degree of reduction. Therefore, the author suggests to abandon that the tug-of-war metaphor in favor of an adaptation Bolinger's famous wave metaphor.

Introduction

Managing and, ideally, explaining phonetic variation has ever since been a key issue in the speech sciences. But, it became even more obvious with the beginning of the "acoustic age" after World War II, when the US military declassified the invention of the sound spectrograph. It made speech a precisely analyzable research object. The radiply developing computer technology made this research object accessible to a growing community of phoneticians (Mattingly 1999), which, in turn, multiplied the number of questions on phonetic variation and their levels of detail and complexity. Phonetic variation supported the development of phonetics and phonology as two different disciplines and later expedited the "divorce" of those disciplines, with phonology taking care of the well-formed structures of clearly defined sound (or intonation) categories and their rule-based changes, and with phonetics measuring the messy, highly variable articulatory and

Proceedings of 7th ExLing 2016, Saint Petersburg, Russia.

acoustic signals and trying to project them across speakers, genders, speaking styles, and communicative situations onto the "ideal-world" categories of phonology. In this context, it was not surprising that phonology soared to dominate phonetics for decades in the 20th century, and that the joint efforts of the two disciplines primarily aimed at marginalizing or abstracting away from phonetic variation by searching for the invariant characteristics of sound (or intonation) categories.

The motor theory of speech perception, which attracted a lot of attention during that time, is a role model of these efforts (which is not to say that the basic idea of "covered mimicry" has no empirical foundation, see, e.g., Watkins et al. 2003). This theory saw phonetic variation as a troublemaker. Thus, the aim of both listeners and researchers had to be to free the sound segments from their variable acoustic ingredients and the resulting "encumbering auditory baggage that would make them all but useless for their proper role as vehicles of language" (Liberman 1982:148). The research paradigm of categorical speech perception nicely reflects this approach to phonetic variation (see Holt & Lotto 2010 for a summary).

The later-emerging articulatory phonology (Browman & Goldstein 1992) is similar to the motor theory in that it is also rooted in the articulatory domain, assuming that consistent patterns are to be found only in articulation, whereas the acoustic patterns of speech are intrinsically variable. However, unlike the motor theory, articulatory phonology was explicitly designed to explain phonetic variation. That is, allophonic variation of speech sounds, for example in terms of voicing/VOT, were made conceptually understandable by means of (changes in) the temporal coordination of glottal and supraglottal gestures, conflicting articulatory commands to the same articulator were used to explain strong coarticulation and the blending of sounds, and the disappearance of sounds at the acoustic level were attributed to an extreme overlap of supraglottal gestures, hence postulating that they are only "masked" but nonetheless consistently there at the level of articulation.

Despite its deserved success, articulatory phonology was also criticized, among others by Kohler (1992). One of his major points was that the rules and restrictions according to which the gestural score is organized are probably not able to explain the full range of variations, especially those that relate to common strong speech reduction patterns in spontaneous speech. Moreover, Kohler points out that the rules and restrictions on which articulatory phonology is built themselves need to be externally motivated and supported by independent empirical evidence. Kohler's suggestion in this context is to go beyond the speaker and explain phonetic variation, i.e. its sources as well as its implications for communication, by means of a

theoretical framework that also takes into account the listener and his/her cognitive abilities and processes.

The latter is exactly what was done by Lindblom (1990) in his very influential H&H theory. "Explaining phonetic variation" (p.403) is the explicit aim of Lindblom's theory. It compares speech communication to a tug-of-war, with speaker and listener pulling the rope that represents phonetic variation in opposite directions, see Figure 1. The speaker follows a basic ethological principle of all mammals, i.e. striving for economy. Accordingly, the speaker's aim is to minimize the articulatory effort invested in speech production and hence reduce the speech signal as much as possible. The extent to which this is possible is defined by the listener at the other end of the rope: The speech signal has to contain at least enough phonetic information to allow the listener understand the message conveyed by the speaker. In other words, speakers want to produce "hypospeech", and listeners want to hear "hyperspeech".

Figure 1: Illustration of the tug-of-war metaphor in the H&H theory of Lindblom (1990). The drawing was made by Nathalie Schümchen.

On this basis, the key concept of the H&H theory is that, at each point of the conversation, the level of speech reduction is an implicitly negotiated compromise along the hypo-hyper scale between speaker desires and listener demands. A further key concept is that this dynamic, adaptive compromise takes into account not only basic factors like speaker physiology (e.g., gender, emotions, pathologies) and the environmental acoustics of the communication situation. The compromise is also made with respect to the listener's metalinguistic top-down knowledge and context-driven expectation about which units, functions, and meanings will be contained in the upcoming speech signal. This allows the speaker to be less clear in or even completely omit those acoustic cues which s/he knows that the listener can

add in the process of speech perception. This idea was probably the H&H theory's most important contribution to speech sciences. It replaces invariance by sufficient contrast and hence goes beyond the common picture of speech as a machine-like self-contained code that is encoded on the side of the speaker and transmitted through the air with all elements that the listener requires to decode it. In contrast, all that speakers need to do in Lindblom's framework is, broadly speaking, to be sufficiently clear, feed their listeners with a sufficient number of acoustic cues, and then let their top-down processes do the rest, i.e. interpret the signal by matching it against knowledge and expectations, and, if necessary, fill in gaps.

Many studies provide empirical support for the H&H theory. For example, Hunnicutt (1985) concluded from the results of a combined production-perception experiment that speakers hyperarticulate more if words are less predictable in a given semantic (sentence-frame) context. Fowler & Housum (1987) showed by means of radio news broadcasts that repeatedly stated words are more hypoarticulated (i.e. reduced) by speakers. Similarly, Wright (2003) found "easy" words, i.e. frequent words with relatively few lexical competitors, to be more strongly hypoarticulated than "hard" words. Finally, we know from a number of experiments that speech produced under adverse conditions such as noise or greater spatial distances between the dialogue partners is produced with more effort both articulatorily and phonatorily (Traunmüller & Erickson 2000; Junqua 1996).

Despite this converging evidence in favor of H&H, we should not lose sight of one crucial fact: Lindblom's framework never aimed at explaining phonetic variation in general. Rather, the framework was developed to explain that phonetic variation that is relevant to and emerges in connection with "successful lexical access" (Lindblom 1990:405). However, we know at least since the rise of intonational phonology (Ladd 2008) that speech communication is not only about words. Lindblom himself notes that speech is "produced not only in the laboratory but also in its natural, ecological settings" (p.418), and he stresses in this context that the assumption of only two antagonistic forces that create the one-dimensional reduction continuum from hypo to hyper is a "deliberate simplification that is likely to be revised in the course of future work" (p.419).

In fact, Lindblom's H&H theory was taken up and further elaborated, for example, in terms of the smooth signal redundancy hypothesis of Aylett & Turk (2004). In simple terms, the hypothesis states that the total degree of reduction used by speakers is understandable as the sum of two types of redundancy: language redundancy (e.g,, due to syntactic order or grammatical agreement) and signal redundancy (e.g., several acoustic cues on the same phonological distinction). Aylett & Turk assume that speakers strive to keep the total redundancy constant, which means that a lower

language redundancy is compensated by a higher signal redundancy (i.e. hyperspeech), whereas a higher language redundancy allows for a lower signal redundancy (i.e. hypospeech). As is obvious from these explanations, Aylett & Turk refined rather than revised Lindblom's H&H framework, keeping intact the central tug-of-war metaphor and its two antagonistic forces, which are called "conservation of effort" and "reliable communication" in Aylett & Turk's terminology. The same applies to many other works and concepts that are inspired by H&H, such as uniform information density, communicative efficiency, and audience design, see Clopper & Turnbull (submitted) for a summary.

In summary, despite Lindblom's own expectation, his deliberate simplification of a one-dimensional hypo-hyper scale has not been significantly addressed. The present paper is intended to pave the way for initial steps in this direction by pointing the readers to two basic aspects in which H&H, in the author's own humble opinion, overestimates and oversimplifies variation in the degree of reduction.[1]

The supposed harmfulness of reduction

A key premise of the tug-of-war metaphor in Lindblom's H&H theory is that reduced articulatory effort on the side of the speaker and the resulting reduction phenomena in the speech signal[2] put pressure on the listener, for example, in that the listener has to rely more on his/her cognitive top-down processes to compensate for missing acoustic cues associated with reduced sound segments or entire meaningful elements. A growing body of production and perception evidence that has accumulated after the H&H theory was published raises doubts about the general validity of this tug-of-war premise.

Nolan (1992) summarizes EPG data collected by W. Barry and P. Kerswill on alveolar-to-velar place assimilation in British English stop consonant sequences like "road collapsed" and "lead covered" (/dK/). The EPG data show a range of more or less strongly assimilated productions, but also cases in without any trace of a tongue contact at the alveolar ridge. These "zero-alveolar" cases (in Nolan's terminology) are indistinguishable

[1]All references involving Niebuhr in the following sections 2 and 3 that are not including in the list of references can be found in Cangemi et al. (in press).

[2]Note that there are actually two different types of reduction: (1) the amount of energy invested in articulation and phonation, and (2) deviations from full/ideal citation forms of consonants, vowels, and words. The two types are equated here. The author is aware of the fact that this is probably a simplification (Yi Xu, pers. comm.), but one that does not affect the line of argument presented here.

from "non-alveolar" cases representing actual "rogue collapsed" and "leg covered" (/gk/) realizations. Yet, Nolan found in a 2AFC word-identification test (based on newly recorded stimuli) that listeners are to a significant degree able to perceive the assimilated /d/ of which there is no EPG trace left and thus keep zero-alveolar cases of "road collapsed" and "lead covered" separate from non-alveolar cases of "rogue collapsed" and "leg covered". These striking results led Nolan to the conclusion that, even in the absence of an alveolar contact or closure, "the tongue configuration in realizing lexical /dg/ sequences [...] is subtly different from that for /gg/ sequences" (Nolan 1992:272). This subtle difference persists as a difference in vowel quality that functions as an acoustic cue to /d/ even when measurements suggest that this sound segment itself has fully disappeared. Later, an acoustic analysis by Local (2003) provided supporting evidence for Nolan's assumption and his impression that, "auditorily, [...] the vowel allophone before the lexical velar is slightly closer than before the lexical alveolar" (Nolan 1992: 272).

A very similar case to that of Nolan (1992) and Local (2003) was found for French by Niebuhr & Meunier (2011). They investigated /s/-to-[ʃ] place assimilation in word sequences like "trousse chargée" (full bag) and "fils charmant" (charming son). Their acoustic spectral center-of-gravity measurements revealed a production continuum from weakly to fully assimilated /s/ sibilants. Moreover, Niebuhr & Meunier found differences in the preceding vowel that were there independently of the degree of /s/-to-[ʃ] assimilation and even remained if the original /sʃ/ sequence was acoustically indistinguishable from an actual /ʃʃ/ sequences. Vowels preceding /s/ were shorter, had a higher acoustic energy, and a less breathy voice than those before /ʃ/. A later pilot perception experiment (Clayards & Niebuhr 2011) based on the identification of pseudo names in forename-surname sequences demonstrated, as in Nolan (1992), that listeners used these vowel cues to identify a /s/ even if the sound segment itself became (according to the measurements taken) indistinguishable from /ʃ/.

Examples like those above stress the relevance of a concept that was developed by Kohler (1990) and is hence as old as the H&H theory: "Articulatory prosodies". At the heart of the concept lies the statement that reduction processes do not necessarily cause a loss of acoustic cues and in this way undermine the richness of the speech signal. In spontaneous speech, reduction is the rule rather than the exception, and the phenomena subsumed under reduction represent processes by which the packaging scheme of acoustic cues in the form of sequences of linear sound segments is broken up. The affected sound characteristics or acoustic cues are then reshaped as long-term resonances, i.e. articulatory prosodies, that are superimposed on the remaining sound segments. Niebuhr (2008) elaborated this concept by

adding the notion of "phonetic essence", see Niebuhr & Kohler (2011) and Kohler & Niebuhr (2011). Phonetic essence is a feature of complex sound sequences like words, and the assumption is that, in speech reduction, those sound characteristics of the sequences are maintained and reshaped as articulatory prosodies that belong to the sequence's phonetic essence.

For instance, the German modal particle "eigentlich" (actually) is characterized by palatality that pervades virtually the entire word: [aɪgŋtlɪc]. An analysis of the Kiel Corpus of Spontaneous Speech (Peters 2005) showed that "eigentlich" can be severely reduced, with only the initial diphthong and, maybe, the middle nasal being left at the segmental level: [aĩj̃(ɲ̃)]. However, in these cases the palatality of the lost sound segments is maintained by strengthening and lengthening the palatality in the initial diphthong. That is, the closed-vowel element is produced longer and with a higher F2 frequency. A perception experiment conducted by Niebuhr & Kohler (2011) showed that listeners have no problems interpreting this articulatory prosody of palatality and distinguishing highly reduced "eigentlich" from the segmentally similar unreduced word "ein" (indef. article). Likewise, the study Kohler & Niebuhr (2011) addressed the word "ihnen" - [i:nʲɪnʲ] (to you) - whose separate segmental representation completely disappeared in the sentence frame "ich kann ihnen das ja mal sagen" (I can mention this to you) produced by speaker TIS in the Kiel Corpus of Spontaneous Speech. Despite the loss of all segments, the phonetic essence of palatality of "ihnen" was kept and superimposed by the speaker on the segments of "kann" and "das" that, as a result, change from [kʰa̲nna̲s] to [k̟ʰɛ̈nʲnʲə̟s]. Evidence from a perception experiment showed that listeners can reliably perceive the entire word "ihnen" on this basis of [k̟ʰɛ̈nʲnʲə̟s] in the sentence frame "Ich ___ ja mal sagen". Moreover, as the sound segments of [k̟ʰɛ̈nʲnʲə̟s] were successively replaced by those of [kʰa̲nna̲s], the perceived wording of the stimulus sentence changed to "ich kann das ja mal sagen" (I can mention this), without "ihnen".

Further phonetic essences that are reshaped as articulatory prosodies and whose perceptual relevance was been experimentally demonstrated are velarization (Niebuhr 2008), glottalization (Kohler 1999), and lip rounding (Niebuhr & John 2014). In all these examples, the articulatory prosodies were reduced representatives of at least entire syllables, and in the case of Niebuhr (2008) the velarization even represented two full words, i.e. "auch noch" (as well).

Articulatory prosodies almost always co-occur with duration cues in the form of a compensatory lengthening of segments in the vicinity of disappeared segments. However, while articulatory prosodies are sufficient to make listeners perceive segmentally disappeared syllables or words, mere segmental lengthening is not sufficient (cf. Niebuhr & Kohler 2011). It must

be temporally coordinated with the articulatory prosodies and/or affect those remaining sound segments that reflect the relevant phonetic essence in order to function as a cue to disappeared syllables or words. Therefore, the question of whether duration or segmental lengthening may be considered a separate articulatory prosody is not yet settled. Interesting in this context is the work of Dilley & Pitt (2010). They manipulated the relative duration of vowels like [ə˜ː] in the middle of phrases like "leisure time", for example, by means of lengthening the vowel. The results of the corresponding perception experiment showed that this change in relative duration makes listeners perceive an additional "or" in the between the two words. That is, "leisure time" became "leisure or time", without adding any further phonetic sound features.

The question has been raised whether this duration-based appearance of words is a one-step all-or-nothing phenomenon, or whether the number of additionally perceived words is correlated with the degree of the increase in relative duration. This question was addressed by Dilley together with Evelin Graupe and the author of this paper in a joint study on German (cf. Graupe et al. 2014). The starting point was the fact that, in German, there are several function words each of which can be reduced to a single alveolar nasal [n]. This includes, for example, "in", "ihn", den", "einen", and "denn". Niebuhr et al. designed stimulus sentences (rhetorical questions) like "Wer braucht Nachrichtensprecher im Radio?" (Who needs newsreaders on the radio) and "Wer findet Nebendarsteller erwähnenswert?" (Who finds supporting actors worth mentioning) and manipulated the relative duration of the initial alveolar nasal of the target nouns, i.e. "Nachrichtensprecher" (newsreader) and "Nebendarsteller" (supporting actor). The semantic contexts of the stimulus sentences basically allowed the relative duration manipulation of /n/ to trigger the appearance of two additional words: "denn" (then, intensifying particle), which corresponds to one syllable, "einen" (indef. article), which corresponds to two syllables, or "denn einen", which corresponds to three syllables. The results clearly show that the stronger the relative duration increase of /n/, the more syllables are perceived by listeners. Weak /n/ lengthening makes the monosyllabic word "denn" appear, strong lengthening the disyllabic word "einen", and very strong lengthening triggers the perception of the whole trisyllable "denn einen".

Results like these emphasize once more that even the most severe segmental reduction need not make the speech signal poorer and ambiguous and speech perception a harder or impossible task for listeners. Moreover, the gradient relationship of relative segment duration and the number of appearing syllables suggests that duration is in fact another independent articulatory prosody rather than just a concomitant feature of palatality, velarization, glottalization, and lip rounding.

Meaningful variation in reduction

The tug-of-war metaphor implies that there are only two parties pulling on the rope: speaker or "articulatory economy" and listener or "sufficient signal contrast". Lindblom himself calls this a deliberate simplification; and, in fact, a growing body of evidence supports this assessment. There is at least one more factor that shifts the degree of speech reduction along the hypo-hyper scale: communicative function.

For example, in the domain of prosody, Niebuhr (2008, 2012) showed that the phonetics of voiceless fricative sounds co-varies with intonation such that the spectral-energy distribution and resulting spectral-pitch impression they convey fit in with the level of the adjacent F0. That is, voiceless fricatives sound "brighter" in high-F0 contexts due to more acoustic energy at higher noise frequencies, and "darker" in low-F0 contexts due to more acoustic energy at lower noise frequencies. Given the fact that the postalveolar sibilant [ʃ] is creates an intrinsically "darker" sound than the alveolar sibilant [s] (also because [ʃ] is produced with lip rounding in German), Niebuhr et al. (2011) wondered whether the degree of /s/-to-[ʃ] (i.e. "bright"-to-"dark") assimilation in German would be affected by the F0 context. The speech-production study they conducted with native speakers of German confirmed this prediction. The degree of /s/-to-[ʃ] assimilation, determined on the basis of spectral center-of-gravity measurements, was weaker in high F0-peak contexts, in this way giving the entire sibilant sequence a "brighter" sound quality. Assimilation was stronger and made the sibilant sequence sound overall "darker" in low F0-valley contexts. Note that the sibilant sequence's total duration did not differ between significantly between the F0-peak and F0-valley contexts. This fact supports the assimilation interpretation as it means that individual sibilants were not simply produced longer or shorter. Together with the results of perception experiments on the integration of F0-based pitch and fricative-based pitch impressions (Mixdorff & Niebuhr 2013; Welby & Niebuhr 2016), the findings of Niebuhr et al. (2011) show that assimilation and hence speech reduction vary in order to support conveying intonational meanings.

An even better demonstration of the fact that reduction patterns are systematically related to communicative functions are the studies of Local et al. (1986). They examined word-final /ptk/ in British English whose realizations can vary from unreduced post-aspirated stops to highly reduced stretches of glottalization. Local et al. scrutinized the claim in the literature that these differences in the degree of reduction are just random, i.e. free variation (cf. Kreidler 1989). In fact, their analysis of a dialogue corpus of Tyneside English revealed quite the opposite: With only a few counterexamples (1-3%, n=206) unreduced stop variants occurred in turn-final position, whereas all reduced variants, including glottalized ones, were

produced turn-internally. Docherty et al. (1997) replicated the findings of Local et al. for a corpus Southern Standard British English. Zellers (in press) analyzed phrase-final consonants in a Swedish dialogue corpus and found that "consonant reduction [...] can further help distinguish between turn change and turn hold contexts in Swedish conversation".

Niebuhr et al. (2013) took up the findings of Local et al. and showed for the Kiel Corpus of Spontaneous Speech that the degree of reduction of the most frequent word ending in German, <-en> /ən/, is highly systematically linked with the distinction between phrase and turn boundaries. Among the approximately 5,700 analyzed tokens, the number of schwas was higher in turn-final <-en> realizations, and even if the schwa was absent, the majority of /n/ nasals showed no place assimilation with the preceding consonant. The opposite applies to turn-internal <-en> realizations whose final nasals were assimilated to either [m] or [ŋ] in about 70% of all cases. Going beyond the studies on English and Swedish cited above, Niebuhr et al. also conducted a perception experiment with degree of <-en> reduction as an independent variable. The experiment was based on a discourse completion task and showed that listeners waited longer with taking the turn and responding to the preceding stimulus if the latter ended in an unreduced <-en> ending. Niebuhr et al. conclude in view of this behavioral evidence that the degree of word-final reduction has a discourse organization - or, more specifically - a turn-taking function. Interestingly, the production data gathered so far point to a language-specific form-function mapping. While in German and probably also in English less or no reduction signals a speaker's turn-yielding intention (in combination with other prosodic cues), Zeller's data suggest the same function could be cued by stronger reduction in Swedish. This is another argument in favor of the non-mechanistic, function-driven nature of phrase-final reduction patterns.

A final example for meaningful variation in reduction is related to speaker attributes or attitudes. Trede (2011) conducted a production study in which she analyzed the phonetic exponents of sarcastic irony in German by means of a comparison of two sets of sarcastic and sincere utterances. Trede replicated previous results (e.g, Bryant 2010) in that she found sarcastic utterances to have lower speaking rates and lower and less variable F0 and intensity contours than sincere utterances. In addition, she counted the number of reductions (assimilations, elisions, lenitions) in each utterance by relating the words' actual realizations to their canonical reference forms. This data showed that sarcastic irony is not only marked by stronger prosodic reductions, but also by stronger segmental reductions. Informal perception tests suggest that these stronger segmental reductions represent a separate cue to sarcasm. This preliminary conclusion fits in well with unpublished perception findings of Niebuhr (in preparation) showing that strong

reduction patterns made utterances sound less sincere. These unpublished data also show that the degree of segmental reduction is significantly positively correlated with a speaker's perceived level of education, clumsiness, scattiness, tiredness, and vanity.

“Offshoring” the tug-of-war metaphor

The examples provided in section 2 of this paper aimed to show that reduction, even when it eliminates entire syllables or words in spontaneous speech, does not necessarily pose a challenge for listeners. This troublemaker-view on reduction is driven by the long-established concepts of 'phoneme' and 'canonical form' both of which are currently being controversially discussed (Cangemi et al, in press). For instance, a phoneme-based, segment-oriented perspective on speech with full canonical word forms at the starting and end points of the speech chain overlooks that critical features of deleted sound segments can still be present in the form of articulatory prosodies, and that canonical forms may not always be proper references, for example, in the sense of the most frequent realization of a word in spontaneous speech.

Furthermore, a second set of examples in section 3 this paper illustrated that the degree of reduction cannot consistently be conceptualized as the result of the two antagonistic forces 'articulatory economy' and 'sufficient contrast', dynamically negotiated on the basis of environmental, social, psychological, and maybe pathological factors. Rather, variation in the degree of reduction can also be meaningful. That is, communicative functions at the levels of intonation, discourse organization, and speaker attitudes/attributes are associated with systematic reduction differences; and, for an increasing number of these production studies, perception experiments show that listeners process and use these differences like any other segmental or prosodic cues in order to identify the corresponding communicative function.

The major contribution of Lindblom's H&H theory was to replace the futile search for invariance by an explainable variance based on the tug-of-war metaphor. The notion of articulatory prosodies and the functional role of reduction both suggest the next steps along the line of argument opened up by Lindblom. Specifically, we need to supplement Lindblom's explanatory framework and revise the speaker-listener conflict that lies at the heart of the tug-of-war metaphor.

The author's suggestion would be to "offshore" the tug-of-war metaphor and replace it by the ocean metaphor of Bolinger (1964), with the ups and downs at the surface of the ocean representing the speaker's variation along the hypo-hyper scale and wavelength corresponding to the time domain of the reduction variation. As is illustrated in Figure 2, the ups and downs are

the combined result of tides, waves, and ripples. Tides are long-term settings in the degree of reduction determined by, for example, the communication channel, the situation, the physiological and pathological properties of speaker and listener and the (acoustic) environment in which their communication takes place. Waves and ripples represent additional meaningful or otherwise systematic (e.g., tailored to integrate the listener's top-down processes) short-term variations along the hypo-hyper scale, associated with phrases, words, or single sounds and syllables. This metaphor is compatible with later refinements of Lindblom's H&H theory, such as the smooth signal redundancy hypothesis.

Figure 2: Reframing the tug-of-war metaphor in the form of the ocean metaphor of Bolinger (1964).

Given the fairly incomplete empirical picture outlined sections 2 and 3, it would be premature to try to associate waves and ripples with different sources of systematic variation in speech reduction. However, as a point of departure, it seems that meaningful reduction variation due to conveying speaker attitudes/attributes as well as reduction variation reflecting the speaker's anticipation and integration of listener knowledge are both more likely to manifest themselves as waves, i.e. at the level of phrases or words, whereas the segmental reduction differences realized in connection with different intonation contours show up as ripples. Reduction variation related to discourse functions, as in the example of turn-final and turn-internal syllables in English and German, could sometimes surface as waves and sometimes as ripples, and may represent a third type of wavelength. The ocean metaphor leaves room for distinguishing additional "wavelengths", for example, longer deep-sea waves and shorter coastal waves; and finding out

whether or not such additional distinctions are necessary will be one of the interesting tasks of follow-up studies on speech reduction.

In fact, the ocean perspective on reduction opens up a completely new field of questions concerning, for example, the temporal interplay (superposition, coordination, alignment) of reduction phenomena with similar/different wavelengths, the limits of wave amplitudes, correlations between types of waves and wave amplitudes as well as between wave amplitudes and the overall (sound) energy level that is fed into the wave system, and, finally, geographical and coastal (i.e. in the case of speech cultural and phonological) differences. These and many other questions have the potential to stimulate, reconsider, and inspire research in speech reduction for many more years.

Acknowledgments

First of all, I would like to thank Meghan Clayards and Meg Zellers for their useful and insightful comments on earlier drafts of this paper. Moreover, I am greatly indebted to Meg Zellers for taking the time to proof-read the paper. Finally, special thanks are due to Yi Xu, Antonis Botonis and many other participants of ExLing as well as all authors and co-editors of the "Rethinking Reduction" volume for inspiring discussions and contributions on the issue(s) of speech reduction.

References

Aylett, M., A.E. Turk. 2004. The smooth signal redundancy hypothesis. Language and Speech 47, 31–56.

Bolinger, D. 1964. Around the Edge of Language. Harvard Educational Review 34, 282-293.

Browman, C.P., Goldstein, L. 1992. Articulatory phonology: An overview. Phonetica, 49, 155-180.

Byrant, G.A. 2010. Prosodic contrasts in ironic speech. Discourse Processes 47, 545-566.

Cangemi, F., M. Clayards, O. Niebuhr, B. Schuppler and M. Zellers (eds). in press. Rethinking Reduction. Berlin: de Gruyter.

Clayards, M., O. Niebuhr. 2011. Production and Perception of Sibilant Assimilation: Do French and English differ? Presentation at the Sound-to-Sense Closing Workshop, Faculty Club Leuven, Belgium.

Clopper, C.G. and R. Turnbull. submitted. Exploring variation in phonetic reduction. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, M. Zellers (eds.), Rethinking Reduction. Berlin: de Gruyter.

Dilley, L.C., M. Pitt. 2010. Altering context speech rate can cause words to appear or disappear. Psychological Science 21, 1664–1670.

Docherty, G.J., J. Milroy, L. Milroy, D. Walshaw. 1997. Descriptive adequacy in phonology: A variationist perspective. Journal of Linguistics 33, 275-310.

Fowler, C. A. and J. Housum. 1987. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. Memory and Language 26, 489-504.

Graupe, E., K. Görs, O. Niebuhr. 2014. Reduktion gesprochener Sprache - Bereicherung oder Behinderung der Kommunikation? In O. Niebuhr (ed.), Formen des Nicht-Verstehens, 155-184. Frankfurt: Peter Lang.

Holt, L.L., A.J. Lotto. 2010. Speech perception as categorization. Atten Percept Psychophys 72, 1218-1227.

Hunnicutt, S. 1985. Intelligibility vs. redundancy - conditions of dependency. Language and Speech 28, 47-56.

Junqua, J.-C. 1996. The Influence of Acoustics on Speech Production. Speech Communication 20, 13-22.

Kohler, K.J. 1992. Gestural Reorganization in Connected Speech: A Functional Viewpoint on "Articulatory Phonology". Phonetica 49, 205-211.

Kohler, K.J. 1999. Articulatory prosodies in German reduced speech. Proc. 14th International Congress of Phonetic Sciences, 89-92, San Francisco, USA.

Kreidler, C.W. 1989. The pronunciation of English. Cambridge: Blackwell.

Ladd, D.R. 2008. Intonational Phonology. CUP.

Liberman, A.M. 1982. On finding that speech is special. American Psychologist 37, 148-167.

Lindblom, B. 1990. Explaining phonetic variation. In W. Hardcastle, A. Marchal (eds), Speech production and speech modelling, 403-439. Dordrecht: Kluwer.

Local, J., J. Kelly, W.H.G. Wells. 1986. Towards a phonology of conversation: Turn-taking in Tyneside English. Journal of Linguistics 22, 411–437.

Local, J. 2003. Variable domains and variable relevance: interpreting phonetic exponents. Proc. TIPS, 101-106, Aix-en-Provence, France.

Mattingly, I.G. 1999. A short history of acoustic phonetics in the U.S. Proc. 14th International Congress of Phonetic Sciences, 1-6, San Francisco, USA.

Niebuhr, O. 2008. The identification of highly reduced words by differential segmental lengthening. Presentation at the First Nijmegen Speech Reduction Workshop, MPI, Nijmegen, The Netherlands.

Nolan, F. 1992. The descriptive role of segments. In D.R. Ladd, G.J. Docherty (eds.), Papers in Laboratory Phonology 2, 261–280. CUP.

Peters, B. 2005. The Database The Kiel Corpus of Spontaneous Speech. AIPUK 35a, 1-6.

Traunmüller, H., A. Eriksson. 2000. Acoustic effects of variation in vocal effort by men, women, and children. JASA 107, 3438-3444.

Trede, D. 2011. Ist Ironie nur Prosodie? Zu lautlichen Reduktionen ironischer und nicht-ironischer Äußerungen. BA thesis, Kiel University, Germany.

Watkins, K.E., A.P. Strafella, T. Paus. 2003. Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia 41, 989–994.

Wright, R. 2003. Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, R. Temple (eds), Papers in Laboratory Phonology VI, 75-87. CUP.

Zellers, M. in press. Prosodic variation and segmental reduction and their roles in cuing turn transition in Swedish. Language and Speech.

Visual search strategies and letter position encoding in Russian

Svetlana Alexeeva
Laboratory for Cognitive Studies, St. Petersburg State University, Russia

Abstract

This article reports a visual search experiment involving Cyrillic letters of the Russian alphabet. Results show that (1) the first and last letters of test arrays are detected faster than neighboring letters and the letter search function looked like M-curve; (2) letter quality influences response latencies. The results argued for parallel letter-position encoding in Russian.

Keywords: visual word recognition, visual search task, Russian, Cyrillic script.

Introduction

Previous studies postulate that identification of letters and encoding their positions within words are essential parts of written word recognition (for a review, Acha and Carreiras, 2014). There are two possibilities how we can identify letters within the words: serially (letter-by-letter) or in parallel (so-called whole-word processing) (Coltheart, 2006). One of the methods that help to shed the light on the low-level orthographic processing is visual search task (Hammond and Green, 1982, Pitchford et al., 2008).

In the task, subjects are asked to decide (press the key) whether or not a predefined target character (letter or non-letter symbol) is the part of a subsequently presented stimulus string. The position in which the cued letter appears in the string is manipulated and the response time is measured. Detection latencies for each position of stimulus strings produce a search function that is considered to reflect strategies of letter position encoding (Ktori and Pitchford, 2010).

If the search function reveals a linear component, then it is thought that serial processing comes into play (Pitchford et al., 2008). Usually, it means that the letters appearing at the beginning of the word (e.g., the *s* and *h* in *shark*) are identified faster than ones, appearing at the end (e.g., the *r* and *k* in *shark*). If the end letter is detected faster compared with the preceding letter (e.g., *k* vs. *r* in *shark*), then it is told about a parallel letter identification (Ktori and Pitchford, 2010).

Previous studies on English show that time-position dependency in five letter strings can be described by an upward-sloping M-form curve: the first position is the fastest, but the reaction time in the second position is slower than in the third one and in the fourth position it is slower than in the fifth (Hammond and Green, 1982). The Greek language shows no latency

decrease in the fifth position compared with the fourth one (Ktori and Pitchford, 2008). The result can be explained with the transparency of the Greek orthography: letters in words are processed serially in the languages with transparent orthography whereas in deep orthography languages (like in English) parallel recognition takes place (Pitchford et al., 2008).

Grapheme-phoneme correspondences in the Russian language is quite regular (but the reverse is not true) (Grigorenko, 2013). Therefore, we can predict that the serial processing dominates and time-position function would be rather line-like than an M-like curve in Russian. This paper reports a visual search experiment in Russian which investigated this claim.

Method

Participants

50 volunteers (age range 18-35 years) participated in the study. All of them were naive to the purpose of the experiment.

Design and material

We conducted an experiment with two within-subject variables: *position of the target* (from 1 to 5) and *target-letter identity* (33 Cyrillic letters). For the half of the trials the cued letter appeared within the stimulus string, for the other part, the target letter was absent. As the number of letters in the Russian alphabet is pretty high, we had five experimental lists. In each list, all 33 letters were shown as a target but only in one of 5 possible positions. We randomly assign a position for every target letter for the list 1 (e.g. *а* in position 1, *б* – in 1, *в* – in *5*, and so on). Then we used the Latin-square principle for counterbalance letters across positions in the remaining lists. In each list, a letter was probed eight times.

We used real words as letter strings. Stimuli words were selected for every letter/position pair based on the Frequency dictionary of modern Russian (Lyashevskaya, Sharov, 2009).

Procedure

Subjects were tested individually in a quiet room. The experiment was run using E-prime software. On each trial, a lowercase target letter was presented in the centre of the screen for a duration of 1000 ms, then the blank screen followed. After 500 ms, the blank was replaced by a lowercase test array, which remained in the centre of the screen until the response. Participants were instructed to push the key '/' if they noticed the cued letter in a string of symbols and the key 'z' in the opposite case. They were encouraged to make a decision as quickly and as accurately as possible.

Results and discussion

The letter search function based on mean latencies for correct responses are presented in Figure 1.

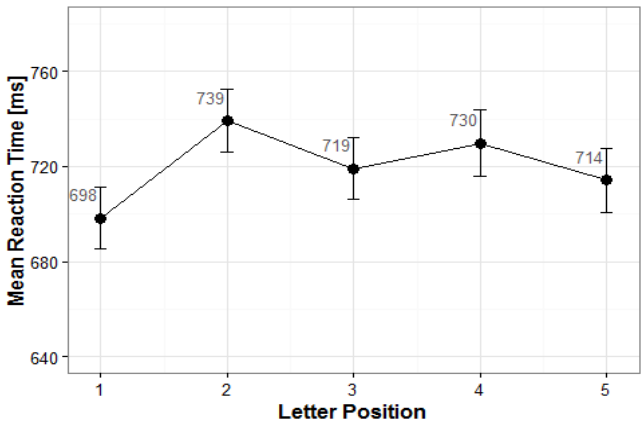


Figure 1. Visual search functions for detection latencies of correct responses across positions (ms).

We performed two linear mixed effects analyses (LMM) of the relationship between detection latencies and letter position. In both analyses, we had intercepts for subjects and items as random effects and letter identity as a fixed effect. Letter identity was coded as a sum contrast (this allowed us to compare detection latencies for each letter against the mean).

In the first analysis, we used letter position as a fixed effect, and it was coded as sliding contrast (this allowed us to compare reaction times in neighboring positions). In the second analysis letter position was entered as a covariate with cubic parameterization (this allowed us to check the significance of linear, quadratic and cubic trends). For all tests, we used the two-tailed criterion ($t \geq 1.96$), corresponding to a 5% error criterion for significance.

The analyses revealed that letters in the first ($t=6.51$) and fifth positions ($t=2.00$) are detected faster than neighboring letters (in the second and fourth positions respectively). There was evidence of a significant quadratic ($t=4.68$) and cubic components ($t=-3.1$), but a linear trend did not reach significance ($t=-0.86$). Contrary to our hypothesis the detection function was M-shaped curve like in English. So we found evidence of parallel letter encoding in the Russian language. We propose two possible explanations for our results: (1) the parallel/serial encoding letter strategy does not depend on transparency of the orthography; (2) the letter-string type biased the results.

We selected real words for the target letter-strings, in previous studies randomly generated nonwords were used (Hammond and Green, 1982, Pitchford et al., 2008).

As for letter quality, we found that ё, о, ж, ш, й, ф, б were recognized significantly faster and letters к, э, и, н, а, ь slower than the mean reaction time across all letters (see Table 1). We think that ascenders/descenders or round elements increase letter identification.

Table 1. Mean reaction times (M) [in ms] and t-test values for positive detections of 33 Russian letters (L.). Effects significant indicated in bold.

L.	M	t	M	L.	t	L.	M	t	L.	M	t	L.	M.	t
а	754	**-2.3**	ж	676	**3.1**	н	754	**-2.4**	ф	700	**2.2**	ы	753	-0.8
б	702	**2.3**	з	722	-0.7	о	662	**5.4**	х	699	1.0	ь	770	**-2.0**
в	717	-0.2	и	743	**-2.5**	п	740	-1.9	ц	749	-0.4	э	752	**-2.9**
г	733	-1.3	й	706	**2.4**	р	713	1.1	ч	720	-1.7	ю	721	0.1
д	703	1.8	к	751	**-3.6**	с	712	0.3	ш	696	**2.4**	я	733	-1.7
е	728	-1.0	л	746	-1.6	т	726	-0.6	щ	702	1.2			
ё	629	**8.3**	м	741	-1.1	у	742	-1.0	ъ	714	-0.5			

Acknowledgements

The project is supported by Russian Science Foundation (#14-18-02135).

References

Acha, J., and Carreiras. M., 2014. Exploring the mental lexicon: A methodological approach to understanding how printed words are represented in our minds. The Mental Lexicon 9. 196–231.

Coltheart. M., 2006. Dual Route and Connectionist Models of Reading: An Overview. London Review of Education 4. 5–17.

Grigorenko. E., 2013. If John were Ivan. would he fail in reading?. in: Handbook of Orthography and Literacy. Routledge. pp. 303–320.

Hammond. E.J. and Green. D.W., 1982. Detecting targets in letter and non-letter arrays. Canadian Journal of Psychology 36. 67–82.

Ktori. M. and Pitchford. N.J., 2008. Effect of orthographic transparency on letter position encoding: A comparison of Greek and English monoscriptal and biscriptal readers. Language and Cognitive Processes 23. 258–281.

Ktori. M. and Pitchford. N.J., 2010. Letter position encoding across deep and transparent orthographies. in: Reading and Dyslexia in Different Orthographies. Psychology Press. pp. 69–86.

Lyashevskaya O.N. and Sharov S.A., 2009. Frequency dictionary of modern Russian. Azbukovnik. Moscow [in Russian].

Pitchford. N.J., Ledgeway. T., Masterson. J., 2008. Effect of orthographic processes on letter position encoding. Journal of Research in Reading 31. 97–116.

Emergence of word prosody in (Seoul) Korean

Angeliki Athanasopoulou, Irene Vogel
Department of Linguistics and Cognitive Science, University of Delaware, USA

Abstract

It has been argued that Korean has recently developed an F0 distinction word-initially partially replacing the VOT distinction of the three stop categories, lax, aspirated, tense. This change has been characterized as tonogenesis, but since the contrast is not on all syllables, it seems to be more consistent with a pitch accent language than a tone language. We investigate the prosodic patterns of trisyllabic words to assess a) whether the VOT-to-F0 change is only word-initial or if it also occurs in other syllables, b) if there is evidence of word level prominence on one syllable supporting a pitch accent interpretation. The data from 10 Korean speakers yield conflicting evidence for both tonal and pitch accent prosodic systems.

Key words: tonogenesis, VOT, pitch accent, Korean

Introduction

Korean is considered a language lacking word prosodic properties (i.e., stress or tone). It has recently been shown that a change is in progress, whereby the three-way stop distinction - lax, aspirated, tense - is being reduced to two (Silva 2006, Wright 2008, Kang 2014).Specifically, word-initially, the VOT contrast between aspirated and lax consonants is being replaced by high and low F0 on the following vowel, respectively. This phenomenon is referred to as tonogenesis; however, for a language to have a fully developed tonal system, we would expect tone contrasts to emerge not only word-initially, but also elsewhere in the word, as for example in Vietnamese (Haudricourt 1954, Thurgood 2002).

In the present study, we examine the acoustic properties of CV syllables with the three consonant types in all positions in 3-syllable words to determine, first, if there is a VOT-to-F0 change in Syllable 1, and then, if there is evidence of such a change beyond the first syllable. Thus, the first prediction is that word initially, the Vowels after a Lax onset (LV) would have lower F0 than those after an Aspirated onset (AV),while the Consonants that are considered Lax (LC) and aspirated (AC) would no longer differ in VOT. The second prediction is that if this process is truly tonogenetic, theVOT-to-F0 change will also be found in syllables 2 and 3.

Method

We collected a corpus of 2700 target vowels (/i, o, a/) in initial, medial and final syllables in real trisyllabic words. The vowels appeared in syllables

with onsets that varied by consonant type, e.g., lax [**pi**gida] ‘draw’, aspirated [**p^{h}i**butε^{h}i] ‘relatives’, tense [**p*i**t*agi] ‘skew’.Two types of simple dialogues were used to elicit the target words in focus and non-focus contexts. The target vowels appeared in the responses, as illustrated in Table 2, where “XXX” is the word containing the relevant vowel.

Table 2. The sentences for the two focus contexts; focus = bold; target = XXX.

Focus:	Chelswu-ka ohu-ey "**XXX**" -lako ha-yss-e. Chelswu afternoon **XXX** said ‘Chelswu said “**XXX**” in the afternoon.’
Non-Focus:	Ani. Chelswu-nun ohwu-ey "XXX" -lako **malha-yss-ci** ku-kes-ul cek-ci-nun ahn-ass-e. No. Chelswu afternoon XXX**said** it write not did. ‘No. Chelswu **said** “XXX” in the afternoon, she didn’t write it.’

Data from 10 native Seoul Korean speakers were collected in Seoul by a native speaker. Participants were recorded individually producing the dialogues presented through PowerPoint presentation.

For each target vowel, duration, intensity, F0 (mean and contour) and vowel centralization were measured and Z-normalized for vowel and speaker intrinsic differences. The data were analysed statistically with Binary Logistic Regression(see Vogel, Athanasopoulou and Pincus 2015). We also measured the VOT of the onset stops for two speakers to verify previous claims that the VOT distinction is being lost.

Results

Our findings corroborate the results of previous studies showing that word-initially, F0 has replaced the VOT distinction between aspirated and lax consonants. In both focus conditions, LC and AC had similar VOTs (~50ms), but LV had a lower F0 than AV. Moreover, the tense stop (TC), as expected, had the shortest VOT (20ms) and the vowel after the tense onset (TV) had a mid F0, roughly between the F0 of LV and AV. In addition to the mean F0, it is interesting to note that while the F0 contour of LV and AVis relatively flat, the F0 of TV has a rising contour, a difference probably due to the longer duration of TV(67ms vs. 49-55ms). The F0 and duration properties are presented in Figure 1.

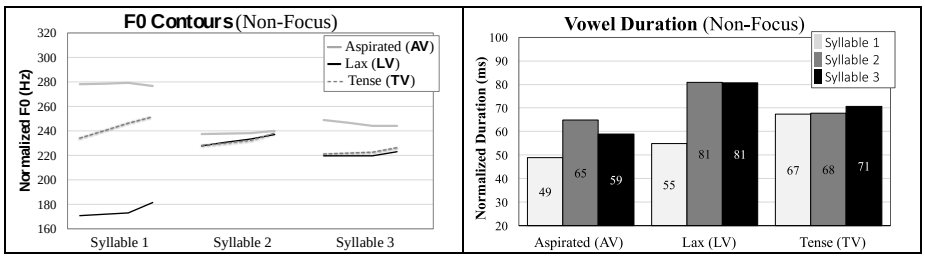


Figure 1. F0 contours and Duration for each onset type and syllable position.

In contrast, as can also be seen by examining the contours in Figure 1, we found essentially no F0 differences between the vowels following the three stop types in Syllable 2, where the VOT distinction between lax and aspirated stops is maintained. This was the case for both focus conditions. In Syllable 3, on the other hand, even though the VOT distinction is also maintained, there were differences in the F0 of the vowels, but smaller than those in Syllable 1. Specifically, the AV had higher F0 (by ~ 20Hz) than LV or TV. In Syllable 3, LV does not have lower F0 than TV and it is much higher than in Syllable 1. In addition, the F0 of AV, although higher than the other two, it is lower than the F0 of the AV in syllable 1. Thus, the slightly higher F0 that we see in the AV of syllable 3 appears to be due to the effect of aspiration on the F0 of the following vowel(e.g., Hombert 1975), and not the replacement of VOT with F0. We can also see this in Syllable 2 but the difference is even smaller. While this may be evidence of the beginning of a tonal difference in syllable positions beyond the first one, such an interpretation requires caution due to the small differences.

As seen above in Figure 1, we additionally found only minimal acoustic evidence of focus, with the strongest distinction between the non-focus vs. focus contexts appearing in Syllable 3. The focused vowels are slightly longer than those without focus (by 10-20ms) in Syllable 3, and F0 is either lower than in the focused Syllable 2 or with a falling contour. Given the combination of somewhat increased duration and the F0 difference on Syllable 3, however, the pattern of these properties appears to be more an indication of a final boundary marker (IP boundary tone in ToBI terms (Jun 2005)) as opposed to evidence of a word prosodic property in that position.

Discussion and Conclusions

Our findings are consistent with previous studies of Korean with regard to initial F0 patterns. That is, there is clearly some development of an F0 contrast word-initially, and we may thus conclude that there is indeed evidence that the language is undergoing a change from one that lacks word prosodic phenomena to one that has such a phenomenon. The limitation of

this phenomenon to word-initial position, however, suggests that "tonogenesis" is not the appropriate characterization of the change at this point, if a tonal language is one that exhibits tone contrasts in different positions throughout a word. Instead, what may be emerging is a restrictive type of lexical stress system with prominence predictably on the first syllable (i.e., asin Hungarian, where the primary cue is also F0(Vogel, Athanasopoulou and Pincus 2015)). Nevertheless, the Korean system is also not yet a full-fledged stress system since we found no evidence of enhancement of the prosodic properties of the first (or any other) syllable under focus, as would be expected on the stressed syllable of a word in a stress language. Finally, it is possible that what is emerging in Korean is a "so-called" pitch-accent system, as in Japanese, where not all words need to bear an accent. This would be consistent with the fact that while the innovative F0 property distinguishes High vs. Low in place of aspirated vs. lax, there remains a non-contrasting pattern in syllables beginning with a tense consonant. Moreover, since the tonal property is not observed with other onsets and on other syllables, there are numerous words that could be considered accentless, as in Japanese.

References

Haudricourt, André-Georges. 1954. "De l'origine des tons en vietnamien." *Journal Asiatique* 242: 69-82.

Hombert, Jean-Marie. 1975. Towards a Theory of Tonogenesis: an Empirical, Physiologically and Perceptually based Account of the Development of Tonal Contrasts in Language. Doctoral Dissertation: University of California, Berkeley.

Jun, Sun-Ah. 2005. "Korean intonational phonology and prosodic transcription." In *Prosodic Typology: The Phonology of Intonation and Phrasing*, by Sun-Ah Jun, 201-229. Oxford University Press.

Kang, Yoonjung. 2014. "Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: a corpus study." *Journal of Phonetics* 45: 76-90.

Silva, David. 2006. "Acoustic evidence for the emergence of tonal contrast in Contemporary Korean." *Phonology* 23: 287-308.

Thurgood, Graham. 2002. "Vietnamese and tonogenesis: revising the model and the analysis." *Diacronica* 19 (2): 333-363.

Vogel, Irene, Angeliki Athanasopoulou, and Nadya Pincus. 2015. "Acoustic properties of prominence in Hungarian and the Functional Load Hypothesis." In *Approaches to Hungarian 14*, by Katalin Kiss, Balázs Surányi and Éva Dékány, 267-292. Amsterdam: John Benjamins.

Wright, Jonathan. 2008. *The phonetic contrast of Korean obstruents.* Doctoral dissertation: University of Pennsylvania.

Voice Activity Detector (VAD) based on long-term phonetic features

Andrey Barabanov[1], Daniil Kocharov[2], Sergey Salishev[3], Pavel Skrelin[2], Mikhail Moiseev[4]

[1]Department of Cybernetics, Saint-Petersburg State University, Russia
[2]Department of Phonetics, Saint-Petersburg State University, Russia
[3]Department of Informatics, Saint-Petersburg State University, Russia
[4]Intel Labs, Intel Corporation, USA

Abstract

We propose a VAD using long-term phonetically motivated features with auditory masking, and pre-trained decision tree based classifier, which allows capturing syllable level structure of speech and discriminating it from common noise types. algorithm demonstrates on test dataset almost 100% acceptance of clear voice for English, Chinese, Russian, and Polish speech and 100% rejection of stationary noises independently of loudness with low computational cost.

Key words: Voice Activity Detector, classification, decision tree ensemble, auditory masking, phonetic features

Introduction

The problem of low complexity accurate VAD is important for many applications in Consumer Electronics, Wearables, Smart Home and other areas, where VAD serves as a low-power gatekeeper for a more complex and energy consuming Automatic Speech Recognition (ASR) system.

Our VAD approach is based on the detection of signal segments with formants in the spectrum. The method cuts off all voiceless consonants and the majority of voiced ones. This should be compensated by considering as a speech the sound signal that includes some unvoiced segments preceding and following vocalized sequence. The duration of such segments is language-dependent. On one hand, it should be long enough to contain consonant clusters. On the other hand, it should be shorter than inter-phrase pauses. Different languages have various consonant-to-vowel ratios and the maximum length of consonant clusters. Thus the length of consonant segment has to vary from language to language. The pause length is less language-dependent and more speaker-dependent. From this point of view the duration of consonant segment should be about 200 – 250 ms.

We propose to use long-term 200 ms speech statistics in combination with pre-trained complex non-linear classifier, which allows capturing syllable level structure of speech and distinguish it from common noises. Proposed algorithm substantially outperforms competitive solutions in

various non-stationary noises and demonstrates on test dataset almost 100% acceptance of clear voice and 100% rejection of stationary noises at the cost of higher latency. The algorithm reuses short-term FFT analysis (STFFT) in ASR front-end; therefore, the complexity increase to MFCC ASR front-end is small.

VAD Algorithm Description

The algorithm consists of feature extraction, feature space dimensionality reduction and two-level classifier (phoneme and syllable levels). It uses Mel band spectral envelope and Mel band peak factor as features.

Spectral envelope is a standard ASR feature usually manifesting as MFCC or Linear Prediction Coefficients (LPC). According to the acoustic theory of speech production by (Fant, 1962), the harmonics of fundamental frequency contain most speech energy, which makes them robust to noise due to high SNR, it distinguishes speech from most types of noise. To improve noise robustness, tonal and temporal auditory masking are applied to spectral envelope (Fastl, Zwicker, 2006). Features are classified by a soft classifier using an ensemble of deep decision trees (Zhou, 2012). For classifier training we used database of continuous English speech TIMIT, noise databases Aurora 2, ETSI and SISEC10.

Comparison

For comparison, we used two state of the art VADs: Google WebRTC VAD and Nuance SREC VAD. For testing, we used sound files completely unused in training. We separately performed False Accept testing on noise database and False Reject testing on speech database with various SNRs. For false accept test, we used DEMAND database containing background noises for 18 environments (Table 1). We conclude that new VAD outperforms competitors. We tested false accept rate on 3 tracks of Rock, Pop, and Classic music genres not used in training. We conclude that new algorithm substantially outperforms competitors, still false accept on music is about 20%.

For false reject testing, we used speech database of 5 min recordings in four languages (English, Chinese, Russian, Polish – in accordance with their consonant coefficient), male and female speakers for each language with manual VAD markup. Noise was synthetically added to with various SNRs calculated as total speech to total noise power after high-pass filter with 100 Hz cutoff. We conclude that new VAD is highly accurate and language and speaker insensitive for high SNR (up to 10 dB). We tested with various noises (Table 2).

Table 1. False accept rate comparison in % for different environmental noises and music.

Noise	SREC	WebRTC	Proposed
dkitchen	11.7	12.4	10.9
dliving	30.7	90.3	4.5
dwashing	20.9	84.5	5
nfield	0	74.1	0
npark	48.1	30	4.6
nriver	0.5	15.3	0
ohallway	23.4	15.4	2.7
omeeting	71.7	67.8	78.3
ooffice	28.6	20.8	0.3
pcafeter	77.4	80.9	38.3
presto	42.8	83.2	1.6
pstation	1	100	0
scafe	75.2	89.7	22.1
spsquare	31.4	73.3	11.8
straffic	10.6	82.9	0
tbus	67.1	77.2	30.4
tcar	1.2	95.5	0
tmetro	27.5	89.1	14.3
rock	91	97.6	11.1
pop	88.8	82.4	19.7
classic	91	90.7	18.1

New VAD algorithm is highly accurate in car noise with FAR about 1% at SNR 0 dB. For non-stationary noises, it demonstrates similar performance up to SNR 10 dB and degrades for lower SNR on babble noise. This correlates with subjective intelligibility of the speech.

Conclusion

The proposed algorithm substantially outperforms competitive solutions in various environments and demonstrates on test dataset almost 100% acceptance of clear voice and 100% rejection of stationary noises with 15% complexity increase compared to MFCC based ASR front-end. The algorithm has a latency of 200 ms, which is not acceptable for some scenarios such as VoIP. The algorithm in some cases falsely accepts some noises as voice: clatter of dishes; sound of flowing water; resonant strokes; tonal beeps; babble noise; bird songs. The algorithm falsely rejects speech in the presence of high amplitude non-stationary noise especially babble noise.

Table 2. Proposed VAD false reject rate in % for different environmental noises and SNR.

		tcar			nriver			presto		
language	gender	20dB	10dB	0dB	20dB	10dB	0dB	20dB	10dB	0dB
English	f	0	0	0	0	1.6	0.7	0	25.6	54.7
	m	0.2	0.2	0.2	0.2	3.7	3.3	0.9	23.1	37.2
Chinese	f	0.1	0	0	0	1.4	5.1	0	27.6	40.8
	m	0	0	0	0	0.4	0.1	0	8.8	51.7
Russian	f	0.1	0.3	0	0.2	4	6.1	0.5	31.1	63.6
	m	0.2	0.2	0.3	0.2	3.1	3.1	0.1	20.4	65.8
Polish	f	0.5	0.7	0.5	0.6	5.1	5.8	1.6	41.1	53
	m	0.2	0.3	0.4	0.3	7.6	11.3	1.3	42.3	57.4
Mean		0.2	0.2	0.2	0.2	3.4	4.4	0.6	27.5	53

Table 3. False reject rate comparison in % for different environmental noises and SNR.

Noise	VAD	inf	20 dB	15 dB	10 dB	5 dB	0 dB
TCAR	SREC	0.9	1.3	1.6	1.9	2.2	2.5
	WebRTC	1.1	1.3	1.3	1.2	0.8	0.4
	Proposed	0.1	0.2	0.1	0.2	0.3	0.6
NRIVER	SREC	0.9	3.5	5.8	11.3	23.4	54.2
	WebRTC	1.1	3.1	4.7	7.1	12.0	22.8
	Proposed	0.1	0.2	0.8	3.4	9.8	27.5
PRESTO	SREC	0.9	2.0	2.6	4.3	9.2	21.4
	WebRTC	1.1	3.3	4.9	5.9	4.8	1.8
	Proposed	0.1	0.2	0.9	4.4	19.8	53.0

References

Fant, G. 1960. Acoustic theory of speech production: With calculations based on x-ray studies of Russian articulations.

Fastl, H., Zwicker, E. 2006. Psychoacoustics: facts and models, vol. 22. Springer Science & Business Media.

Zhou, Z.H. 2012. Ensemble methods: foundations and algorithms. CRC Press.

The identification of two Algerian Arabic dialects by prosodic focus

Ismaël Benali
CLILIAC-ARP, Université Paris Diderot, France

Abstract

The purpose of this research is to show that it is easier to identify the prosody of Algiers and Oran dialects when a focus is produced. For this study, we compared prosodic features associated with different types of focus: broad focus, emphatic narrow focus, contrastive narrow focus and interrogative focus. It appears from the acoustical analysis that recurrences of prosodic patterns that differentiate the two dialects were observed in narrow and interrogative focus. The analysis of the interaction between the identification of the two dialects and the four types of focus showed that Algiers and Oran speakers are better identified when their utterances are produced with narrow focus when it is placed at the edge of an intonation phrase and interrogative focus.

Key words: dialectal variations, Algerian Arabic, intonation, focus

Introduction

Several studies have shown that dialectal varieties can be differentiated only on the basis of prosody. The suprasegmental parameters such as speech rate, F0 register and range, F0 excursion and F0 alignment are sufficient to distinguish and identify dialects.

The Algiers and Oran varieties are two urban dialects of Algeria. They are characterized by regional accents marked segmentally and prosodically.

In a previous studies (Benali, 2004), it appeared that Algiers speakers produced more melodic variations than Oran speakers who tended to produce more syllabic lengthening. We found also that intonation patterns which characterize Algiers and Oran varieties are marked more clearly when the speaker spoke with emphasis and implication. To study this phenomenon, we compared prosodic features (mainly F0 movements) associated with different types of focus: first, broad focus (emphasis on the whole or a part of utterance); then, emphatic narrow focus (strong emphasis on a specific item of an utterance); then, contrastive narrow focus (emphasis on a contrasting item in an utterance) and finally, interrogative focus (emphasis of a linguistic element on which the question bears).

In most languages, narrow focus is marked by F0 rise and often accompanied by an increase of duration and intensity (Hirst and Di Cristo, 1998). In a comparison of the acoustic realizations of contrastive focus carried on three Arabic dialects: Moroccan Arabic, Kuwaiti Arabic and

Yemeni Arabic, Yeou and al (Yeou et al., 2007) have shown that these dialects share the same strategy in the realization of contrastive focus consisting in a rising falling movement. This melodic contour was more locally defined in Yemeni and Kuwaiti Arabic while it may span the entire focused word in Moroccan Arabic. Moroccan Arabic is distinguished by a significant effect of the syllabic structure on F0 peak alignment: It occurs within the accented syllable when it is closed and outside when it is open. In Kuwaiti and Yemeni Arabic, this peak occurs within but near the end of the accented vowel either in open or closed syllable. In Egyptian Arabic, S. Hellmuth (Hellmuth, 2011) showed an increase of F0 in focus and a compression of it in the following words. Also in Tunisian Arabic (Bouchhioua, 2009), focus affects positively the duration of both the stressed syllable and the unstressed syllable. Stressed final syllables are more lengthened and the F0 and intensity of the stressed syllable increase in effect of focus.

Methodology

20 Algiers speakers (15 men and 5 women) and 20 Oran speakers (10 men and 10 women) were recorded in their respective cities. There is spontaneous and read speech. Focus was either naturally produced or provoked.

In a first experiment we isolated the prosodic information, using a method of delexicalization by filtering speech frequency above 400 Hz.

In a second experiment we manipulated non filtered speech: we transposed F0 variations and vowels durations of the read statements of one dialect onto the other and vice versa. We submitted these two types of stimuli to 30 listeners (neither from Algiers nor from Oran) who had to identify the dialects.

The analysis and the acoustic manipulations were carried either on the speech analysis/resynthesis program ‘WinPitch’ (Martin, 2000), or on ‘Praat’.

Results

Acoustic analysis results

Spontaneous speech

Narrow focus is marked prosodically in both dialects. Algiers variety is characterized by rising falling contours and especially by a final melodic drop. Oran dialect is characterized by a lengthening of stressed syllables with lowered contours which are generally flat. F0 peak alignment is usually on pre-nuclear syllable in Algiers dialect.

Read speech

The statement used in read speech is: "Ali (he) is sick." [ʕali rah mri:dˁ]; the speakers were asked to vary the type of focus.

It appears from the acoustical analysis that recurrences of prosodic patterns that differentiate the two dialects were observed in only two types of focus: the emphatic narrow focus when it is at the edge of an intonation phrase and interrogative focus. Emphatic narrow focus is produced in the Algiers dialect by a high and falling contour on the last stressed syllable. In the Oran dialect, this focus is realized either with a contour which is flat or slightly rising on the last stressed syllable (figure 1). In both dialects the stressed syllable is lengthened.

In the interrogative focus Algiers speakers produce an amplified rising-falling contour while Oran speakers produce on the last syllable a rising contour preceded by a falling one (figure 2). The realization of contrastive focus varied across speakers of the same dialect. Broad focus was realized with similar intonation patterns for both dialects.

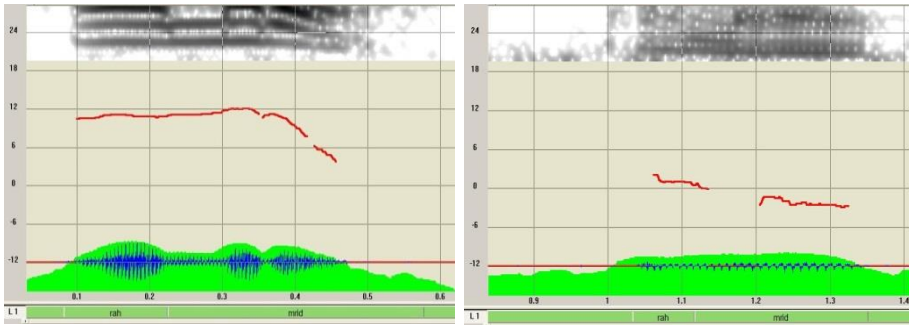


Figure 1. Emphatic narrow focus produced by Algiers (left) and Oran speakers (right).

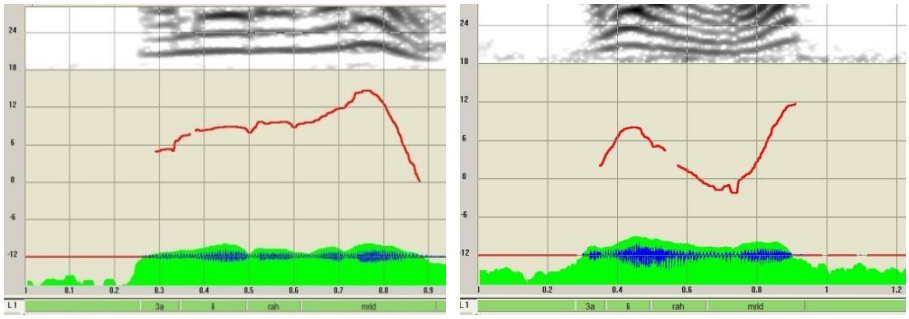


Figure 2. Interrogative focus produced by Algiers (left) and Oran speakers (right).

Perception test results

The interaction between identification of Algiers and Oran dialects and types of focus is significant: p <0.0001 (figure 3). Algiers speakers were clearly identified in all types of focus. Algiers and Oran speakers were better identified in interrogative focus (80%). Only Algiers speakers were better identified in emphatic narrow focus (90%) while 53% of Oran speakers were identified.

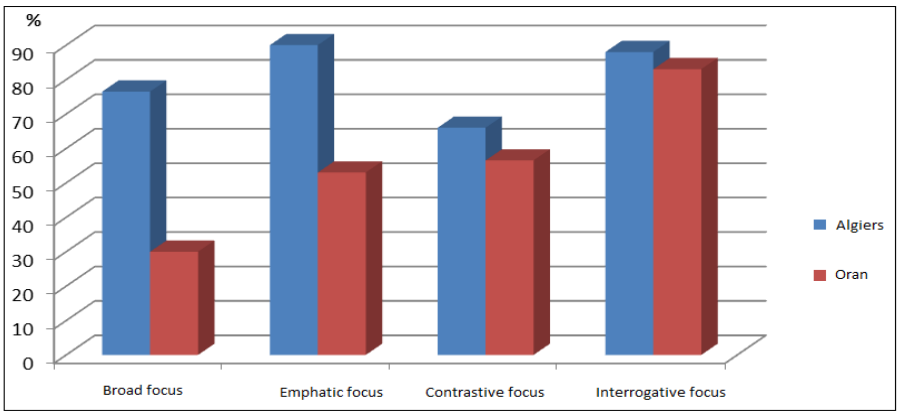


Figure 3 Interaction of identification rate and type of focus

Conclusion

The narrow emphatic focus and the interrogative focus distinguish Algiers and Oran dialects and they are better identified in these types of focus.

References

Benali, I. 2004. Le rôle de la prosodie dans l'identification de deux parlers *algériens: l'algérois et l'oranais. Workshop MIDL.*

Bouchhioua, N. 2009. Stress and Accent in Tunisian Arabic. First International Conference on Intonational Variation in Arabic, 28-29.

Helimuth, S. 2011. Acoustic cues to focus and givenness in Egyptian Arabic. *Instrumental Studies in Arabic Phonetics*, 319**,** 301.

Hirst, D., Di Cristo, A. 1998. *Intonation systems: a survey of twenty languages*, Cambridge University Press.

Martin, Ph. WinPitch 2000: a tool for experimental phonology and intonation research. Proceedings of the Prosody 2000 Workshop, 2000.

Yeou, M., Embarki, M., Al-Maqtari, S. 2007. Contrastive focus and F0 patterns in three Arabic dialects. *Nouveaux cahiers de linguistique française***,** 317.

Intonation and polar questions in Greek revisited

Antonis Botinis[1], Anthi Chaida[1,2], Olga Nikolaenkova[3], Elina Nirgianaki[1,2]
[1]Lab of Phonetics & Computational Linguistics, University of Athens, Greece
[2]Faculty of Primary Education, University of Athens, Greece
[3]Department of General Linguistics, Saint Petersburg State University, Russia

Abstract

This is a production study of intonation and polar questions in Greek. The results indicate that there is a fairly invariable rising-falling tonal structure at the right edge of polar questions. However, the alignment of both tonal rising and tonal peak depend on the position of focus as well as lexical stress production. Thus, in the context of initial and medial focus productions, the tonal rising is aligned with the onset of the final stressed syllable whereas, in the context of final focus production, the tonal rising is aligned with the onset of the last syllable regardless of the position of lexical stress. On the other hand, the tonal peak is aligned with the post-stressed syllable in the context of initial and medial focus productions whereas, in the context of final focus production, the tonal peak is aligned with the nucleus of the last syllable. However, the earlier the lexical stress production, the earlier the tonal rising as well as the tonal peak in all focus contexts.

Key words: polar questions, intonation, Greek, focus, tonal associations

Introduction

Sentence type intonation in Greek shows several basic characteristics. Thus, statements and polar questions, much like in other languages, such as Italian and Russian, hardly have any other correlates except for intonation. Lexical stress production in statements may be associated with a tonal rise in prefocus position whereas, in focus position, a local tonal expansion is followed by a postfocus tonal flattening (e.g. Botinis 1989). In polar questions, on the other hand, there is no tonal expansion of focus application but a rising-falling tonal characteristic at the right edge (e.g. Chaida 2010).

In a recent study (Chaida, Sotiriou, Kontostavlaki 2016, this volume), the rising-falling tonal characteristic of polar questions at the right edge is fairly evident, but a tonal expansion of focus application is also evident, leading us to the question: what are the basic characteristics of polar question intonation in Greek? In addition, particular questions are addressed with reference to the associations of variable lexical stress as well as focus applications. In this study, we have developed a new question-question methodology, according to which a first wh-question elicits a polar (yes-no) question. We think that this methodology is straightforward and may be applied to other languages in principle, especially to languages with no other morphological and/or syntactic means to produce polar questions but intonation.

Experimental methodology

The speech material of the present study consists of two serries of a question-question methodological paradigm sequence. The first question is an elicitation Wh-question, in order to assign a specific focus at the second question, i.e. the target question, which is a polar question with variable final lexical stress assignment at one of last three syllables (Table 1). In the first series, the target question is a full question, whereas, in the second series, the target question is an eliptical one, corresponding to the final prosodic word of full question (Figures 1-4).

Five female speakers, 20-40 years old, with standard Athenian pronunciation, produced the speech material at a normal tempo in a sound-treated room at Athens University Phonetics Laboratory. The speakers red the speech material from a piece of paper, first the elicitation question followed by the target question.

The speech material analysis was carried out with Praat, with several annotation tiers. In this report, we have concentrated on one tier, i.e. the stressed vs. unstressed prosodic distinction, and the speech material has been normalized with Prosody Pro tool (Xu 2013).

Table 1. Elicitation questions (left) and target questions (right) with variable final lexical stress as well as neutral and variable focus assignments (bold letters).

1.1	[pça ðuˈlevi sti ˈmadova]? ‘Who works in Mantova’?	1.1	[i ˈ**nana** ðuˈlevi sti ˈmadova]? ‘**Nana** works in Mantova’?
1.2	[pça ðuˈlevi sto miˈlano]? ‘Who works in Milano’?	1.2	[i ˈ**nana** ðuˈlevi sto miˈlano]? ‘**Nana** works in Milano’?
1.3	[pça ðuˈlevi sto banaˈma]? ‘Who works in Panama’?	1.3	[i ˈ**nana** ðuˈlevi sto banaˈma]? ‘**Nana** works in Panama’?
2.1	ti ˈkani i ˈnana sti ˈmadova? What does Nana in Mantova?	2.1	[i ˈnana **ðu**ˈ**levi** sti ˈmadova]? Nana **works** in Mantova?
2.2	ti ˈkani i ˈnana sto miˈlano? What does Nana in Milano?	2.2	[i ˈnana **ðu**ˈ**levi** sto miˈlano]? Nana **works** in Milano?
2.3	ti ˈkani i ˈnana sto banaˈma? What does Nana in Panama?	2.3	[i ˈnana **ðu**ˈ**levi** sto banaˈma]? Nana **works** in Panama?
3	pu ðuˈlevi i ˈnana? Where works Nana?	3.1	[i ˈnana ðuˈlevi **sti** ˈ**madova]**? Nana works **in Mantova**?
		3.2	[i ˈnana ðuˈlevi **sto mi**ˈ**lano]**? Nana works **in Milano**?
		3.3	[i ˈnana ðuˈlevi **sto bana**ˈ**ma]**? Nana works **in Panama**?

Results and discussion

In accordance with the aim of the study and the questions addressed in the introduction, the results are presented in Figures 1-4.

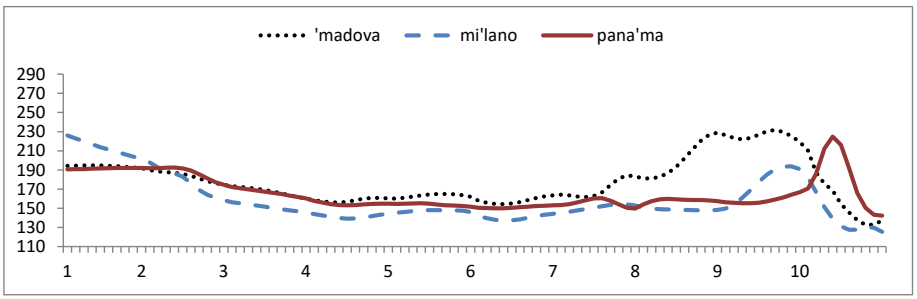


Figure 1. Intonation of polar questions in Greek as a function of initial focus and variable final lexical stress assignments. Numbers on vertical and horizontal axis indicate values in Hz and syllable boundaries, respectively.

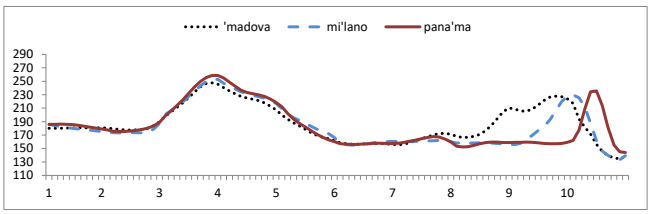


Figure 2. Intonation of polar questions in Greek as a function of medial focus and variable final lexical stress assignments. Numbers on vertical and horizontal axis indicate values in Hz and syllable boundaries, respectively.

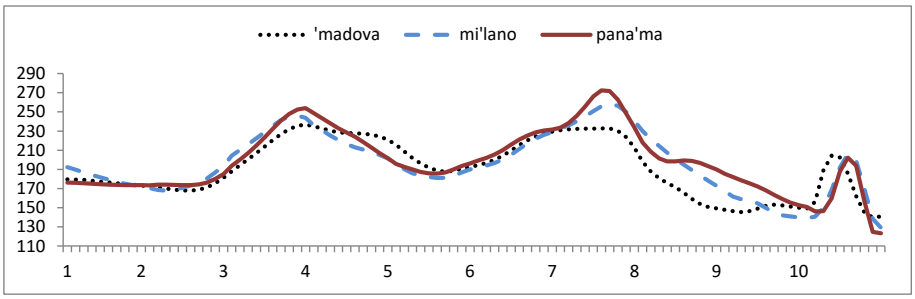


Figure 3. Intonation of polar questions in Greek as a function of final focus and variable final lexical stress assignments. Numbers on vertical and horizontal axis indicate values in Hz and syllable boundaries, respectively.

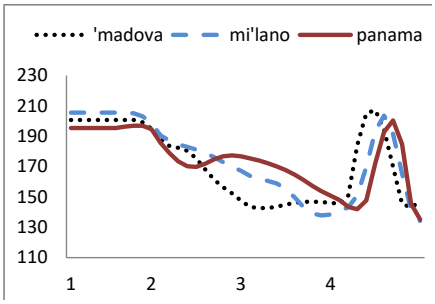


Figure 4. Intonation of one prosodic word polar questions in Greek as a function of variable lexical stress assignments. Numbers on vertical and horizontal axis indicate values in Hz and syllable boundaries, respectively.

Initial focus productions (Figure 1) show only one right edge tonal peak, which is aligned with the nucleus of the last lexical stress whereas the tonal rise is aligned with the onset of respective syllables. Medial focus productions (Figure 2), in addition to a tonal peak at the right edge, show another peak aligned with the post-stressed syllable of the first word. Final focus productions (Figure 3), show three tonal peaks, the first two aligned with the post-stressed syllables of respective words whereas the third one is aligned with the last syllable of the sentence. However, the earlier the lexical stress the earlier the alignments of both tonal rise and tonal peak, which is also evident in one prosodic word polar questions (Figure 4).

The rising-falling tonal production at the right edge of polar questions has also been reported in other languages, including Greek (Grice, Ladd, Arvaniti, 2000) and Brazilian Portuguese (Castelo, Frota, forthc.). These studies are in the framework of Autosegmental-Metrical phonology and argue for a sequence of three tones, i.e. LHL, with reference to the rising-falling tonal production at the right edge of polar questions. However, in accordance with our results, this is too broad an approach and further research of polar questions and sentence types in general is in place.

Acknowledgements

Thanks to Athens University S.A.R.G for economic support.

References

Botinis, A. 1989. Stress and Prosodic Structure in Greek. Lund University Press.

Castelo, J., Frota, S. Forthc. The yes-no question contour in Brazilian Portuguese.

Chaida, A. 2010. Production and Perception of Intonation and Sentence Types in Greek. PhD Thesis, University of Athens.

Chaida, A., Sotiriou, A., Kontostavlaki, A. 2016. Intonation and polar questions in Greek. (this volume).

Grice, M, Ladd, R., Arvaniti, A. 2000. On the place of phrase accents in intonational phonology. Phonology 17, 143-185.

Xu, Y. 2013. ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. Proc. TRASP 2013, 7-10. Aix-en-Provence, France.

The imprint of disposition in social interaction

Mark Campana
Dept of English and American Studies, Kobe City University, Japan

Abstract

This study considers how listeners perceive and interpret the disposition of others through non-linguistic vocal cues. Changes in F0 and pitch span (measured against a 'running' mean of the previous 15 seconds), constellations of sequential tones, and emergent speech rhythms index recognizable states of positive/negative valency, desire, knowledge and/or processing, which together constitute emotional display (these same states correlate with mental predicates in the composition of emotion words. Excerpts of natural conversation were converted to 'iterant speech', i.e. speech devoid of lexical content. Listeners were invited to identify speaker disposition, and their ability to do so was remarkably accurate. The results lend support to a theory of vocal affect based on sound-types, rather than sounds.

Keywords: disposition; emotional display; mental predicates; iterant speech

Introduction

This paper addresses the imprint of disposition in social interaction. Disposition is taken to mean something like a frame-of-mind which both governs the behavior of the person who has it, and is evaluated by those who witness it. One can *have* a certain disposition (where *certain* is replaced by an adjective) or *be of* a certain disposition (idem). The question we address here is how the listener comes to realize that a speaker has (or is of) a certain disposition based on tone-of-voice. In principle, the answer will be the same as how the listener grasps the shifting mental states of the speaker in the course of interaction, but there are some differences.

To illustrate the phenomena, a person can have e.g. a *sunny* disposition or a *surly* one, be of a *grumpy* or a *fearful* disposition. Other plausible/attested collocations are *thoughtful, cheerful, kind, easy-going*—with positive valency—or *angry, taciturn*—with negative. One can also be *predisposed* towards a proposition with positive/negative content—e.g. judging someone *harshly* or with *kindness*.

Consider next the concept of an 'imprint', which is different from an impression. An impression of e.g. a person's character can be formed after a single encounter. An imprint typically results from several encounters, i.e. it includes memories of previous ones. In it, impressions are weighed and integrated in a more substantive schema. Basically, it takes longer to make an imprint of disposition, but in theory it can be appraised after a single encounter.

Social interaction is something that everybody understands. The focus here is on talk-in-interaction, considering utterances that are highly affective. These are called stances, or spoken actions in which a speaker displays his/her thoughts or feelings about some object, and communicates them to the listener with inevitable social consequences. We will concentrate on the prosodic features of stance utterances, paying special attention to pitches and pitch combinations (in sequence or together) rhythms, tempos and timbres. Different types of vocal are suitable conveyors of mental states, which in turn constitute emotions and dispositions.

Theory

What is tone-of-voice? First, it is 'about' tones (or pitches) but the array of sounds at the disposal of the speaker has a temporal aspect, an organizational one, and then there is the issue of voice quality. At the same time, we understand tone-of-voice to be the audible analogue of emotion. The litany of speech sounds in social interaction is essentially infinite. The oral cavity alone is designed such that even minute flexions of a single muscle (or muscle-group) can produce a complex, distinctive sound that is potentially 'meaningful' for the assessment of the speaker's mental state. In terms of efficiency, it would make sense for such sounds to be organized into sound-types for the purpose of transmitting and understanding vocalized meaning. Categorization is a cognitive skill at which humans (and some other species) have proved to be adept. In this paper, we test a specific theory of sound-types that index mental 'sub-states' of positive/negative valency, desire, knowledge and processing, which together constitute an emotional display (Wierzbicka 1999). Inasmuch as changes in perceived disposition correlate with controlled modulation of sound-type parameters, the theory can be verified. What then is emotion? This is not a simple question either, but we may start by following Wierzbicka (1999) and others in assuming that most 'emotions' include a 'thinking' part, as well as a 'feeling' one. In her model of semantics (NSM), words like *disappointment, afraid, happiness,* etc. are cast as 'cognitive scenarios', short narratives made up of simple words and propositions. Among the set of 'mental predicates' which play a key role in every scenario are *want, know, feel* and *think*. Together with *good, bad* and and *not* (also from the metalanguage) we derive the following mental states, any combination of which can be heard in the expression of emotion itself (abbreviated as WXYZ):

(1) Mental states (adapted from Weirzbicka 1999)

W wanting/not wanting (takes an object)

X knowing/not knowing (takes an object)

Y feeling good/bad (about something)

Z thinking (no negative counterpart)

The next step is to match the types of sounds that make up tone-of-voice with WXYZ. This is only an approximation, whereby a given sound type is just a 'leading indicator' of a mental state, not necessarily the only one. Combinations of sounds (as well as the meaning of words) can also index a mental state. That said, we propose that voice qualities—broadly defined—are used to signal states of *wanting* or *not wanting*. Intensity of F0 (volume) counts as a voice quality, along with upper partials (timbres) and non-standard vocal gestures, such as 'clipped' endings, etc.

Short tunes or melodies—sequences of tones—are used to signal *knowing* or *not knowing. Aizuchi* (backchannels) are typical: even when the 'tune' appears to have a single tone, it is juxtaposed against that of previous speech. Consider what it sounds like to say "I don't know" in your language. Echoes of the same can be heard in longer stretches of speech as well.

Next, consider the mental states of *feeling good* or *feeling bad.* These correspond most closely to valency, as it is known in emotion research. Pitches and pitch combinations are primarily responsible for signaling these states. Cook (2003) develops the idea that valency follows from three-tone chordal structure, and there is no reason to dispute this. Emotional displays do unfold quickly, so it is likely that even tones in sequence are perceived as simultaneous, i.e. in the 'psychological now'.

Finally, we propose that rhythms and timing units in general (tempos, pauses, hesitations etc.) accurately reflect the mental activity of *thinking.* It is not enough to simply demonstrate that thinking is taking place; the style presentation and grouping of syllables is important too, influenced in part by the choice of words.

To summarize, the mental predicates that serve to characterize emotion words in Wierzbicka's semantic system correspond to real mental states that occur in the display of emotion. In theory, such states could be indicated by facial expression, body movements (including gesture), or simply words. Tone-of-voice is just another means of expression, where each mental state/activity is indicated by a sound type, shown below (wxyz):

(2) Mental states and leading sonic indicators (sound types)

W	wanting/not wanting	w	voice qualities
X	knowing/not knowing	x	short tunes/melodies
Y	feeling good/bad	y	pitches/pitch combinations
Z	thinking	z	timing units (rhythm, tempo)

Given that at least one display of emotion is necessary to appraise a speaker's disposition, it follows that the same elements listed here will contribute to it. In the following section, we outline how such events can be discerned in a controlled experiment.

Data, methods

In the course of daily interaction, listeners can appraise the disposition of a speaker based on tone-of-voice. Can naïve subjects reach similar conclusions in a clinical experiment? Possibly, but not necessarily: every action depends on individual experience, social consequences, and other factors. It isn't fruitful to devise an experiment along these lines. Nevertheless, listeners may be able to recognize repeated patterns in a speaker's voice on different occasions, and trained ones can identify and describe them. Gathering such data from a longitudinal study is optimal, but impractical. In the tasks reported on here, listeners were presented with stance utterances from speakers over a range of topics, and asked to appraise their disposition. In order to control for word meaning though, the stance utterances were converted to 'iterant' form, leaving only prosody.

In its core meaning, a stance is a physical event whereby the stance-taker assumes a bodily position that signals a clear intention to the audience. One can easily imagine how something like 'defiance' is acted out by assuming a defensive posture. In current sociolinguistics, the concept of stance has been extended to talk-in-interaction. Many researchers refer to the seminal work of DuBois (2007), who proposes that every stance has a subjective dimension (i.e. about the speaker), an objective one concerning the person or thing being evaluated, and an intersubjective dimension which pertaining to the social relationship between speaker and hearer. He refers to this as the "stance triangle". A stance utterance encapsulates the stance, and can be regarded as its core element. Stance utterances make good objects for study because a) they are usually short and succinct, and b) they tend to summarize a speaker's story or narrative. Typical stance utterances might be "I'm sorry, but that's *not* exactly what I had in mind", "There's a *reason* why we do this", or "I don't even know if that's *enough*" (emphasis added). Further examples are given below, with purported effects (punctuation omitted):

(3) Typical stance utterances (all negative valency)	TOPIC
a. The worst is yet to come	[global warming]
b. Hillary (Clinton) does not inspire confidence	[politics]
c. Frankly, I can't understand how people put up with this	[migration]
d. The Internet hasn't enriched my life in any significant way	[modern life]
e. Keeping up relations takes a lot of work	[social obligations]
f. Every day I eat the same thing	[food]

Judgements of disposition are based on tone-of-voice as well as words, however. In order to test for it, it is necessary to expunge all lexical content. Nooteboom (2000) suggests using 'iterant' speech, that is substituting nonsense syllables for words, thus preserving prosodic features. At present this can only be done by humans, and is most effective when the forms are

produced immediately after voicing. To illustrate, the same utterances in (3) are repeated below as iterant speech:

(4) Iterant speech

a. da**DA** da*Da* da***Da:***

b. ***Dadada*** da**DA** da**Da**da **Da**dada

c. **Da**da | da**Da** dada***Da*** da dada da*Da* dada

d. *daDadada* dada dada **DA**da dadada ***daDadada***

e. ***dadada*** da*Da*da da dadada **DA**

f. dada**Da** dada dada **DA**

'Prominent' syllables appear in in upper case letters, with two degrees of prominence (onset or onset+vowel). These are all stressed syllables in English which might be represented by some other prosodic feature in another language. Prominence, or sentential stress is itself a kind of voice quality, pointing to extremely rapid displays of *wanting* or *not wanting*—[W] in the syntax of mental states). Metrical structure—and some hint of rhythm—is preserved in the grouping of syllables ([Z]). Most of the prominent syllables in (4)—and some non-prominent ones—are show relative pitch levels: bold (non-italic) stands for highest, bold italic for lowest, and italic for mid. The intervals between the tones are significant, but cannot be depicted in this transcription system. Tones in sequence and in harmony are responsible for the communication of melody and valency—[X] and [Y] in the theory of emotion we are assuming).

One Japanese and one English speaker produced scripted, 'emotional' utterances in reference to several topics. For each topic, one utterance was characterized by positive valency, another by negative valency. These were then converted to iterant speech and presented to separate groups of Japanese and American subjects. In one test, subjects were asked to appraise the disposition of the speaker (same and different languages). Only speech forms of one valency ([±]) were presented; no choices were offered. In a control test, speech forms of both valences were 'mixed'.

Subjects were prompted with a lexical 'introduction' to each topic, before hearing converted (iterant) utterances. Samples included *He was real bastard, didn't give a fig about the people who elected him* ([–]) vs. *Actually, he didn't do anything that everyone else before him had done* [+] (in reference to Masuzoe, the former mayor of Tokyo); *It doesn't taste the same, and it kills off all the nutrition* [–] vs. *I use it all the time* [+] (RE food/microwave ovens); *It sucks. Worst thing to hit the planet* [–] vs. *It's raining now, but it should be better soon* [+] (weather), etc.

Discussion

The results of these tests were predictable. Subjects could easily determine valency based on their choice of terms to describe perceived disposition, e.g.

grumpy, *cheerful,* or Japanese *ganko* 'stubborn', *rakutenteki* 'optimistic', etc. The 'mixed' test of utterances with positive/negative valency produced no consensus as to what kind of person the speaker was. While it is unfortunate that more nuanced appraisals of disposition beyond valency could not be obtained, to do so would be difficult given limited exposure to the speakers' tone-of-voice, the varied experiences of the participants, and the different conceptualizations of emotion in the languages themselves.

Listeners gather their impressions though repeated verbal exchanges. Not only through words (lexis), they may rely on prosodic features to build an imprint. Experiments have shown that listeners can do this based on iterant speech where lexical/semantic meaning has been stripped away. We have proposed that disposition is indeed analyzable in the same terms as 'emotions' generally, where the latter are understood as composites of mental states WXYZ related to types of sound (wxyz): voice qualities, sequential tones, tones produced simultaneously, and timing units.

What distinguishes 'disposition' from rapid, continuous displays of mental states is time. Given the similarity of (theoretically quantifiable) frequent displays, the listener will store them economically in terms of a general impression or 'imprint' with regard to the speaker. To judge someone's disposition then, is to have such an imprint. Regardless of topic, a speaker with a certain attitude will voice similar prosodic outlays over time. This can be shown with a more precise examination of interval sizes and 'harmonic' effects that arise between and among prominent tones. Listeners can recognize previously-heard constellations of sounds, and base their appraisal of speaker disposition on them. Speakers may also gravitate towards topics that facilitate the expression of their attitudes. This implies they sometimes choose words based as much on how they sound as on the meaning of words themselves. It is certainly a topic worthy of future study.

References

Cook, N. 2003. *Tone of Voice and Mind.* John Benjamins Publishing Co., Amsterdam.

Crystal, D. 1975. *The English Tone of Voice.* Edward Arnold, London.

DuBois, J. 2007. The stance triangle. In *Stancetaking in Discourse.* R. Englebretson (ed.), John Benjamins Publishing Co., Amsterdam.

Laver, J. 1980. The Phonetic Description of Voice Quality. CUP.

Nooteboom, S. 2000. The prosody of speech: Melody and rhythm'. MS, Research Institute for Language and Speech, Utrecht.

Wichmann, A. 2000. The attitudinal effects of prosody and how they relate to emotion. *Proc. of ISCA Workshop on Speech and Emotion;* Cowie, R., E. Douglas-Cowie, & N. Schroder (eds.)

Wierzbicka, A. 1999. Emotion Across Languages and Cultures. CUP.

Intonation and polar questions in Greek

Anthi Chaida, Angeliki Sotiriou, Athina Kontostavlaki
Laboratory of Phonetics and Experimental Linguistics, University of Athens, Greece

Abstract

The present study focuses upon the effects of lexical stress and focus on Greek polar (yes/no) questions. According to the results of a production experiment, the tonal structure of neutral questions presents striking similarities with the tonal structure of questions with focus on the final element. Questions with focus in the first element display a different tonal structure and do not show the typical F0 fall on the stressed syllable of the nucleus. The peak of the tonal boundary in these questions aligns with the last stressed syllable, while in neutral questions and in questions with focus in the final element it aligns with the last syllable of the utterance.

Key-words: intonation, polar questions, lexical stress, focus, Greek.

Introduction

This study aims to investigate the interaction of lexical stress and focus with the intonation of polar (yes/no) questions in Greek. Although different sentence types and specifically polar questions as well as focus have been the objective of several studies (e.g. Chaida 2010), the effect of the position of lexical stress on tonal contours, and especially on tonal boundaries, still remains an open question (see Botinis et al. 2016). According to previous studies, the tonal structure of polar questions consists of a low nuclear tone, followed by a risin-falling tonal movement at the right edge of utterances. More specifically, the tonal peak has been found to align either with the last syllable of the sentence when focus in the last word or with the last stressed syllable when focus earlier (Grice et al., 2000, Arvaniti 2002, Baltazani 2007, Chaida 2010).

Experimental methodology

One simple sentence was crossed with 3 focus renditions (no focus, focus on the first element, focus on the final element), and 3 lexical stress placements on the final element (Table 1). The speech material was placed in 3 lists with random order, and was produced by 10 female speakers aged 20-40 years old with standard Athenian pronunciation. The speakers were given verbal instructions and provided with contextual information and a suggested answer for every question. The total corpus of the recorded utterances consisted of 270 utterances (3 sentences X 3 focus renditions X 10 speakers X 3 repetitions).

The speech productions were directly recorded into a computer hard disk at the isolated sound recording booth of the Laboratory of Phonetics and Computer Linguistics of the University of Athens. The speech material was analyzed with Praat software, and the relevant data were automatically generated through the script Prosody Pro (Version 5.6.0) (Xu 2013). MS Excel and a Python script were used for the creation of graphs.

Table 1. Speech material of polar questions used for recordings, based on 3 different lexical stress placements (ˈ)crossed with 3 focus renditions (in bold)

STRESS-SYLLABLE	FOCUS	TARGET UTTERANCE
antepenultimate	No focus	[to peˈði ˈmeni ˈmonaxo]? (The child lives in Munich?)
antepenultimate	First element	[to **peˈði** ˈmeni ˈmonaxo]?
antepenultimate	Final element	[to peˈði ˈmeni ˈ**monaxo**]?
penultimate	No focus	[to peˈði ˈmeni moˈnaxo]? (The child lives alone?)
penultimate	First element	[to **peˈði** ˈmeni moˈnaxo]?
penultimate	Final element	[to peˈði ˈmeni **moˈnaxo**]?
ultimate	No focus	[to peˈði ˈmeni monaˈxo]? (The child lives alone?)
ultimate	First element	[to **peˈði** ˈmeni monaˈxo]?
ultimate	Final element	[to peˈði ˈmeni **monaˈxo**]?

Results and discussion

Figures 1-3 show the results of the present investigation. In accordance with these results, the position of lexical stress in the last word of the sentence and the position of focus affect the tonal structure of the question. In final focus and in neutral utterances, the position of lexical stress affects the alignment of the peak of the large F0 rise and consequent fall. The earlier the last lexical stress of the utterance, the earlier the aforementioned peak aligns within the utterance and the larger the inclination of the pitch curve.

As far as focus-first utterances are concerned, the position of the lexical stress of the final word is related with the position of the peak of the tonal boundary, which co-occurs with the final stressed syllable. In addition to this and contrary to the results of previous studies (e.g. Grice et al. 2000, Arvaniti 2002), instead of a fall of the value of the F0 on the stressed syllable of the word in focus, there is a rise in pitch. Furthermore, there appears to be no tonal range expansion associated with the word in focus, as the case is for declarative sentences (Botinis et al. 2001).

Neutral utterances and final focus utterances display similar tonal structures, because the nucleus is aligned in both cases with the right prosodic edge and not with the verb. Consequently, there can be no post-focal de-accenting. Regarding the alignment of the nucleus in neutral questions, the results of the present study differ from the results of previous studies, where the nucleus is in the verb (Baltazani 2007, Chaida 2010).

As to the effect of focus on tonal boundaries, the peak of the tonal boundary in questions with nucleus on the final word is aligned with the last syllable of the utterance in all cases. On the other hand, in questions with an early nucleus, the peak of the tonal boundary is aligned with the last stressed syllable. This finding is in line with previous studies (Grice et al. 2000, Arvaniti 2002, Baltazani 2007, Chaida 2010).

Focus last and neutral questions were produced, in general, in a consistent way and focus was given as expected. On the contrary, in early focus questions, focus was either given in the first element, as required (~62% of the utterances) or in both the beginning and the end of the utterance (~38%). Considering the above, further research on polar question prosodic features is required, since it seems that it still remains an open issue.

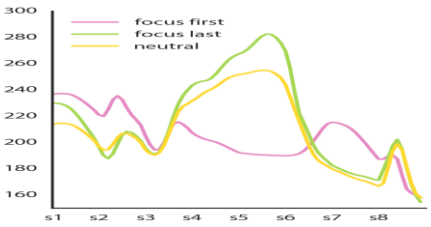


Figure 1a. Intonation of polar questions with stress on the antepenultimate syllable in 3 focus renditions.

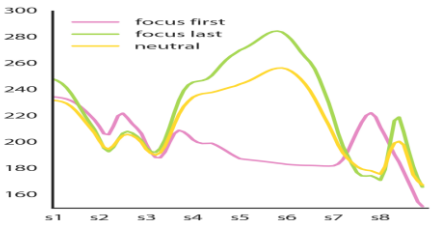


Figure 1b. Intonation of polar questions with stress on the penultimate syllable in 3 focus renditions.

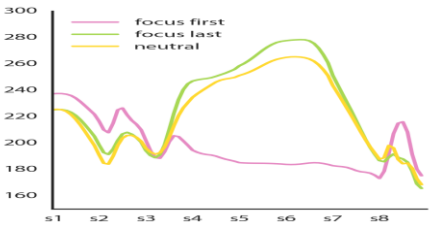


Figure 1c. Intonation of polar questions with stress on the ultimate syllable in 3 focus renditions.

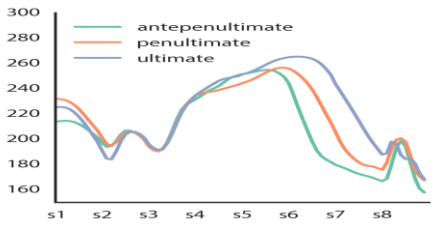


Figure 2a. Intonation of polar questions without focus, with 3 lexical stress placements on the final element.

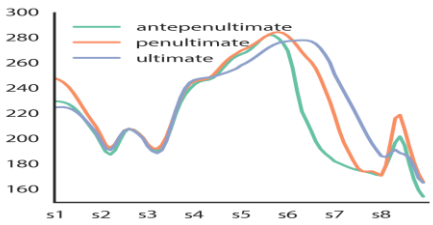


Figure 2b. Intonation of polar questions with focus as well as 3 lexical stress placements on the final element.

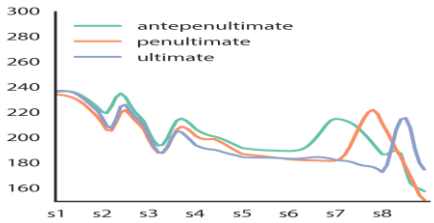


Figure 3a. Intonation of polar questions with focus on the first element and 3 lexical stress placements on the final. (Realisation A, 62% of the utterances).

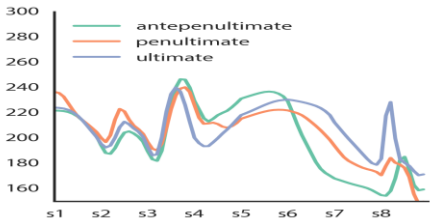


Figure 3b. Intonation of polar questions with focus on the first element and 3 lexical stress placements on the final (Realisation B, 38% of the utterances).

References

Arvaniti, A. 2002. The intonation of yes-no questions in Greek. In M. Makri-Tsilipakou (ed.), Selected papers on theoretical and applied linguistics 71-83. Thessaloniki: Aristotle University.

Baltazani, M. 2007. Intonation of polar questions and the location of nuclear stress in Greek. In Carlos Gussenhoven & Tomas Riad (ed.), Tones and tunes, Volume II: Experimental Studies in Word and Sentence Prosody (387-405). Berlin: Mouton de Gruyter.

Botinis, A., Chaida, A., Nikolaenkova, O., Nirgianaki, E. 2016. Intonation and polar questions in Greek revisited. Proc. ExLing 2016 (this volume).

Botinis, A. Granström, B., Möbius, B. 2001. Developments and paradigms in intonation research. Speech Communication 33 (4), 263-296.

Chaida, A. 2010. Production and Perception of Intonation and Sentence Types in Greek. PhD Thesis, University of Athens.

Grice, M, Ladd, R. & Arvaniti, A. 2000. On the place of phrase accents in intonational phonology. Phonology 17, 143-185.

Xu, Y. 2013. ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. In Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, France.7-10.

Contextual predictions and syntactic analysis: the case of ambiguity resolution

Daria Chernova, Veronika Prokopenya
Laboratory for Cognitive Studies, St. Petersburg State University, Russia

Abstract

We test the hypothesis that syntactic analysis is based on contextual predictions and is guided by discourse salience of the referents. The head of the complex noun phrase tends to be more prominent in discourse as native speakers expect the continuation of the story to refer to N1 more often than to N2. It corresponds to the data on adjunct attachment interpretation.

Key words: contextual prediction, syntactic analyses, referent activation

Introduction

The problem of syntactic ambiguity resolution is widely discussed in psycholinguistics being a testing ground for different parsing models (Traxler 2014). The question is what guides the choice of the interpretation when grammar allows several possible variants.

Adjunct attachment ambiguity (*I met the servant of the countess that was on the balcony*) is particularly widely discussed cross-linguistically as different preferences in different languages contradict the idea of universality and are inconsistent with the Late Closure Principle (Cuetos & Mitchell 1988, Grillo & Costa 2014). Previous studies (Sekerina 2003 Yudina et al. 2007), show high attachment preference for Russian.

We test the hypothesis that syntactic analysis is guided by discourse salience of the referents and is based on contextual predictions (Rohde et al. 2011). We presuppose that the listener/the reader expects further information about a more salient (activated) referent (Chafe 1994). The referent which is mentioned in a story-continuation task more often is more discourse salient and thus is more likely to attract the adjunct.

Method

Materials and design

12 experimental stimuli were constructed for fill-in-the-blank task. Each stimulus consisted of two sentences, the first sentence contained a complex noun phrase and the second sentence was the continuation of the first one and could refer either to N1 or N2 equally plausible (as in (1)). N1 and N2 had the same number, gender and animacy. The subject of the second

sentence was omitted and substituted by a gap which the participants were asked to fill by any appropriate word.

The questionnaire also included 62 fillers which contained no ambiguity.

(1) На улице я встретил служанку графини. Много лет _____ жила в доме неподалеку.

'I met the servant of the countess in the street. For many years _____ lived nearby'

Participants and procedure

40 native speakers of Russian naïve to the aim of the study were asked to fill in the questionnaire on voluntary basis.

Results

The gap was filled with one the following variants: N1 or its periphrasis, N2 or its periphrasis, a noun which could refer both to N1 and N2, 3rd person pronoun or any other word (see Table 1).

Table 1. Types of answers.

Type	Number of answers
N1 (служанка 'the servant')	33 (6.9%)
Periphrasis of N1 (горничная 'the maid')	70 (14.7%)
N2 (*графиня* 'the countess')	58 (12.2%)
Periphrasis of N2 (эта благородная дама 'the noble lady')	17 (3.6%)
Noun (эта женщина 'the woman')	32 (6.7%)
3rd person pronoun (она 'she')	233 (48.9%)
other	33 (6.9%)

As we see, most of the continuations are ambiguous as they contain 3rd person pronoun which can refer to N1 or N2 (or another noun that can refer to N1 or N2), ambiguous continuations actually prevail: χ^2=6.89, p=0.009.

If we consider unambiguous continuations only, N1 is mentioned in 57.9 % cases (103 answers) whereas N2 is mentioned in 42.1% cases (75 answers), this difference is statistically significant: χ^2=7.89, p=0.006.

Discussion

From our data we can draw two main conclusions:

- Pronominalization takes place in 50% of cases despite the potential referential conflict in the sentence which can be explained by the use of egocentric strategy (Kibrik 2011) and leads to ambiguity. Thus, native speakers tend not to avoid ambiguity despite the risk of potential communicative failure.

- N1, being the head of the complex noun phrase, tends to be more prominent in discourse. The continuation of the story is expected to refer to N1 1.3 times more often than to N2. It corresponds to the data on adjunct attachment interpretation and explains them: adjuncts, being the continuations of the sentence, tend to be attached to head of the complex noun phrase which is more expected to be modified, so syntactic analysis of potentially ambiguous sentences is affected by contextual predictions.

Acknowledgements

The study was supported by grant from Russian Humanitarian Scientific Fund, #14-04-00586

References

Chafe, W. 1994. Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing. Chicago, University of Chicago Press.

Sekerina, I. 2003. The Late Closure Principle in Processing of Ambiguous Russian Sentences. In: The Proceedings of the Second European Conference on Formal Description of Slavic Languages. Potsdam: Universität Potsdam.

Cuetos, F., & Mitchell, D.C. 1988. Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. Cognition, 30, 73-105

Traxler M. 2014. Trends in syntactic parsing: anticipation, Bayesian estimation, and good-enough parsing. Trends in Cognitive Science, 18(11), 605-611.

Yudina M.V., Fedorova O.V., Yanovich I.S. 2007. Sintaksicheskaja neodnoznachnost' v eksperimente i v zhizni. Dialogue, Moscow.

Kibrik A. A. Reference in discourse. Oxford: Oxford University Press, 2011

Grillo, N., & Costa, J. 2014. A novel argument for the universality of parsing principles. Cognition, 133, 156- 187.

Rohde, H., Levy, R., & Kehler, A. 2011. Anticipating explanations in relative clause processing. Cognition, 118, 339-358.

Vocal fatigue in voice professionals: collecting data and acoustic analysis

Karina Evgrafova, Vera Evdokimova, Pavel Skrelin, Tatiana Chukaeva
Department of Phonetics, Saint Petersburg State University, Russia

Abstract

The present study examines acoustic manifestations of the vocal fatigue in three groups of voice professionals (pronunciation teachers, professional speakers and tourist guides) who seem to be particularly susceptible to vocal loading. In the paper data collecting and the non-fatigue/fatigue speech corpus are described. The detailed acoustic analysis of the data obtained is presented. The results of the acoustic analysis showed a consistent dependency between acoustic parameters and vocal fatigue in terms of F0, jitter and shimmer values. The results can contribute to objective voice examinations and automatic voice pathology detection.

Key words: vocal fatigue, acoustic analysis, voice professionals, speech corpora

Introduction

Vocal fatigue is a voice disorder which particularly concerns professional voice users and can lead to serious pathological conditions Teachers, singers, actors, guides and all types of professional speakers that require prolonged voice use are identified as an at-risk group for developing vocal disorders. The symptoms of vocal fatigue are various and explained by the physiologic mechanisms of vocal production. There exist many studies on vocal fatigue providing various concepts of the phenomenon. However, there is no universally accepted definition. It can be viewed either as a voice disorder caused by other pathological voice conditions or as a separate voice problem resulting from prolonged and excessive voice use [10]. In this study the vocal fatigue is understood as a separate phenomenon caused by excessive professional voice load which results in auditory perceptual and acoustic changes in the voice signal and can lead to serious pathological conditions. The present study paper is aimed to describe the data collecting for the non-fatigue/fatigue speech corpus and to present the results of acoustic analysis.

Methods

The methodologies that attempt to induce vocal fatigue in experiment participants vary across numerous works on the vocal fatigue [1-9]. In most studies the vocal fatigue is induced artificially as a result of reading or speaking tasks of various types. The results described are inconsistent and often conflicting. The conditions of our experiment seem to be more

realistically challenging. 20 male and female subjects were recorded. They involved pronunciation teachers with average work experience of 7 years, professional speakers (broadcasters) and tour guides with the work experience not less than 5 years. No one had pathological voice problems. The participants were asked to read at habitual loudness a four minute phonetically representative text.

The teachers were recorded before and after a 7 hour teaching day. The tour guides were recorded before and after 3 hour non-stop excursion and professional speakers – before and after 3 hour non-stop interview/3 hour non-stop recording of a book. All the subjects were asked to fill in a special questionnaire before each type of the recordings. In the questionnaire they evaluated their physical state, mood and a level of activity. The recordings were made in the recording studio at the Department of Phonetics, Saint Petersburg State University.

Results

We calculated (in Praat) a number of acoustic parameters based on formant values, jittter, shimmer, pitch and loudness which can help detecting the absence/presence of voice fatigue in a given speech sample. The parameters which seem to be most important for automatic detection are the mean value of F0, jitter and shimmer values.

The calculations showed that the main tendency for both male and female speakers was the increase in the mean value of F0 in the fatigued speech across all the speaker groups. However, the jitter values become lower. As to the shimmer value, there can be seen the decrease in fatigued female voices and the increase in fatigued male voices. The tables 1-3 below show the results.

Table 1. F0 and duration mean values. Non-fatigue vs. fatigue speech.

		Duration (sec)	Unvoiced parts (%)	Mean F0
Female	non-fatigue	214	45,7	209
	fatigue	220	47,0	212
Male	non-fatigue	217	48,1	124
	fatigue	213	46,1	130
All	non-fatigue	215	46,4	185
	fatigue	218	46,7	188

Table 2. Jitter mean values. Non-fatigue vs. fatigue speech.

		Jitter				
			local, absolute (seconds)	rap %	ppq5 %	ddp %
Female	non-fatigue	2,283	0,00011	1,002	1,051	3,008
	fatigue	2,208	0,008578921	0,97	1,036	2,91
Male	non-fatigue	3,239	0,000272208	1,273	1,421	3,82
	fatigue	2,888	0,000228958	1,085	1,229	3,254
All	non-fatigue	2,556	0,000156442	1,08	1,157	3,24
	fatigue	2,403	0,006036776	1,003	1,091	3,008

Table 3. Shimmer mean values. Non-fatigue vs. fatigue speech.

		Shimmer					
		local %	local, db (dB)	apq3 %	apq5 %	apq11 %	dda %
Female	non-fatigue	8,022	0,833	2,653	4,068	7,871	7,96
	fatigue	8,108	0,837	2,666	4,168	8,008	7,998
Male	non-fatigue	11,003	1,063	3,777	5,775	12,18	11,33
	fatigue	10,377	1,015	3,521	5,387	11,28	10,56
All	non-fatigue	8,874	0,898	2,974	4,556	9,103	8,923
	fatigue	8,756	0,887	2,91	4,516	8,943	8,731

There is also difference in the amount of pauses and their duration between the female and male fatigued recordings. The whole number of pauses tends to increase in the female fatigued speech while the number of pauses in the male fatigued speech decreases. The duration of pauses in the fatigued speech increases both in the male and female recordings.

Conclusions

The results of the voice acoustic analysis of the fatigued speech in comparison with the non-fatigued speech showed a consistent dependency between acoustic parameters and vocal fatigue. The parameters which are affected by the vocal fatigue are the F0, jitter and shimmer values, the duration and number of pauses. The differences in the acoustic parameters before and after vocal loading mainly seem to reflect increased muscle activity as a consequence of excessive vocal loading.

The results can contribute to objective voice examinations and automatic voice pathology detection.

References

Boucher, V.J. 2008. Acoustic Correlates of Fatigue in Laryngeal Muscles: Findings for a Criterion-Based Prevention of Acquired Voice Pathologies. Journal of Speech, Language, and Hearing Research, vol. 51, 1161–1170.

Caraty, M.J., Montacié, C. 2010. Multivariate Analysis of Vocal Fatigue in Continuous Reading, Proceedings of Interspeech 2010, 470-473.

Kostyk, B.E., Rochet, A.P. 1998. Laryngeal airway resistance in teachers with vocal fatigue: a preliminary study. Journal of Voice, vol. 12, 287–299.

Sala, E., Airo, E., Olkinuora, P. et al, 2002. Vocal Loading among Day Care Center Teachers". Logoped Phoniatr Vocol,vol. 27, 21–28.

Schneider, B. 2006. Effects of Vocal Constitution and Autonomic Stress-Related Reactivity on Vocal Endurance in Female Student Teachers. Journal of Voice, vol. 20, No. 2, 242–250.

Scherer, R.C., Titze, I.R. et al. 1986. Vocal fatigue in a professional voice user. In Transcripts of the Fourteenth Symposium: Care of the Professional Voice, New York: The Voice Foundation, pp.124–130.

Scherer, R.C., Titze, I.R. et al. 1991. Vocal fatigue in a trained and an untrained voice user. Laryngeal Function in Phonation and Respiration, San Diego, Singular Publishing Group, pp. 533–555.

Titze, I. , Lemke, J., Montequin, D. 1997. Populations in the U.S. workforce who rely on voice as a primary tool of trade: a preliminary report. Journal of Voice, vol. 11, 254–259.

Laukkanen, A.M. 1995. On speaking voice exercises. PhD dissertation, Acta Universitatis Tamperensis, ser A, vol. 445, Tampere: University of Tampere.

Creating a subcorpus of a heritage language on the example of Yiddish

Valentina Fedchenko[1], Ilia Uchitel[2]
[1]Department of Jewish Culture, St. Petersburg State University, Russia
[2]School of linguistics, Higher School of Economics in Moscow, Russia

Abstract

The paper presents a Yiddish heritage subcorpus on the basis of the Corpus of Modern Yiddish. The contemporary status of the Yiddish language and the absence of monolingual speakers nowadays makes it perfect candidate for research within the framework of heritage languages. Yiddish exists in different sociolinguistic contexts and forms plenty of bilingual pairs. Corpus-linguistic approach, especially in corpora with multimedia utilities and L2 component, enlarges the variety of possible instruments and subjects of research. The paper discusses practical issues of creating and using a multimodal corpus of the Yiddish language with a special focus on the more recently added subcorpus of recorded interviews with L2 speakers of Yiddish, while analyzing the corpus architecture, the corpus representativity, L2 corpus marking.

Key words: corpus linguistics, Yiddish, multimedia corpus, L2 corpus, heritage language.

Yiddish as a heritage language and L2

This short article presents the project of developing corpora tools aimed at producing quantitative research of Yiddish as a heritage and learned as second language (L2).

Yiddish language is a West-Germanic language spoken mainly by Ashkenazi Jews. In the beginning of XXth century it was main language of communication, both oral and written, for Jews in central and eastern Europe, including vast lands of former Russian and Austro-Hungarian Empires. Since then the dialectal diversity in Yiddish remains very high. After the Holocaust and mass migration to Israel and the US, Yiddish speakers has nearly disappeared in their original language areas (i.e. today Poland and Ukraine). Still, language situation in different areas, vary. While in Poland and Lithuania Yiddish ceased to be language of communication shortly after the WW2, in Ukraine, Belarus and Moldova, where the number of Jewish survivors was much bigger, Yiddish continued to be used and even be acquired by post-war children till the 1980s – the beginning of mass migration to Israel and other countries.

Having lost a great number of its speakers, the Yiddish language is still spoken by some communities in Israel and the US. It is still considered to be

an important part of Jewish identity, with some scholars qualifying it as a language which is in some cases used as "post-vernacular" [Avineri 2012: 25] and surpasses the ordinary communicative use by developing extra-functions (i.e. cultural, symbolical). As an example of such use we can mention widespread use of Yiddish in several language programs, such as "Yidish Vokh" ('Yiddish Week') or "Yidish dorf" ('Yiddish Village'), where participants (numerously not native speakers) are supposed to speak only Yiddish. This "post-vernacular" use continues in the Internet with heritage activists which use Yiddish, for example, in their everyday Facebook activities.

In the past decade, the subject of heritage Yiddish use (mainly in the US) have been studied by several scholars in such works as [Shandler 2008], [Avineri 2012], [Sadan 2011], [Levine 2000]. In addition to these thorough sociolinguistic descriptions of the current practices, there are small amount of works focusing on analysis of specific features of Yiddish as a heritage language. We can note, for example, [Safadi 2000], discussing noun gender and case differences between groups of heritage and native speakers, and, partially, [Levine 2000] with discussion on choice of auxiliary in perfect tense among heritage and non-heritage speakers.

Corpus data

One of the most valuable projects made during recent years are Corpus of Modern Yiddish (http://web-corpora.net/YNC/search/) and Yiddish Multimedia Corpus (http://web-corpora.net/YiddishMultimediaCorpus/search/) . First one includes documents representing language of press and fiction of the XIXth till late XX centuries, including modern documents. Second corpus presents annotated audio records of authentic Yiddish speech, with speakers coming from various dialect areas. It includes 10 files: lectures and field recordings. With online search available, these sources give a great possibility for performing quantitative studies of Yiddish language, as well as learning Yiddish as second language.

CMY contains currently 4 150 933 tokens from 3662 documents. The largest part of it is press with a share of 78.43% is mostly represented by archive of "Forverts" newspaper, publishing in the US, with issues dating from 2004. Some of the press text authors are not native speakers, or, to some extent, are heritage speakers.

Therefore, the first step for construction of heritage and L2 subcorpus is looking for sociolinguistic characteristics of the authors. Such characteristics should include, at least, type of language knowledge (native/heritage/second), first language (if appreciable), year of birth of the speaker. Some additional information, however, would enrich the set: for

example, details about parents' place of birth and their knowledge of Yiddish.

Though there are many parsed documents of this type in CMY and there can be some findings results just after adding sociolinguistic information, the overall result cannot be considered as fully reliable and representative for several reasons: 1. texts published in press are pre-edited, even if in some of them (i.e. ultraorthodox Hasidic newspapers) some idiolect features are tolerated; 2. print text genre represent different linguistic peculiarities, comparing with colloquial language examples. Yiddish multimedia corpus either cannot provide data from heritage and L2 speakers.

In order to collect relevant information, oral interviews with heritage speakers can be produced. While the advantages of this approach are clear, there are some serious shortcomings. The most important problem is the time-consuming manual transcription of interviews. This can be avoided by interviewing consultants in writing, primarily, using Internet. However, in the written data we can hardly find some traces (i.e spelling mistakes) to phonological features of consultant's speech. In addition, such interviews can be conducted with a limited number of speakers, who use internet for some communication in Yiddish, with arising problem of transcription processing. Moreover, there is a possibility to proceed automatically some amount of independently produced texts by a limited set of "language activists", who use Yiddish in their public communication on Internet (basically, in communication on Facebook). This material which demonstrates a real language use, can be very useful.

Text processing and access

A well-developed tagging engine already exists for Corpus of Modern Yiddish, so that the inserted texts can be quickly morphologically parsed (with certain inaccuracy due to homonymy). The heritage subcorpus can be built into the CMY, with adding search by sociolinguistic metadata, or it can be hosted independently, but will share the CMY platform.

One of the greatest difficulties in building a L2 subcorpus is the marking of mistakes or "non-standard language features". It takes remarkably much time to find and classify a mistake, which can occur at all levels of language. Even for Russian Learning Corpus team of native Russian speakers a special error-checking engine was necessary for facilitating the mistake marking.

As the stage of gathering data for corpus is not fulfilled yet, the time-consuming process of mistake marking should be postponed. Nevertheless, the option to search by mistakes (or "distinct features") is very important not only for theoretical reasons, but for applied purposes (as language teaching) as well.

However, when applying this principle to Yiddish data, we can face a different kind of problem. The error tagging is quite difficult even for native speakers. A vague standard language and a high dialectal diversity makes this task nearly impossible. Probably, another way to highlight distinct features of heritage Yiddish should be used. One simple way to do it is to use bi-grams and tri-grams of some lemmas and then compare results of heritage and non-heritage language. This method was used in automatic error-detection tool made for Russian Learner Corpus [Klyachko et al. 2013]. A spell checking engine would be very useful for tracing mistakes in certain documents. There are some Yiddish spell checkers, elaborated by YIVO.

One more problem concerns the analysis of phonological features in the heritage speech. In several corpus projects (i.e. YMC) the original speech was transcribed according to standard Yiddish rather than phonetically. That leads to difficulties in quantitative analysis of phonology. Unfortunately, transcribing audio according to IPA alphabet would take tremendous efforts and can't be done at the time. However, YMC interface allows to look for individual word and then find it easily in the audio file. Therefore, inserting new heritage records into the YMC engine and tagging it seems to be the best solution at time.

Acknowledgements

The authors thank for financial support the Russian Science Foundation (project 15-18-00062, St. Petersburg State University).

References

Avineri, N. R. 2012. Heritage language socialization practices in secular Yiddish educational contexts: the creation of a metalinguistic community. Disertation in Applied Linguistics, University of California.

Klyachko E., Arkhangelskiy T., Kisselev O., Rakhilina E. 2013. Automatic error detection in Russian learner language. Corpus Linguistics. Lancaster, UK, July 22–26, 2013.

Levine, G. S. 2000. Incomplete L1 acquisition in the immigrant situation: Yiddish in the United States. Tübingen, Niemeyer.

Sadan, T., 2011. Yiddish on the Internet. Language and Communication, 31(2), 99-106.

Safadi, M. 2000. Yiddish: its survival in an English-dominant environment. Dissertation, University of California.

Shandler, J. 2008. What is American Jewish Culture? In Raphael, M. L. (ed.), The Columbia History of Jews and Judaism in America, 337-365. Columbia University Press, New York.

Affricates in the spontaneous speech of Aromanians in Turia

Anastasia V. Kharlamova
General Linguistics Department, Saint-Petersburg State-University, Russia

Abstract

This paper deals with the affricate inventory of Aromanian spontaneous speech, using the spoken materials collected in Turia (Greece) in 2002 for the Small Dialectological Atlas of the Balkan Languages. The purpose is to analyse the affricates present in the Turia Aromanian dialect and their development. The texts, which had been previously put down in Romanian-based Aromanian orthography with the help of a native Aromanian speaker, were transcribed using computer programs Sound Forge and Speech Analyzer. The instrumental analysis shows that there are eight affricates to be found in our materials: [t͡s], [d͡z], [t͡s'], [d͡z'], [t͡ʃ], [d͡ʒ], [t͡ɕ], and [d͡ʑ]. However, there is evidence of these sounds – most notably [t͡s] and [d͡z] – being in the process of losing their stop phase. On the other hand, there are also instances of an opposite process, namely a fricative phase appearing after [t]. Both processes are fairly well-known typologically and among the Indo-European languages, but there has been previously little to no research on affricate development in Aromanian.

Key words: instrumental phonetics, Aromanian language, Turia, affricates, stops

The Aromanian language

Native speakers of the Aromanian language, a Romance language that belongs to the Eastern Romance subgroup, live in Greece, Albania, Romania, FYRM, Bulgaria and Serbia. Their exact number is unknown mostly for the reason that they usually identify themselves as people of the titular nation of their country (Nedelkov 2009: 247).

Academic research of Aromanian phonetics is mostly done by Romanian dialectologists – according to the Romanian linguistic tradition, Aromanian is a Romanian dialect (Capidan 1932). However, there have previously been no papers dedicated specifically to Aromanian affricates and/or Aromanian spontaneous speech, the only exception being the author's pilot research project (Харламова 2015).

The phonological system of Aromanian is stated to include four affricates – /t͡s/, /d͡z/, /t͡ʃ/, and /d͡ʒ/ (Нарумов 2001: 641). The original Romance affricates are a result of palatalization of Latin stops in the position before front-row vowels (Meyer-Lübke 1890: 318-342). Affricates also occur in Slavic, Turkic (Rothe 1957: 62), Greek, and Albanian (Gołąb 1984: 40) borrowings. It should be noted that the presence of /d͡z/ and /d͡ʒ/ in the

consonant system is one of the chief differences between Aromanian and Romanian on the phonological level, for in Romanian these sounds have long lost their stop phase (Meyer-Lübke 1890: 318-342).

Turia Aromanian

Kranea (Greek), or Turia (Aromanian), is a village with a population of circa 600, located in the Pindos Mountains in Greece, on the border between the administrative districts of Western Macedonia and Thessalia (Бара и др. 2005: 16). The inhabitants of the village identify themselves as Greeks, but call themselves "Vlachs" (*Βλάχοι*), and their language *limba noastrā* 'our language', *vlāhești* 'Vlach', and *armānești* 'Aromanian'. There is a widespread opinion among them that Aromanian can't be written (Бара и др. 2005: 17).

The Turia variety of Aromanian is given a full and highly detailed description in (Бара и др. 2005). We shall here only summarize the phonetic characteristics of this dialect.

It represents many of the chief features of Southern Aromanian dialect zone, among them the reduction of non-accented /e/ and /o/ into /i/ and /u/, the occurrence non-syllabic /^{u}/ and /^{i}/ after final consonants, syncopes, etc (Бара и др. 2005).

The recordings of the Turia Aromanians' spontaneous speech that we used in our research are available on a CD attachment to (Бара и др. 2005). The list of speakers is given in (Бара и др. 2005: 20-22). We mostly used the recording of the speech of Anastasia Pissoni (born in Turia in 1931, housewife).

The first transcription of the analyzed texts had been made by M. Bara, one of the authors of (Бара и др. 2005), herself a speaker of Aromanian. However, it was based chiefly on her language intuition, and therefore often reflects her interpretation of the sounds rather than what really was recorded.

Our own transcription was made with the help of two programs developed for phonetic and acoustic research – Sound Forge and Speech Analyzer. Sound Forge was used for building oscillograms and writing down the transcription, while Speech Analyzer was used for spectrograms.

The affricate inventory

We have found out that the affricate inventory of Turia Aromanian spontaneous speech consists of the following sounds: /t͡s/, /d͡z/, [t͡ʾs'], [d͡ʾz'], and /t͡ʃ/. There are also several doubted occurrences of /d͡ʒ/ and alveolo-palatal affricates [t͡ɕ] and [d͡ʑ]. The affricates found in our materials are considered among the most widespread affricate sounds typologically (Berns 2014: 382).

However, sometimes affricates didn't appear in the positions where they ought to have been (according to (Papahagi 1974) and (Бара и др. 2005)), instead being replaced by homorganic fricatives. There were other occasions, of [t͡s] occurring in place of /t/.

The loss of stop

The loss of stop in affricates in our materials didn't occur regularly. We have found occurrences of complete change of [t͡s] and [d͡z], as well as their palatal equivalents, into fricatives, and of weakened stop in [t͡ʃ]. There is too little data on the rest of the affricates to draw any conclusions from it.

The change of affricate into a fricative sound occurred usually in short frequent words, most notably *ți* 'what', and *ḑāsi* 'said'. The resulting fricative in place of an alveolar affricate could be [s], [s'], [ʃ], [z], or [z'].

In the transcription made by M. Bara, there are 50 occurrences of *ḑ* and 94 *ț*. Of them, 24 *ḑ* (about 50%) and 38 *ț* (about 40%) were found to have lost the stop phase in pronunciation.

As for the weakening stop in [t͡ʃ], there are no statistics on it for now, mainly because there is of yet no definite scale of stop strength; therefore, we should first set a border between a strong stop and a weak one.

If we look at the facts of phonetical typology, we find that the disappearance of stop phase has been observed or reconstructed in many languages, according to (Kümmel 2007). In (Żygis et al. 2012: 299) it has been suggested that the voiced affricates are more likely to lose their stop phase, due to their complicated articulation.

Affricated [t]

There are several clear cases of affrication of [t] in the analysed data: 3 occurrences of full affrication and 10 appearances of an audible fricative phase. All of them are recorded before front vowels.

In (Kümmel 2007) this process is mostly found in reconstructions. However, there are two well-known and notable examples of [t]-affrication, one of them occurring in late Latin and influencing the whole subsequent Romance group, and the other being one of the results of the High German consonant shift.

Therefore, although affrication is not as widely spread as loss of stop phase, it is still not a rare process typologically. Most importantly, it has already taken place once in the history of the Romance languages.

Future research

Our main perspectives for future research include: first, observation of this dialect's development over the years; second, collection of spontaneous speech data from other Aromanian dialects; third, use of our knowledge of contacts of Aromanian with other languages to better describe and predict its language changes.

References

Berns, J. 2014. A Typological Sketch of Affricates. Linguistic Typology, 18 (3).

Capidan, Th. 1932. Aromânii. Dialectul aromân. București, Imprimeria națională.

Gołąb, Z. 1984. The Arumanian dialect of Kruševo in SR Macedonia, SFR Yugoslavia. Skopje, Macedonian Academy of Sciences and Arts.

Kümmel M. 2007. Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion. Wiesbaden, Reichert Verlag.

Meyer-Lübke, W. 1890. Grammatik der Romanischen Sprachen, Bd. 1. Leipzig, Fues's Verlag.

Nedelkov, J. 2009. The Ethnic Code of the Vlachs at the Balkans. EthnoAnthropoZoom 6, 221-253.

Papahagi, T. 1974. Dicționarul dialectului aromân general și etimologic. Ediția a doua augmentată. București: Editura Academiei Republicii Socialiste România.

Rothe, W. 1957. Einführung in die historische Laut- und Formenlehre des Rumänischen. Tübingen, Max Niemeyer Verlag.

Бара М., Каль Т., Соболев А. Н. 2005. Южноарумынский говор села Турья (Пинд). München, Biblion Verlag.

Нарумов, Б. П. 2001. Арумынский язык/диалект. In Жданова Т. Ю. и др. (ред.) 2001, Языки мира. Романские языки, 636–656. Москва, Academia.

Харламова, А. В. 2015. Опыт фонетического анализа арумынской спонтанной речи. In Чердаков Д. Н. (ed.), XVIII Международная конференция студентов-филологов. Тезисы докладов. Санкт-Петербург, Филологический факультет СПбГУ.

L1 transfer, definiteness and specificity of determiners in L2 English

Sviatlana Karpava
University of Central Lancashire, Cyprus

Abstract

This study investigates L1 transfer from Cypriot Greek (CG), definiteness and specificity of determiners in L2 English. 100 CG undergraduate students (ages 17-23) participated in the study. The linguistic (socio-economic) background questionnaires were used. Their written corpus (100 essays) was analysed in terms of determiner production. They were also offered an elicitation task based on Ionin et al. (2003, 2004), which was focused on elicitation of definite determiner *the* in [+def; +spec] and [+def; –spec] environments and indefinite determiner *a* in [–def; +spec] and [–def; –spec] environments. The results of the study showed that the most problematic condition for CG students was [–def; +spec] with target indefinite determiner as they fluctuated in their written production between target and non-target settings.

Key words: determiners, definiteness, specificity, L1 transfer

Introduction

It was found that L2 English acquisition of articles is a very difficult process (Huebner, 1983; Master, 1987; Parrish, 1987; Robertson, 2000; Leung, 2001; Ionin et al., 2008). L2 leaners make omission or substitution errors (Larsen-Freeman, 1975; Thomas, 1989; Parodi et al, 1997; Hawkins et al., 2006). L2 learners either have access to Universal Grammar (UG), directly or via their L1, which is in line with the domain-specific view of L2 acquisition, or they use general learning mechanisms such as statistical learning, which is in line with the domain-general view (Ionin et al., 2008).

Definite articles are presuppositional expressions, while indefinite articles are quantificational expressions, as for the latter there is no prior presupposition or mentioning (Heim, 1991). In English, definite article *the* presupposes that the referent has been established by prior knowledge or discourse and this knowledge is shared by both a listener and a speaker (Ionin, 2003, 2006). Learning of articles involves form-meaning mapping. Definiteness is one of the cross-linguistic semantic universals, the other is specificity. L2 learners have access to both universals and they fluctuate between them. Ionin et al. (2003, 2004, 2008) observed that L2 learners of English have more accurate performance on [+def; +spec] and [–def; –spec], when there is agreement between definiteness and specificity, than on [+def; –spec] and [–def; +spec], when the two universals are in conflict. English

articles encode definiteness rather than specificity, therefore L2 English input provides target-like definiteness patterns and L2 learners with a higher level of proficiency might be more successful than those with a lower one.

L1 Cypriot Greek (CG) has articles, which means that L2 learners of English with CG background would either transfer semantics of Greek article into English or fluctuate between definiteness and specificity semantic universals provided by UG (Ionin et al., 2003, 2004, 2008).

The aim of this study is to examine L2 acquisition of English definiteness and specificity of determiners, whether L1 transfer overrides fluctuation or fluctuation overrides L1 transfer and whether amount and quality of L2 input, level of proficiency and age affect L2 learners' production with respect to definite and indefinite articles.

Study

100 CG undergraduate students (ages 17-23, L2 proficiency: beginners, intermediate and advanced) participated in the study. The linguistic (socio-economic) background questionnaires were used. Their written corpus (100 essays) was analysed in terms of determiner production. They were also offered an elicitation task based on Ionin et al. (2003, 2004), which was focused on elicitation of definite determiner *the* in [+def; +spec] and [+def; –spec] environments and indefinite determiner *a* in [–def; +spec] and [–def; –spec] environments. The participants were offered to choose from three options each time (*the*, *a* or *Ø*), there were 10 items for each condition. The task also investigated whether L2 learners of English transfer from L1 and they were asked to choose the appropriate variant (*the*, *a* or *Ø*) in such semantic and syntactic environments, where CG and English differ in terms of article use (Holton et al., 2004; Buschfeld, 2013). There were also distractor items focused on the use of various tenses.

Results and discussion

The results of the study showed that the most problematic condition for CG students was [–def; +spec] with target indefinite determiner as they fluctuated in their written production between target (42.55%) and non-target (57.45%) settings. They mainly substituted indefinite article *a* by *the* (52.12%) or used null determiner (5.31%). As far as other conditions concerned, for [+def; +spec] condition they had 76.38% target *the* and 23.62% non-target (12.55% indefinite article or 11.07% omission); for [+def; –spec] condition they used target *the* (73.40%) and 26.60% non-target (20.21% indefinite article and 6.39% omission); and for [–def; –spec] condition they had target *a* (78.29%) and 21.71% non-target (12.34% definite article and 9.37% null article), see Table 1.

Table 1. Definite vs. indefinite article production in four environments.

Environment	target *the*	non-target	non-target *a*	non-target Ø
[+def; +spec]	76.38%	23.62%	12.55%	11.07%
[+def; –spec]	73.40%	26.60%	20.21%	6.39%
Environment	target *a*	non-target	non-target *the*	non-target Ø
[–def; +spec]	42.55%	57.45%	52.12%	5.31%
[–def; –spec]	78.29%	21.71%	12.34%	9.37%

According to *one-way ANOVA*, age seems to be an important factor for the production of target *the* determiner in [+def; +spec] environment/condition: *Sig 2-tailed .005*. Age of onset to L2 English seems to be important for the target production of definite determiner *the* in [+def; –spec] condition *Sig 2-tailed .047*.

According to *paired samples t-test*, there is a statistically significance between the target production of indefinite determiner *a* in [–def; +spec] and [–def; –spec] conditions: *t(99)=11.861, p=.000*; between the target production of definite determiner *the* in [+def; +spec] condition and target production of indefinite article *a* in [–def; +spec] condition: *t(99)=8.702, p=.000*; between target production of definite article *the* in [+def; –spec] condition and target production of indefinite article *a* in [–def; +spec] condition: *t(99)=6.290, p=.000*.

It was found that L2 learners of English transfer from L1 CG, but the rate of transfer is low: they used definite determiners with *proper names* and *places* (24.69%), before *time expressions* (17.66%), with nouns that are additionally modified by *a demonstrative* and *possessive* (12.77%), quantifiers *all* and the *whole* (36.18%), with *most of* (54.47%). They tend to omit indefinite articles in predicate DPs after verbs *to be* and *to become* (21.28%), with expression *like* (21.71%), in direct object position with the verb *have* (32.77%), see Table 2.

According to *one-way ANOVA*, age is an important factor for the omission of articles in time expressions due to L1 transfer: *Sig 2-tailed .005*. The results of the study showed that fluctuation overrides L1 only for [–def; +spec] condition, when two semantic universals, definiteness and specificity are not in agreement. This finding is in line with Ionin et al. (2008) and Trenkic (2000) as L2 learner had an overall better performance in the use of definite than indefinite articles. Age is a statistically important factor for definite/indefinite article acquisition in L2 English, but not the level of proficiency, quantity and quality of input. CG participants transfer from L1 and might not pay attention to discourse-based triggers in L2 English.

Table 2. L1 transfer from CG.

target Ø	non-target	non-target Ø	non-target *the*	non-target *a*
with proper names/place names				
75.31%	24.69%	12.12%	7.68%	4.89%
preceding time destination, hours, weekdays, months, years and before seasons				
82.34%	17.66%		10.63%	7.03%
with nouns which are additionally modified by a *demonstrative* and *possessive*				
87.23%	12.77%		4.89%	7.88%
with nouns that are additionally modified by the quantifiers *all* and *whole*				
63.82%	36.18%		21.70%	14.48%
in predicate DPs after the verbs *to b*e and *to become*, predicate structures, simple DPs				
78.72%	21.28%	16.18%	5.10%	
with *like*				
78.29%	21.71%	15.33%	6.38%	
with most of				
45.53%	54.47%		49.36%	5.11%
in direct object position with the verb *have*				
67.23%	32.77%	30.63%	2.14%	

References

Buschfeld, S. 2013. English in Cyprus or Cyprus English? An Empirical Investigation of Variety Status. Amsterdam: John Benjamins.

Heim, I. 1991. Artikel und Definitheit. In Stechow, A and Wunderlich, D. (eds.), Semantik: Ein internationales Handbuch der zeitgenossischen Forschung, 487-535. Berlin: de Gruyter.

Ionin, T. 2003. Article Semantics in Second Language Acquisition. Unpublished doctoral dissertation, MIT.

Ionin, T. 2006. This is definitely specific: specificity and definiteness in article systems. Natural Language Semantics 14, 175-234.

Ionin, T., Ko, H. and Wexler, K. 2004. Article semantics in L2-acquisition: the role of specificity. Language Acquisition 12, 3-69.

Ionin, T., Zubizarreta, M.L. and Maldonado, S.B. 2008. Sources of linguistic knowledge in the second language acquisition of English articles. Lingua 118, 554-576.

Trenkic, D. 2000. The acquisition of English articles by Serbian speakers. Unpublished PhD dissertation. University of Cambridge.

Writing-based wordforms vs. spoken wordforms

Vadim Kasevich[1], Iuliia Menshikova[2]
[1]Faculty of Asian and African Studies, SpbU, Russia
[2]Faculty of Philology, SpbU, Russia

Abstract

This study addresses a very important problem of reshaping Russian Grammar in conformity with its real acoustic realization rather than with the traditionally written expression plane. In this way, one can switch from an absolutely abstract coding of wordforms to acoustic entities, first phonological and then phonetic, which underlie the real processes of speech production and speech perception. The multiple approach to grammar writing makes it necessary to develop a special database for the phonologically represented wordforms of Russian. Typically, the respective paradigms are reduced. More generally, the links camouflaged by the traditional orthography are made visible. E.g., the Adjective Gender paradigm, normally made up of three genders, is reduced to a two-item paradigmatic structure, because Neuter Gender and Feminine Gender just merge.

Key words: linguistics, Russian language, grammar, phonetics, morphemics.

Introduction

Natural language grammars as we know them may differ in many ways depending on the theories that underlie them. However different, the vast majority of the existing grammars share at least three important things in common, viz. (i) practically all of them are designed to account for the formal structure of the language rather than for its functioning, (ii) even where the grammars somehow model the dynamic nature of the language, the sets of rules are typically intended for the speakers rather than for the hearers, (iii) most grammars present their paradigms etc. in terms of standard orthography rather than in terms of phonological representations.

Unlike the prevailing tradition referred to above, we choose an approach where the grammar (of Russian) is modelled as a set of rules designed for the hearer. Since the hearers operate with sound patterns of linguistic entities the expression plane of the entities is expected to be presented in terms of the phonemes. E.g. the wordform КУПАТЬСЯ (kupat's'a) 'bathe' is normally written with the so-called particle –СЯ (–s'a). However, if we switch to its sound shape, we find that the hearer must be prepared to recognize, in addition to /kupal-s'a/ 'bathed', also /kupal'i-s'/ '[they] bathed', and /kupac-ca/ '[to] bathe'. In many cases, the phonology-based representation reshapes the paradigm as compared with its writing-based version, cf. НОВОЕ Neuter 'new' and НОВАЯ 'new' Feminine which just merge in /novaja/.

Methods

As our goal is to "redress" Russian morphology in such a way that its expression plane would be consonant with the phonology, our first step is to provide all the (nominal) wordforms of the Russian lexicon with a phonological transcription. E.g., ОДЕЯЛ-О 'blanket' → /ad'ijál-a/. To make sure that our phonological transcription faithfully reproduces the expression plane of the wordforms chosen, our Ss. were asked to filter out the output of the transcribing routine.

The second step is developing a database for nouns where all the relevant information about individual nouns would be stored (see Kasevich *et al.*, this volume).

Results and discussion

The results of reshaping Russian nominal wordforms along phonological lines make it possible to see the morphological structure of Russian 'as it is', with a "distorting" influence of the traditional writing totally eliminated.

For instance, it is a well known fact that in many, if not all, languages where morphological component is sufficiently developed, the paradigms include at least two homomorphic inflections, cf. DOM 'house', Nominative and DOM 'house' Accusative. (One could add, rather parenthetically, that if such pairs would be the only means to express given meanings, there would be every reason to classify Russian with Ergative languages.) When we base our analysis on spoken (phonologically represented) wordforms, two more homomorphic forms should be added to the DOM-paradigm, viz. /dom-i/ 'house' Genitive and /dom-i/ 'house' Locative. Using our database, one can easily trace all the types of paradigm reduction due to the spoken-form orientated approach. What is more important, in this way we can try to bring to light the regularities that underlie the functioning of the grammar. For instance, we can see that Neuter is a 'weak' point of the paradigm it enters, as it tends to merge with Feminine (cf. above).

We are not going to claim that the traditional writing based grammars are just "cultural artifacts" with no prototype in the real world. However, we do claim that spoken language should be given priority, if one sets an ambitious goal of looking into inner mechanism of language. That would be consonant with the insights from linguists like Jan Baudouin de Courtenay, Lev Scherba and Charles Hockett who insisted on an absolute necessity to discriminate between differently aimed grammars.

A typological note would be appropriate. For quite a few languages, the problems discussed in this paper are simply irrelevant, because the languages are pre-literate. As a matter of fact, compiling special Russian grammars intended for the hearers treats the Russian language as if it were pre-literate.

Another situation is met where there is a wide gap between writing and sound systems. If we compare, say, Russian and English, we will see that the Russian writing system is relatively simple and systematic, while the English system is notorious for its very unsystematic, sometime extravagant, relationship between writing and sound. This means that the analyst will be confronted with very different tasks depending on the language.

It is also interesting to study the sound-writing relation from the point of view of how writing reflects diachronic shifts. For Russian, it could be hypothesized that, at least in some cases, the reduction phenomena described above synchronically recapitulate diachronically important development (like Weak Vowel Drop, etc.).

Finally, a few more words about our problem from the applied linguistics perspective could be added. Stripping the wordform of its writing 'dress' is not the end of the story, although it is surely a prerequisite to writing computer programs for automatic speech perception and speech production. A phonologically transcribed speech, especially when it is a piece of the fluent text, is still very far from the real acoustic speech signal with all its redundancy on the one hand and imperfections and missing portions on the other. It is quite typical to be exposed to a speech signal so impoverished that only a good deal of guesswork makes an adequate perception possible.

There is one more very important problem that cannot be neglected, given the goal of our study. We mean the prosodic (here accentual) characteristics which are indispensable for any wordform of Russian. It has been demonstrated in lots of experiments that the lexical stress (accent) is an indepedent parameter in speech perception. According to our findings, quite typical is the situation where accent recognition scores are much higher than those for the phonemes or syllables. It is much likely that the overall language system contains a separate, relatively independent prosodic subsystem. This subsystem comes into play first in speech perception and in language acquisition, too, the stress strategies are well developed even prior to all the other subsystems.

Here again, typological aspects are also essential. To begin with, there exist languages, like Mongolian, where they have no lexical accent (stress) at all (vowel harmony being a partial functional substitute). No statistics are available, but it seems safe to argue that the number of unaccentual (lacking lexical stress) languages are much less. However, if we turn to standard written texts, where no accents are shown, we will see that the two language types discussed above (with and without stress) become very much closer. Within one language as well as cross-linguistically, various subsystems and compensatory strategies are used to achieve an approximately the same level of efficiency both in perception and production, writing being one of the factors in play.

Writing to some extent makes obscure the real number of the homonyms to be found in the language. According to our data, in Russian one finds more than four thousand words which are written the same but differs due to different positions of the stressed syllable, e.g. L'UBIM ~ L'UBIM' '[we] love ~ '[he is] loved'. These are, so to speak, writing-made homonyms although 'in reality' they are a clear case of minimal pairs.

In some cases, the writing/spoken dichotomy may determine the very deep typological features making the language typologically the way it is.

According to a witty observation of Professor EugenyJakhontov, Semitic languages are typologically close to the isolating class when the languages are written, but acquire most features of inflexional languages when the languages are spoken, The thing is that in Semitic languages the so-called schemata whose function is to express grammatical meanings are not "visible" when written, that is KiTaB 'book' and uKTub 'write' where KTB is a root, i-a and u-u schemata, are reduced to writing in the same way.

Of cause, it is a comforting idea to believe in the unique grammar for each language, our duty being to discover it. In reality, the situation is much more complicated and the written word/spoken word dichotomy adds a lot to its complexity.

References

Baudouin de Courtenay J.A. 1912. On the Relation of Russian Writing to the Russian Language. In Baudouin de Courtenay J.A. Selected Works on General Linguistics. Vol. 2, 209-235. Moscow

Shcherba L.V. 1957. Baudouin de Courtenay and His Contributions to Linguistic Studies. In Shcherba L.V. Selected Papers on the Russian Language, 85-96. Moscow

Hockett Charles F.1961. Grammar for the hearers. In Structure of language and its mathematical aspects. In Proceedings of symposia in applied mathematics, vol. 12, 220-236.

On the buildup of an integrated database for the formal description of grammars for the hearers

Vadim Kasevich[1], Iuliia Menshikova[2], Maria Khokhlova[2], Elena Shuvalova[2], Anna Lastochkina[3]

[1]Faculty of Asian and African Studies, SpbU, Russia
[2]Faculty of Philology, SpbU, Russia
[3]Faculty of Liberal Arts and Sciences, SpbU, Russia

Abstract

Grammars for the hearers often significantly differ from those for the readers as traditional orthographic notation of wordforms is unable to fully represent the actual expression of the morphological categories and, consequently, the real composition of the paradigms. As a first step for the construction of a grammar for the hearers, one needs a database containing the information on the spoken (phonological) expression of the morphological units. At present the part of the database with the information on Russian nouns is completed. The subjects in the database are Russian noun forms of different declensions and accent paradigms expressing all the types of the stem endings that are able to shape the actual spoken realization of a form.

Key words: linguistics, Russian language, grammar, phonetics, morphemics.

Introduction

The idea of the project is based on two articles published in the 1970s: L.V. Bondarko, L.A. Verbitskaya "On Phonetic Characteristics of Post-tonic Vowels in the Modern Russian Language" and L.V. Bondarko, L.A. Verbitskaya, M.V. Gordina, L.R. Zinder, V.B. Kasevich "Styles of Pronunciation and Types of Pronouncing". The experiments on which these publications were based showed, in particular, that native speakers do not distinguish "by ear" such word forms as, for example, новая, новое: they merge into новая. And it is not a singularity, because such "merges" are found in many different segments of the system of the modern Russian language.

Baudouin de Courtenay was first to call the problem of describing the grammar of a language on the basis of oral (primary) speech one of central fundamental problems of descriptive grammar in particular and of theoretical linguistics in general. However, more than a century after the publication of Baudouin's works this problem remains unsolved. It explains the academic novelty of this project. For a long time solving this problem was considered problematic, because it required having developed and application-proven phonological and grammatical theories. Present-day linguistics in Russia has all the prerequisites for a systematic description of the grammatical structure

of the contemporary Russian language on the basis of its oral form, and the problem of creating this description is of great current interest.

The authors are not aware of any Russian or foreign research teams that would work on the problems raised in this paper. At the beginning of the XXth century there existed an international scholarly journal LE MAITRE PHONETIQUE, where all publications were printed in phonetic transcription. However, it was a purely empirical project the aim of which was to popularize the usage of transcription.

Methodology

The specific problems that are to be solved within the project are the development of two basic problematic areas. The first one is the creation of databases that would reflect changes in inflectional paradigms of Russian words that depend on their sound/orthographic codes. The second one is to reveal shifts in the system of Russian morphosyntax caused by this recoding. Using the projected databases will allow effectively establishing basic trajectories of changes in paradigms after the change of the code (modality) of the plane of expression of linguistic units. In order to solve the formulated problem we use methods of classical structural linguistics with its focus on revealing formal paradigms that consist if oppositive word forms; categorical analysis; neutralization of oppositions in specific contexts etc. The formal paradigms that are analyzed are seen as semantisized structures, where the plane of expression and the plane of content are inseparable, and shifts in semantics normally correlate with shifts in the content plane, and vice versa. Considerable attention is given to the exploration (both theoretical and experimental) of the category of neutralization in its complex relationship with the category of homonymy.

The expected general outcome of the methods and approaches briefly described above is a model that would allow tracing all the changes of the language system that it undergoes in the transition from orthographically oriented to phonologically oriented representation.

Results and Discussion

Within the framework of the project we have created a prototype of the database filled with word forms of different parts of speech that allows tracing consistent patterns in the reduction of paradigms caused by transition from orthographic to phonological code. Working on the database will allow determining trajectories connecting “orthographic” and “phonological” word forms and, consequently, correlate the grammar of the speaker and the grammar of the hearer.

The first stage of the project is data collection and presentation of data in the frame of the existing database. It will be build “around” separate inflected parts of speech (nouns, adjectives, numerals, pronouns and verbs). At the same time, we are going to use the results of database processing to prepare material for perceptive experiments.

The results of the project are to be on open access, so choosing the data format was an important decision. We have selected the XML format as the most universal and well adapted to future conversion for the developing database. Below is an example of a fragment of XML representation of the lexical item «окно».

```
<entry id="n50" author="yum" time="2016-05-28">
<word>окно</word>
<orth>окно</orth>
<grammar>1d*</grammar>
<accent>B</accent>
<url>http://ru.wiktionary.org/wiki/окно</url>
</entry>
```

We have chosen the platform Microsoft SQL Server for database maintenance because of its reliability, scalability and productivity.

The chosen database format is based on client-server architecture; the server side provides most functionality while the client presents a graphic interface for the users. The client applications contact the server via the standard HTTP protocol. The server part is build up from small parts called servlets that allow for the composition of all servers from modules. Each servlet provides functionality, e.g. the database access, search, morphological analysis and connection to various corpora if required. Thanks to different commands it is possible to receive various results corresponding to queries (including combined queries). For example:

- a paradigm member or the initial form in orthography;
- a paradigm member or the initial form in phonological transcription;
- a grammatical characteristic on one morphological category;
- a grammatical characteristic on a given set of morphological categories;
- information on the homonymy of inflectional elements in orthography;
- information on the homonymy of inflectional elements in phonological transcription;
- information on the allomorphism of inflectional elements in orthography;
- information on the homonymy of inflectional elements in phonological transcription etc.

In the database there is search with wildcards support (of the language of regular expressions), so it is possible to search for parts of words or expressions. At present we are working on creating algorithms of data processing for the database of the selected type on the basis of a completely filled fragment of the nouns database. The objects of the database are the word forms that represent Russian nouns of different types of declensions and accent paradigms and demonstrate all the types of stem endings that can influence the phonetic image of the word form. The fields of the database contain information about the orthographic and phonetic image of a word form, about all of its morphological characteristics, variability of morphological forms and accent patterns, inflection indexes and accent types. Different fields contain the orthographic and phonetic images of stems and inflectional affixes included in each word form.

References

Bondarko L.V., Verbitskaya L.A. 1973. On Phonetic Characteristics of Post-Tonic Inflexions in the Contemporary Russian Language. In Problems of Linguistics, No 1, 37-49.

Bondarko L.V., Verbitskaya L.A., Gordina M.V., Zinder L.R., Kasevich V.B. 1974. Styles of Pronunciation and Types of Pronouncing. In Problems of Linguistics, No 2, 64-70.

How to write an oral dialect or about some problems of the Tsakonian Corpus

Maxim Kisilier
Hellenic Institute, Saint-Petersburg State University; Department of Comparative and Areal Linguistics, Institute for Linguistic Studies (RAS), Russia

Abstract

Hellenic Institute of the Saint-Petersburg State University in collaboration with the Institute for Linguistic Studies of the Russian Academy of Sciences organized more than twenty expeditions to South Kynouria in Peloponnese (Greece) in order to describe the Tsakonian dialect. During these expeditions its participants collected a large number of oral texts in Tsakonian and it was decided to create a Tsakonian corpus so that this very interesting linguistic material could be easily accessed. This paper provides the first description of the project and discusses its current problems.

Key words: Modern Greek dialectology, Tsakonian, language corpus.

Introductory remarks

Modern Greek dialectology has a rather long history. Many institutions in- or outside Greece possess large collections of Modern Greek dialect materials from various Greek speaking regions. Unfortunately the major part of them remains unknown and unused not only by typologists, but even by specialists in Modern Greek dialectology. Short dialect texts from these collections are sometimes published as supplements to linguistic papers (cf.: Kisilier 2009: 406–411; 2014: 342–344), but they can hardly be used for serious linguistic analysis as they provide just a general idea of the dialect and may lack some very important features. More often certain samples from these collections appear in linguistic articles to illustrate a statement of the author.

However when the statement is false, the reader may be led to incorrect interpretations of the example or even to erroneous conclusions in general since he has no opportunity to check this example or statement. Thus Russian linguist Mikhail Sergievskiy who was the first to describe the verb system of Azov Greek found perfect forms in this dialect (Sergievskiy 1934: 582–583). So Azov Greek could be grouped together with other few Modern Greek dialects that have perfect/pluperfect along with aorist. All other descriptions of Azov Greek never mention perfect forms, while the analysis of the modern state of the verb in the dialect based on recently collected data doesn't let to discover any trace of perfect forms or any appropriate place for them within the verb system (Kisilier 2009: 193–205). This ambiguous situation can be easily explained. Sergiyevsky found perfect forms in the

poems by Georgy Kostoprav who tried to create a special language for Azov Greek literature based both on local idioms and on some Demotic features that in fact did/do not exist in the dialect like perfect forms (Kisilier 2009: 13–14).

The progress of modern technologies gives hope that one day there will be no need to look for dialect examples in books and articles, but in text corpora. Nowadays there is still no open access corpus of any Modern Greek dialect that can be really helpful for linguistic research (cf.: http://griko.project.uoi.gr/), but many attempts in this direction are already made. In this paper I am going describe briefly the project of Tsakonian corpus and some problems I had to face.

About Tsakonian dialect and Tsakonian project

Tsakonian is one of Modern Greek dialects of Peloponnese. It is generally believed that Tsakonian can be traced back directly to the Ancient Doric Laconian. Different sources provide different number of speakers — from 200 in (Salminen 2007: 271–272) up to 8000 in (Kontosopoulos 2001: 3). Since 2008 Hellenic Institute of the Saint-Petersburg State University in collaboration with the Institute for Linguistic Studies of the Russian Academy of Sciences (http://iling.spb.ru/index.html?language=en) organized more than 20 expeditions to the Tsakonian speaking area and now disposes approximately 250 hours of audio and 30 hours of video recordings and 147 linguistic and ethnographic questionnaires (Kisilier 2014). The most interesting texts were transcribed and supplied with detailed interlinear morpheme-by-morpheme glossing that takes into account all inflectional peculiarities.

One of the goals of these expeditions was to collect lexical data using the questionnaire of "Minor dialectological Atlas of Balkan languages (Domosiletskaya et al. 1997). At the present stage the words are put into Field Linguist's Toolbox together with subdialectal variants and grammatical forms, for example:

1. **mountain:** *ʃína* masc. [Vaskina, Kastanitsa], *sína* masc. [Melana]; pl. *ʃínu*; genitive *tu ʃínu — korfá tu ʃínu* 'top of the mountain'.
2. **to bite:** *kat͡sʲínu* [Vaskina, Melana, Tyros], *gat͡sʲínu* [Prastos], *tat͡sʲínu* [Kastanitsa]; fem. *kat͡sʲína*; neutr. *kat͡sʲínda*; aorist 1Sg *ekat͡sʲíka*/*ekat͡sʲía*, 3Sg. *ekat͡sʲítse*/*ekat͡sʲíe*; perfect participle *kat͡sʲitˣé*; subjunctive imperfective 1Sg. *na=kat͡sʲínu*, 3Sg. *na=kat͡sʲíni*, 3Pl. *na=kat͡sʲínoi*; subjunctive perfective 1Sg. *na=kat͡sʲíu*, 3Sg. *na=kat͡sʲí*, 3Pl. *na=kat͡sʲínoi*.

Local community, especially the Tsakonian Archives (http://www.tsakonianarchives.gr/) always tried to help us and demonstrated a deep interest in our activities and results. It became evident that they should have access to our materials and they ought to be able to use them in their attempt to provoke interest towards Tsakonian in the younger generation. Some years before we started our activities in Tsakonia Demotic School of Leonidio (municipal center of the region) created online dictionary of Tsakonian (http://dim-leonid.ark.sch.gr/?page_id=28). They used the "Dictionary" by Michael Deffner (1923) and asked pupils to go their grandparents in order to have some words translated from Modern Greek into Tsakonian. This online dictionary is, certainly, very small (only 4635 entries) and inconsistent because the quality of the data collected by pupils is very uneven. However this attempt demonstrated that Tsakonian youngsters are ready to deal with Tsakonian, especially if they see its application and if they realize that Tsakonian is not so old-fashioned as they used to believe and it can be accessed by means of the modern technologies.

About Tsakonian corpus

When it was decided to create a Tsakonian corpus, it became evident that corpus must be available for both linguists and local community. This decision has its advantages and disadvantages. Collaboration with the Tsakonian Archives, on the one hand, makes it possible to incorporate some already published dialect texts together with the recently collected ones. So the corpus will become more diachronically representative and rich in its vocabulary. On the other hand, it presupposes that non-linguists should be able to read Tsakonian dialect and we cannot content just with IPA.

Famous specialist in Tsakonian Thanasis Costakis has invented Tsakonian alphabet based on Greek graphics. The varieties of this alphabet are widely used in local editions of dialect texts. Despite certain inconsistencies, Costakis alphabet can be easily transformed into IPA, but it is totally inapplicable for corpus because it makes use of particular diacritics that is absent from standard fonts. This problem can be somehow solved for printed editions or by means of uploaded fonts or virtual keyboards. However I am sure that most local users are going to visit the site from their mobile devices and that is why a different alphabet is required.

On October, 30th 2015 I introduced a new alphabet to local community. I thought that it was to be very easy and it had to avoid any possible ambiguity. The user should not depend on his knowledge of Ancient Greek and choose among ι, η, ει, οι and υ as he looks for a word with /i/ in it. Actually I followed the example of Russian linguists who created Greek-based alphabet for Georgian Pontic and Azov Greek (cf.: Kisilier 2009: 11–12). Thus the new Tsakonian was supposed to get rid of traditional digraphs

in vowel system (/i/ is expressed only by means of ι, ω is not used at all, υ is /u/). Tsakonian has a number of peculiar consonants (cf.: Haralampopoulos 1980: 26–83), and I decided to adopt some clusters for them: γκ /g/ (vs γ /ɣ/), ζζ /ʒ/, πχ /p^h/, σσ /ʃ/ (as the dialect has no geminates), τζ /d͡z/, τσ /t͡s/, τσσ /t͡ʃ/, and τχ /t^h/. Sometimes it is important to indicate palatal consonants (Kisilier, Fedchenko 2011) and I proposed to introduce η as palatal index: λη /l^j/, νη /n^j/, ρη /r^j/, τση /t^j/ etc.

However local intellectuals did not approve of the new alphabet because it is totally different from the one by Costakis which they regard as an important ground of the modern Tsakonian culture. And this stalemate is still not resolved.

Acknowledgements

This research was supported by the Russian Science Foundation (project No 15-18-00062) and the Russian Foundation for Humanities (project No 14-04-00581).

References

Salminen, T. 2007. Europe and North Asia (Ch. 3) In Moseley, C. (ed.) 2007. Encyclopedia of the world's endangered languages, 211–280. London; New York, Routledge.

Deffner, M. 1923. Lexicon tis Tsakonikis dialektou. Athens, Typography “Estia”; Meissner & N. Karagadouris.

Kontosopoulos, N. G. 2001. Dialektoi kai idiomata tis neas ellinikis. Athens, Grigori Publishers.

Haralampopoulos, A. L. 1980. Fonologiki analysi tis Tsakonikis dialektou. Doctoral dissertation. Arostotle University of Thessaloniki, Faculty of Philosophy. Thessaloniki

Domosiletskaya, M. V., Zhugra, A. V., Klepikova, G. P. 1997. Malyy dialektologicheskiy atlas balkanskikh yazykov. Lexical questionnaire. St. Petersburg, Institute for Linguistic Studies.

Kisilier, M. L. 2014. Tsakonskiy dialekt: novyy vzglyad In Vydrin, V. F., Kuznetsova, N. V. (eds.) 2014. Ot Bikina do Bambalyumy, iz varyag v greki. Ekspeditsionnye etyudy v chest’ Eleny Vsevolodovny Perekhval’skoy, 330–348. St. Petersburg, Nestor-Istoria.

Kisilier, M. L. (ed.) 2009. Lingvisticheskaya i etnokul’turnaya situatsiya v grecheskikh selakh Priazov’ya. Po materialam ekspeditsiy 2001–2004 g. St. Petersburg, Aleteya.

Kisilier, M.. L., Fedchenko, V. V. 2011. K voprosu o myagkikh soglasnykh v tsakonskom dialekte novogrecheskogo yazyka. Indo-European Linguistics and Classical Philology Yearbook XV, 259–266.

Sergievskiy, M. V. 1934. Mariupol'skie grecheskie govory. Opyt kratkoy kharakteristiki. Izvestiya AN SSSR. Otdelenie obshchestvennykh nauk 7, 533–587.

Some aspects of /r/ articulation in French Vocal Speech

Ulyana Kochetkova
Department of Philology, Saint-Petersburg State University, Russia

Abstract

This study analyses some common and individual strategies in choosing /r/-variants in French vocal speech. The problem of the /r/ pronunciation is approached from a new side by considering deviations from singers' main /r/ articulation model. The following analysis has been done: examination of the individual and common preferences of 2 different generations of singers in /r/ articulation in French lyric songs and operatic arias; study of deviations frequency in different phonetic contexts in its relation to musical phrase boundaries.

Key words: singing, French, phonetic-phonological analysis, pronunciation models.

Introduction

The question of the French /r/ articulation is today one of the most discussed subjects both due to its variety in the contemporary standard French and to the existing differences of views on pronouncing this consonant on stage in Opera as well as in Art Songs (Melodies). There is no absolute agreement among singers, singing teachers, accompanists and coaches about which variant is preferable: the "Italianate" apical alveolar trill or flap suggested from 17th century onward as the only correct pronunciation in singing (Bacilly 1679, Garcia 1851, Duval 1878, Lavoix, Lemaire 1881, Grubb 1979, Yarbrough 1991); or the conversational uvular consonant, the latter having been criticized for its "vulgarity" and destroying effect that it produces on surrounding vowels and airflow projection in general (Nedecky 2015) or recommended only to French native singers (Vennard 1967). However, the uvular consonant is consistently observed not only in some famous modern French singers' performances (Nedecky 2015),but can also be found (though seldom) even in the interpretations by renowned artists of the past, who themselves crucially criticized it.

Today most of non-French contemporary singers face certain problems and difficulties when performing an opera or lyric song written by a French composer, for it is one of the most complicated languages for a non-native speaker to sing. The modern performing art standards are high, but there is a lack of panoramic theoretical and experimental works in this field, so that the current study will be a contribution to it.

Methods and material

At the first stage, in order to provide data for the identification of the preferences in /r/ articulation by contemporary French singers in comparison with their previous centuries colleagues, 87 French lyric songs, 36 operatic arias and 3 whole operas in stage production were analysed. The material includes recordings of 25 French singers (12 male and 13 female voices), who can be divided in 2 groups: 14 singers born after 1950 and 11 singers born before 1950. However the above mentioned works were not interpreted by every of the 25 singers, because the diversity and the heterogeneity of the material caused a range of problems: 1) some old recordings have a quality which is not sufficient for the appropriate analysis; 2) several recent live or broadcast recordings may contain some noise; 3) the repertoire of different singers is specific and restricted to a concrete style of music, depending on the singer's voice and background.

The singer's preferred model in two different genres (Opera vs. Art Songs), as well as in two different styles (French Lyric Opera vs. French Baroque Opera) was established in the following way: singers were evaluated as [ʁ]or [r] preferring, if they used a certain model in more than 50% of analysed works.

Although the use of the uvular [ʁ] is commonly regarded as a recent trend, it was observed that 5 contemporary singers choose alveolar [r] in both styles and both genres, and 3 other contemporary singers of the same group performed baroque opera using only the alveolar variant of /r/, even though they articulate often or mostly the uvular consonant in the Romantic opera. On the other hand performances of the previous generations of singers contained the uvular consonant (even in Baroque music). It allowed us to suppose that the Baroque educational background developing leads to a more elaborate and conscious way of working on the articulatory aspect.

At the second stage of this study the relation of phonetic contexts with the deviations from the singer's preferred model was examined in 58 Art Songs, in which at least one deviation occurred (songs with no deviations, as well as operatic arias were excluded from the material in order to avoid other factors' potential impact).From 1 to 10 interpretations of each piece were considered. This part of material included performances of ten contemporary and ten early 20th century singers; five of them preferred [ʁ]-model (only contemporary performers),fifteen singers preferred [r]-model.

Results

In the studied material 15 different types of phonetic contexts were defined, presenting 5 main groups: 1) intervocalic – VRV ("horizons"); 2) musical phrase initial position – RV ("reviens"), RCV ("roi"), CRV ("cri"), CRCV

("trois"); 3) final position before breath pause – VR ("Lahor"); 4) pre- or postconsonantal before or after 1 consonant – VRC ("courtes"), VCRV ("tendre"), VRCV ("en ruines"); 5) interconsonantal in different types of clusters – VCCRV ("ellecrie"), VRCCV ("lorsque") including clusters with semivowels – VCRCV ("endroit"), VCCRCV ("ellecroit") or another "r" – VRRV ("coeurregrette"), VRCRV ("arbre"). Total amount of contexts with /r/ in the analysed vocal texts was 1814. The following types of phonetic contexts were the most frequent ones: "VRV" (34%), "VRC" (32%), "VCRV" (19%).

Deviations from different /r/ articulation models were considered separately for singers preferring alveolar articulation and for those preferring the uvular one. In order to normalize data, deviation in each concrete context was counted as 1, if it occurred at least in one singer's interpretation. Then, the percentage of the deviations in each type of phonetic context to the total amount of these phonetic contexts in the material was obtained.

Figure 1 represents the normalized frequency of the deviations from the individual's main alveolar [r] articulation model in different types of phonetic contexts. Deviations occurred more frequently in the following phonetic contexts: before breath pause (VR, 27%), in the consonant cluster with another /r/ (VRCRV, 26%), in intervocalic position (VRV, 23%), in the initial position in musical phrase after one consonant (CRV, 20%). Frequency of other types of contexts is less than 20%.

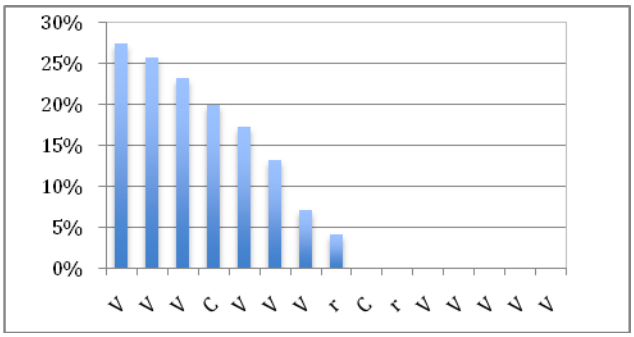


Figure 1. Frequency of deviations from [r]-model in different types of phonetic contexts.

As it was mentioned above, at this stage of analysis the French uvular consonant was observed as the main /r/-articulation model only for 5 singers in the studied material (see table 1); four of them had deviations in their performances. As long as deviations from the uvular articulation occurred only in 27 cases, it is impossible to make a reliable comparison with the results obtained for [r]-model. But it is an interesting fact that the deviations

from [ʁ]-articulation model occurred in the contexts, which were considered as favorable for the deviations from the model with the alveolar consonant.

Table 1. Number of deviations from [ʁ]-model in different types of phonetic contexts.

Phonetic contexts	Singer1	Singer 2	Singer 3	Singer 4	Total number
VR	1	2	0	0	3
VRC	5	7	0	2	14
VRV	0	5	1	0	6
VCRV	0	2	0	0	2
VRCRV	0	0	0	2	2

Conclusion

This study allowed to make the following observations: 1)different /r/-articulation preferences in singing exist in both singers' age groups; 2) deviations from two different models are possible in both groups; 3) some contemporary performers never choose the model with the uvular consonant (even in lyric songs); 4) some of the contemporary singers use different /r/-articulation main models in different styles (Romantic vs. Baroque); 5) some phonetic contexts, as well as the initial or final position in a musical phrase may influence the occurrence of deviations from the chosen /r/-articulation model.

References

Bacilly, B (de). 1679. L'art de bien chanter de M.de Bacilly. Paris, Bacilly.

Garcia, M. 1851.Ecole de Garcia: traité complet de l'art du chant. Paris, Mayence.

Grubb, Th. 1979. Singing in French: a Manual of French Diction and French Vocal Repertoire. Belmont, Schirmer Books.

Duval, G. 1878. Artistes etcabotins. Paris, Ollendorf.

Lavoix, H., Lemaire, Th. 1881. Le chant. Sesprincipeset son histoire. Paris, Heugeletfils.

Nedecky, J. 2015. French Diction for Singers. A Handbook of Pronunciation for French Opera and Melodie. Toronto, Book POD.

Vennard, W. 1967. Singing. The Mecanism and the Technic. New York, Karl Fischer.

Yarbrough, J. 1991. Modern Languages for Musicians. Stuyvesan, Pendragon Press.

Different acoustic cues for emphasis in teaching English word stress to Hong Kong Cantonese ESL learners of different proficiencies

Wience Wing Sze Lai[1,2], Manwa Lawrence Ng[2]
[1]Hong Kong Community College, The Hong Kong Polytechnic University, Hong Kong
[2]Speech Science Laboratory, Division of Speech and Hearing Sciences, The University of Hong Kong, Hong Kong

Abstract

The present study examined English word stress produced by twenty-two (11 highly proficient and 11 less proficient) native adult speakers of Hong Kong Cantonese (CS) learning English as a second language (ESL), in comparison with that produced by five native English speakers (NS). All participants read four English donor words, and CS also read the corresponding Cantonese loanwords. The three acoustic cues for stress, namely pitch (F0), duration (length) and intensity (loudness) values of the vowels were obtained from all syllables. While vowel duration was found to be the dominant cue, followed by F0, in distinguishing stressed and unstressed syllables in all speakers' production, HCS may have overused F0 and LCS may have underused vowel duration.

Key words: English word stress, Cantonese loanwords, acoustic cues, speaker proficiency

Introduction

To Cantonese speakers (CS) who have been using English as a second language (ESL), English word stress could be a challenge, because Cantonese, as a tone language, makes use of pitch to distinguish lexical meanings while English, as a stress language, makes use of not only pitch (fundamental frequency, F0) but also intensity (loudness) and duration (length). With regard to Cantonese speakers' English word stress acquisition, previous studies investigated either (1) Cantonese loanwords borrowed from English (Lai, 2004; Lai, Wang, Yan, Chan, & Zhang, 2011; Silverman, 1992; and Zhang, 1986) or (2) CS's pronunciation of English words (Chan, 2007; Lai & Ng, 2014a; 2014b; and Luke, 2000).

All studies in (1) agreed that loanword syllables corresponding to stressed ones in English were assigned a high level (55) tone. Epenthetic loanword syllables were assigned a low-mid (22) tone (Lai, 2004; Zhang, 1986), but loanword syllables corresponding to unstressed ones assigned a mid (33) (Zhang, 1986) or low-mid (22) tone (Lai, 2004; Lai, et al., 2011).

With regard to (2), while Chan (2007) found that CS could effectively represent word stress by manipulating duration, intensity and F0, Lai and Ng (2014a; 2014b) identified F0, rather than duration and intensity, as the dominant cue for producing stress in HCS and LCS. Luke (2000) reported stressed syllables as being assigned an H tone and unstressed ones an M or L tone.

As revised from Lai and Ng (2014a), which compared only HCS and LCS (excluding NS) and measured parameters by segmenting syllables instead of vowels, this study examines CS's production of English word stress in English donor words and corresponding Cantonese loanwords by identifying the most dominant acoustic cue, among pitch, intensity and duration of the vowels, for HCS and LCS, when compared with NS.

Methodology

Twenty-two Cantonese ESL speakers (F=11; M=11), aged 18-24, were recruited as target participants, known as CS. All CS were born in Hong Kong and had lived there since birth. Among them, 11 were highly proficient in English (with a grade "C" in HKALE UE or a grade "5" in HKDSE English, equivalent to an IELTS score of 6.51, or above), and 11 were less proficient (with a grade "E" in HKALE UE or a grade "3" in HKDSE English, equivalent to an IELTS score of 6.02, or below) (Hong Kong Examination Authority, 2004; 2010). All CS were recruited from the Hong Kong Community College (HKCC), The Hong Kong Polytechnic University (PolyU) community. Five native speakers (F=2; M=3) of British English were recruited as controls, known as NS. They were all residents of the United Kingdom. All participants had normal hearing, speech and language ability by self-report.

All participants were instructed to read four English donor words (sauna /ˈsɔːnə/, guitar /gɪˈtɑː/, carnivals /ˈkɑːnɪvəlz/ and vanilla /vəˈnɪlə/), and CS also the corresponding Cantonese loanwords (桑拿 /sɔŋ55 naː21/, 結他 /kit33 tʰaː55/, 嘉年華 /kaː55 nin21 waː21/ and 吔哩拿 /wɐn22 nei55 laː35/). The speech samples were recorded using AUDACITY in a quiet room with a high-quality unidirectional dynamic microphone fixed at 10 cm from each participant's mouth for consistency.

The recording of each participant was first processed using Praat (Boersma & Weenink, 2010). Each syllable in the pronounced English donor words and Cantonese loanwords was extracted and stored. The extracted syllables of both the English donor words and Cantonese loanwords were then classified into two types, (1) stressed syllables or those corresponding to stressed syllables in the English donor words, and (2) unstressed syllables or those corresponding to unstressed syllables in the English donor words.

The vowels were segmented manually by one of the authors, with ten percent repeated for intra-judge reliability measure, regarded as satisfactory with the Spearman's correlation coefficient between the duration of segmented vowels as 0.997 ($p < 0.001$). Three acoustic parameters: average fundamental frequency (F0) (in Hz), duration (in ms), and average intensity (in dB) of the vowel were measured from each sound sample.

Results

Concerning the production of the English donor words, vowel duration (instead of F0 in CS as identified previously) was found to be the dominant cue in distinguishing stressed and unstressed syllables in both NS and CS. However, HCS (with a difference of 32% between stressed and unstressed English syllables) appeared to be more similar to NS (with a difference of 51%) in relying on vowel duration when compared with LCS (with a difference of only 15%). While F0 was the next dominant cue for both NS and CS, HCS (with a difference of 20%) relied on F0 more than both NS and LCS (with a difference of 13% and 10% respectively) did.

Since Cantonese makes use of tones but not stress to contrast meanings, Cantonese loanword syllables corresponding to stressed and unstressed English syllables are supposed to differ only in F0 but not in intensity and vowel duration. Surprisingly, vowel duration was still the dominant cue, followed by F0 and intensity, in both HCS and LCS's production. Despite this, the small difference of only 2% in HCS in the use of F0 in distinguishing the (originally) stressed and unstressed syllables in the English donor words and Cantonese loanwords and the marked difference of 28% in LCS in the use of vowel duration in distinguishing them further confirm HCS's overuse of F0 and LCS's underuse of vowel duration in realising English word stress.

Conclusion

In short, unlike previous findings, vowel duration was found to be the dominant cue, followed by F0, in distinguishing stressed and unstressed syllables in all speakers' production. Also, HCS may have overused F0 and LCS may have underused vowel duration. This implies the need for different approaches in teaching English words stress, with less emphasis on F0 for HCS, and more emphasis on vowel duration for LCS.

Acknowledgements

The work described in this paper was substantially supported by a grant from the College of Professional and Continuing Education, an affiliate of The Hong Kong

Polytechnic University, jointly undertaken with Division of Speech and Hearing Sciences, The University of Hong Kong.

References

Boersma, P., & Weenink, D. 2010. Praat: doing phonetics by computer. Retrieved July 20, 2011, from http://www.fon.hum.uva.nl/praat/

Chan, M. K. K. 2007. The Perception and Production of Lexical Stress by Cantonese Speakers of English. M.Phil Dissertation. Hong Kong: The University of Hong Kong.

Lai, W. 2004. Tone-stress Interaction: A study of English Loanwords in Cantonese. M.Phil Dissertation, The Chinese University of Hong Kong, Hong Kong.

Lai, W. W. S., & Ng, M. L. 2014a. English Donor Words and Equivalent Cantonese Loanwords Pronounced by Hong Kong Cantonese ESL Learners - Implications for Teaching English Word Stress. Proceedings of International Teacher Education Conference, Dubai (pp.19-28). Dubai: Ankara University.

Lai, W. W. S., & Ng, M. L. 2014b. The Use of Acoustics-based Teaching Software in Hong Kong Cantonese ESL Speakers' Learning of English Word Stress Production. Proceedings of the 6th Annual International Conference on Education and New Learning Technologies, Barcelona (pp. 5773-5781). Barcelona: IATED.

Lai, W. W., Wang, D., Yan, N., Chan, V., & Zhang, L. 2011. Influence of English Donor Word Stress on Tonal Assignment in Cantonese Loanwords - An Acoustic Account. In W. Lee & E. Zee (Eds.), Proceedings of the 17th International Congress of Phonetic Sciences (pp. 1162-1165). Hong Kong: City University of Hong Kong.

Luke, K. K. 2000. Phonological Re-interpretation: The Assignment of Cantonese Tones to English Words. ICCL-9 Conference Paper. Singapore: National University of Singapore.

Hong Kong Examination Authority. 2004. IELTS (2004). Retrieved from http://www.hkeaa.edu.hk/en/ir/Standards_of_HKEAA_qualifications/IELTS/

Hong Kong Examination Authority. 2010. Results of the Benchmarking Study between IELTS and HKDSE English Language Examination [Press Release]. Retrieved from http://www.hkeaa.edu.hk/DocLibrary/MainNews/press_20130430_eng.pdf

Silverman, D. 1992. Multiple Scansion in Loanword Phonology: Evidence from Cantonese. Phonology, 9, 289-328.

Zhang, R. 1986. Xianggang Guangzhouhua Yingyu yinyi jieci de shengdiao guilü [= the tonal patterns of English loanwords in Hong Kong Cantonese]. Zhongguo yuwen, 1, 42-50.

Cognitive approach to translation and interpreting teaching methods

Julia Levi
Department of the English, MGIMO University, Russia

Abstract

Nowadays translation/interpreting studies are focused upon human mental processes, cognition, the role of the interpreter/translator. According to the human activity theory, each action is purpose - oriented, thus a complex act of translation/interpreting which can be described as a secondary process of human activity is goal - oriented as well. It means that the act of interpreting/ translation corresponds to the main principles of human activity, has its own purpose and is aimed at achieving the same result as an ordinary act of communication, i.e. a communication effect. We believe, it is critical to start an account of the text for translation purposes by making a deliberate pre-translation text analysis (PTA), which according to most experts, may consist of several activities.

Key words: cognition, the act of interpreting/ translation, pre-translation text analysis

Introduction

A new paradigm of language studies allowed linguists in the late XX – early XXI centuries to consider the language as a dynamic phenomenon, rather than a static product, so nowadays experts in translation/ interpreting studies have become more interested in exploring the basic principles of the process of translation/interpreting, which is characterized by the shift to the study of human mental processes, cognition, the role of the interpreter/translator. At the first stage of the development of translation/interpreting science scholars focused on the analysis and description of some objective laws and rules of transformations. But later a new approach with the focus on the nature of the process of translation/interpreting was put forward, which became possible due to advancement in research in the fields of psycholinguistics, sociolinguistics, cognitive linguistics, anthropology, and etc. The roots of a cognitive approach can be traced back to the ideas of such renowned linguists as F. de Saussure, L. Vigotskyi, L. Sherba, A. N. Leontiev, A. A. Leontiev, and many others. In fact they developed and implemented the strategies of linguistic studies which consider the language as a part of human activity with a human playing the central role in it. According to the human activity theory, each action is purpose - oriented, thus a complex act of translation/interpreting which can be described as a secondary process of

human activity is goal - oriented as well. It means that the act of interpreting/ translation corresponds to the main principles of human activity, has its own purpose and is aimed at achieving the same result as an ordinary act of communication, i.e. a communication effect.

What allows translators/interpreters to achieve the same communication effect, evoke the same feelings and emotions in the target recipient? We believe that a profound comprehension of the original text, successful meaning construction produces a communication effect envisaged by the author of the original text.

Methodology

According to J. Field, central to meaning construction is the distinction between 1) the words on the page or in the ear; 2) the propositional information that a text contains (loosely, its literal meaning); and 3) the enriched and selective interpretation which a reader or listener takes away. In processing a text, a comprehender performs a number of operations. At a sentence level they 1) extract propositional information; 2) make any necessary inferences; 3) enrich the interpretation by applying word knowledge; 3) integrate the new information into their mental representation of the text so far; 4) monitor their comprehension in case of misunderstanding.

At discourse level, they also have 1) to recognize the hierarchical structure of the text; 2) identify patterns of logic which link the parts of the text; 3) determine which parts of the text are important to the speaker/writer or relevant to their own purposes.

Numerous accounts of discourse comprehension which attempt to describe how text information is built into an overall meaning representation have proved to be useful both for scholars and learners. A cognitive approach to text studies help linguists perceive information processing mechanisms better and therefore work out some strategies to secure a full understanding of the text.

Nevertheless, comprehension is one of the stages that the model of translation/interpreting comprises. In fact, the model consists of three stages: comprehension, the act of translation/interpreting, and text production.

At the level of comprehension the translator/interpreter builds the concept of the text. When they perceive the original text in a foreign language, they search for semantic frame equivalents to their knowledge. Charles Fillmore believes that, “meanings are relativized to scenes”. According to him, meanings have an internal structure which is determined relative to a background frame or a scene. What is more, during the text processing a so-called process of anticipation plays an important role as it helps to predict the final unfolding of the text through the explanation of

dynamic sematic frames. It goes without saying that anticipation is critically important in simultaneous interpretation.

At the level of translation/interpreting the translator/interpreter builds dynamic frames in his/her mind on the basis of the original text and relates them to their frame equivalents in the target language. They find prototype equivalents on the basis of prototype semantic frame structures and try to find a solution if they are missing, in which case, they apply a certain strategy to compensate for them.

At the final stage the translator/interpreter produces a text in a foreign language taking into account all its syntactic features.

Successful translation/interpreting requires a profound comprehension of the original text, retrieval of adequate equivalents corresponding to dynamic frames and scenes, and finally a text production in a target language. Therefore, from the perspective of a cognitive approach, the role of *the translator/interpreter* and the *text* remains the major focus of linguists' attention. Many scholars believe that the translator/interpreter performs the role of a reader, analyst, linguist, text creator, editor and, finally critic of it. But the text still needs a thorough examination, especially in terms of a pre-translation analysis in a written activity which helps work out and apply special translation strategies.

Results

Most experts suggest that a pre-translation text analysis (PTA) may consist of several activities: 1) considering factors external to the linguistic text; 2) establishing the style and genre of the text; 3) designating the type of the information represented in the text. The succession of these stages may vary, but all the existing models of PTA illustrate 1) textocentric (linguistic); 2) functional; 3) communicative approaches to this process.

On the basis of the U. Breus and N. Valeeva conceptions of PTA, we present a full PTA, which ensures a better comprehension of the text and a well-balanced approach to the selection of a translation strategy.

1. Identify the type of the text (narration, description, etc.) and its functional style (scientific, publicist, official, colloquial, etc.).
2. Outline the basic communication goal of the author, his/her intention, cultural/situational factors.
3. Specify the primary and secondary functions of the text (to inform, communicate, exert influence), which can be understood through explicit or implicit markers.
4. Outline the context.
5. Define the main topic of the text.
6. Specify the stylistic devices of the author.

7. Identify the examples of cultural dissimilarities, which are evident in the text, and make predictions concerning potential difficulties the translator might confront.
8. According to the type of difficulties choose a variant of translation (find the logical focus of the sentence, generalization, specification, logical development, the shift of focus, etc.)

Conclusion

A cognitive approach has proved to be efficient in translation/interpreting teaching methodology as it explains the cognitive functions of the humans' mind, provides a profound analysis of the translation/interpreting model, through a well elaborated pre-translation analysis (PTA) helps learners apply the right strategy of translation, and thus, master the art of translation/interpreting.

References

Alekseeva, I.S. 2004. Vvedenie v perevodovedenie. – M.: Izd. centr «Akademiya».

Breus, E.V. 2007. Kurs perevoda s anglijskogo yazyka na russkij. Uchebnoe posobie. – M.: Valent.

Field, J. 2004. Psycholionguistics The key concepts, Routedge Taylor & Francis Group: London and New York.

Fillmore, C. 1977. The case for case reopened. In Syntax and Semantics 8: Grammatical Relations, ed. P. Cole, 59 – 81. New York: Academic Press.

Nefedova L.A., Remhe I. N. Kognitivnye osobennosti perevodcheskogo processa. - Chelyabinskij gosudarstvennyj universitet. - S. 64-72

Shvejcer A.D. Teoriya perevoda (status, problemy, aspekty).

Valeeva, N.G. 2010. Teoriya perevoda: kul'turno-kognitivnyj i kommunikativno-funkcional'nyj aspekty: Monografiya. – M.: RUDN.

Zimnyaya, I.A. 2001. Lingvopsihologiya rechevoj deyatel'nosti. — M.: Moskovskij psihologo-social'nyj institut, Voronezh: NPO «MODEHK». (Seriya «Psihologi Otechestva»).

Perception of reduced words: Chunking and predictability

David Lorenz[1], David Tizón-Couto[2]
[1]English Department, Albert-Ludwigs-Universität Freiburg, Germany
[2]Facultade de Filoloxía e Tradución, Universidade de Vigo, Spain

Abstract

This is a first report on a word-monitoring experiment to examine how frequency-based chunking and predictability affect recognition of reduced speech. The effect of reduction on recognition of the word *to* was tested in English V *to* V_{inf} constructions of varying frequencies (e.g. *have to go, prefer to stay*). Our first results suggest that in types of mid-high frequency, predictability aids the recognition of a reduced item. In very high frequency sequences, however, reduction seems to encourage chunking, that is, accessing the sequence as a single unit.

Key words: chunking, reduction, frequency, speech perception

Introduction

It has long been noted that certain multi-word sequences undergo phonological reduction and contraction to a single word (e.g. *want to* > *wanna*). In usage-based approaches, this is seen as a matter of coalescence, or chunking, which in turn has been linked to frequency (i.a. Bybee 2006, Ellis et al. 2009). Thus high-frequency sequences will be stored in the mind as a single unit. They have a propensity for reduction due to neuromotor routines (Bybee 2006), but the reduced forms may be more or less strongly represented in the language user's mind, on a gradient cline from on-line reduction in articulation to stored, fixed variants (Connine & Pinnow 2006, Lorenz 2013).

Most of the evidence of chunking and the gradient status of reductions regards language production only, which raises the question how they affect speech perception. There is some evidence that full canonical forms generally serve the listener best (Tucker 2011, Pitt et al. 2011). In a word recognition experiment, Sosa & MacFarlane (2002) show that listeners treat highly frequent sequences as chunks, leading to a delayed recognition of elements of the sequence (e.g. *of* in *kind of*). Their design did not, however, consider these sequences' propensity for reduction (e.g. "kinda") and its effect on word recognition. In a similar study Kapatsinski & Radicke (2009) find a U-shaped frequency effect, such that word recognition is delayed in sequences of both very high and very low frequency. They suggest that frequent co-occurrence increases the predictability of a word, hence

facilitates its recognition, and that this is offset by chunking and low salience in collocations of very high frequency.

The present study builds up on this, testing the import of string frequency and reduction on speech perception. It employs constructions of the type V *to* V_{inf} (e.g. *need to work, dare to go*) to measure response times to the word *to*.

The crucial question is how frequency and reduction interact. In high frequency collocations, listeners may have an active knowledge of the high probability of *to* based on frequency, leading to a higher expectation of reduction (cf. Jurafsky et al. 2001); in this case reduction would not strongly affect recognition times. On the other hand, listeners may have a chunked item available; in that case a reduced form would lead them to access this chunked variant and considerably delay recognition of *to*.

Experiment design

The stimuli consist of 126 recorded sentences in American English. 42 of these contain a V *to* V_{inf} construction (the target items), 42 contain *to* in a different construction (control items), 42 do not contain *to* at all (distractors). Native speakers of American English were asked to respond to the presence or absence of *to* as accurately and quickly as possible. Response times were measured from the onset of *to*.

The V *to* V_{inf} sequences are of varying frequencies, as taken from the Corpus of Contemporary American English (COCA, Davies 2008-) – e.g. *trying to* V_{inf} (high frequency), *deign to* V_{inf} (low frequency). Participants were assigned to one of two groups; each group heard half of the target items with a full pronunciation, the other half with a reduced *to* (e.g. *need to* as "needa"). This reduction and the frequency of the sequence serve as independent variables whose effect on response times is tested.

At the time of writing, the study is still ongoing. We present here a sketch of the results from 22 participants, which gives a first impression of the interplay of frequency and reduction.

Results

Overall, participants correctly identified *to* within 2000 milliseconds in 89.7% of cases (1658/1848). When comparing conditions, however, the accuracy rate is significantly lower for reduced items than for fully articulated ones (82.7% vs 94.4%).

There is also a clear difference between full and reduced stimuli in the response times of the correct responses. Recognition of reduced items is significantly delayed compared to full items. The mean response times are:
Full *to*: 636 ms – Reduced *to*: 786 ms – Control: 683 ms

Response times to full and reduced items of different frequencies are shown in Fig.1. The four frequency bins are derived from the surface frequencies of the V *to* V_{inf} types in COCA, '1' being the lowest frequency (up to 1.5 occurrences per 1 million words), '4' the highest (over 290 per million).

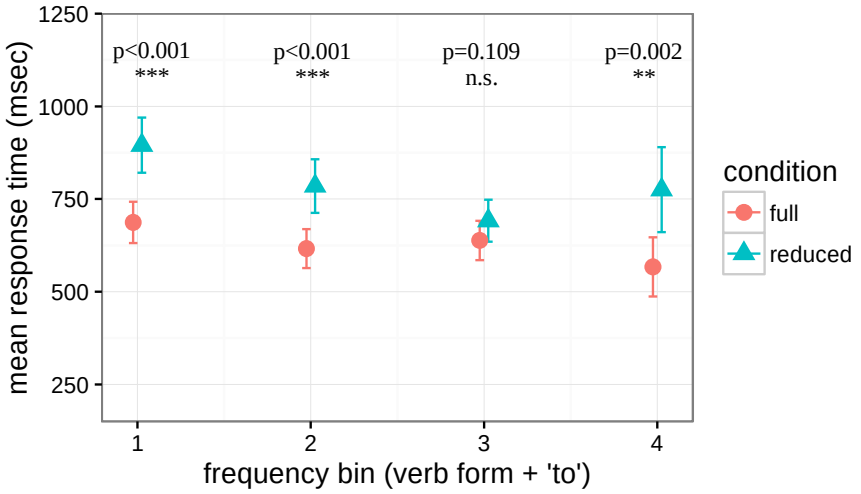


Figure 1. Response times to full and reduced *to* by frequency of V *to* V_{inf} type. The p-values refer to Mann-Whitney U test of difference between 'full' and 'reduced' in each frequency bin.

As Fig.1 shows, there is a clear difference between response times to full and reduced items, except at mid-high frequencies (bin 3). Recognition of reduced *to* is slowed down at low and very high frequencies. The pattern is less clear for the fully pronounced items, where recognition appears to be less sensitive to frequency.

Discussion

In low frequency collocations (bin 1), *to* is least predictable from context, and reduction will be least expected; here its recognition is slowest in both full and reduced forms.

Regarding the pattern for reduced items in Fig.1, our tentative interpretation is that there is a frequency range (around or within bin 3) at which *to* is highly predictable and reduction can be expected; therefore, reduction does not inhibit recognition. At higher frequencies (bin 4), a chunking effect sets in which inhibits recognition of the element and which is reinforced by a reduced rendering. Possibly, this chunking also implies an expectation of reduction, such that a reduced input leads the listener onto a

non-compositional access path (making it more difficult to retrieve the element *to*), whereas the non-reduced form encourages a compositional interpretation and thus does not inhibit recognition of the element.

These results need to be checked against possible other factors such as the form and length of the verb preceding *to*. It also remains to be seen how the frequency measure employed here – surface frequency of construction types – compares to measures of transitional probability or mutual information. In general, the findings suggest that hearers use probabilistic and frequency information to cope with reduction in the flow of speech.

References

Bybee, J. 2006. From usage to grammar: The mind's response to repetition. Language 82(4), 711-733.

Connine, C. and Pinnow, E. 2006. Phonological variation in spoken word recognition: Episodes and abstractions. The Linguistic Review 23, 235-245.

Davies, Mark. 2008-. The Corpus of Contemporary American English: 450 million words, 1990-present. Available online at http://corpus.byu.edu/coca/.

Ellis, N., Frey, E. and Jalkanen, I. 2009. The psycholinguistic reality of collocation and semantic prosody (1): Lexical access. In Römer, U. and Schulze, R. (eds.) 2009, Exploring the Lexis-Grammar Interface, 89-114. Amsterdam, John Benjamins.

Jurafsky, D., Bell, A., Gregory, M. and Raymond, W. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J. and Hopper, P. (eds.) 2001, Frequency and the Emergence of Linguistic Structure, 229–254. Amsterdam, John Benjamins.

Kapatsinski, V. and Radicke, J. 2009. Frequency and the emergence of prefabs: Evidence from monitoring. Formulaic Language 2, 499–520.

Lorenz, D. 2013. Contractions of English Semi-Modals: The Emancipating Effect of Frequency. NIHIN Studies. Freiburg, Rombach.

Pitt, M., Dilley, L. and Tat, M. 2011. Exploring the role of exposure frequency in recognizing pronunciation variants. Journal of Phonetics 39, 304-311.

Sosa, A. and MacFarlane, J. 2002. Evidence for frequency-based constituents in the mental lexicon: collocations involving the word *of*. Brain and Language 83, 227-236.

Tucker, B. 2011. The effect of reduction on the processing of flaps and /g/ in isolated words. Journal of Phonetics 39, 312-318.

Neurological state manifestation in infants' and children's voice features

Elena Lyakso, Olga Frolova
Child Speech Research Group, Saint Petersburg State University, Russia

Abstract

This study has the aim to find out the data about the reflection of the neurological state in the voice features of infants and children. Two types of experiments were conducted: comparing of vocalizations of 0-3 months old infants having neurological disorders (n = 45) and typically developed (TD) infants (n = 50); comparison of speech features of TD children (n=30) with vocalization and speech features of 5-16 years old children with autism spectrum disorders (ASD) (n=30). The results of the study showed that the infant's vocalizations contain features important for determination of the risks of development. Differences between children with ASD and TD on the basis of higher values of pitch, pitch variability and formant characteristics were revealed.

Key words: voice features, children, RAS, neurological state.

Introduction

The human voice contains the characteristics important for different states and developmental risk determination. Since 50 years of the last century the study of infants cry and pain vocalizations for purpose to diagnose neurological conditions were beginning (e.g. Wasz-Hockert, et al., 1996; Xie, et al., 1996). More recent studies have focused on the acoustic properties of speech production in autism spectrum disorders (ASD). Abnormal prosody has been identified as a core feature for ASD (Bonneh, et al., 2011), however in respect of pitch values and pitch variation, the data are contradictory (Nakai, et al., 2014). The goal of this study is to find out the acoustic features specific for developmental risk and ASD children vocalizations and speech.

Method

Data collection

Participants in the study were -3 months old infants with neurological disorders (ICD -10, 91.8, 91.9) (n = 45) and typically developed (TD) infants (n = 50), 5-14 years old TD children and children with ASD (F84.0; n=30). ASD children have varying degrees of neurological disorder severity. They were divided into two groups: presence of development reversals at the age

1.5 - 3.0 years (group-1- ASD -1) and developmental risk diagnosed at the infant birth (group-2 – ASD -2).

Two types of experiments were conducted: comparing of vocalizations of infants with neurological disorders and TD infants; and speech features of TD children with vocalization and speech features of ASD children. Different emotional states were used for comparing TD children and ASD children that allowed finding the variable characteristics of the voice.

Data analysis

The recording of vocalizations and speech was executed. Perceptive analysis of vocalizations and speech was made (200 adults). Spectrographic analysis of speech was carried out in the Cool Edit (Syntrillium Soft. Corp. USA) sound editor. The duration of vocalizations and pauses were measured. Pitch values, spectral maximums, their amplitude, and spectrum types were determined. Pitch values (F0), min and max pitch values, pitch range (F0 max - F0 min), formant frequencies and their amplitudes of vowels were measured in speech. All procedures were approved by the Health and Human Research Ethics Committee (HHS, IRB 00003875, St. Petersburg State University).

Result

Infant's vocalizations features

The "noise" spectrum frequently presents ($p<0.01$ –Mann- Whitney test) in the vocalizations of infants with neurological disorders than in vocalizations of TD infants (figure 1).

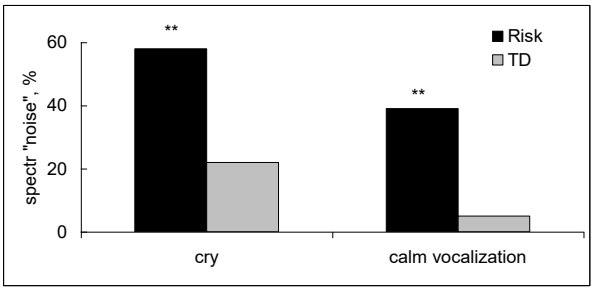


Figure 1. The duration of "noise" spectrum fragments in cry and calm vocalizations of TD infants and infants with neurological risk. ** $p<0.01$ – Mann - Whitney test.

The severity of the child's disease is reflected in the duration of vocalizations and the pauses between the phonation, the pitch values, and the predominance of vocalizations with "noise" spectrum.

Acoustic features of TD and ASD

Spectrographic analysis revealed that speech interpreted by listeners as discomfort, neutral and comfort is characterized by a set of acoustic features. Discomfort TD children's speech samples are characterized by highest maximum pitch values ($p<0.01$), average pitch values ($p<0.05$) and pitch variation values (F0max-F0min) ($p<0.05$) vs. neutral speech sample. Correctly recognized by adults discomfort and comfort speech do not differ in pitch variation values. Discomfort state is mostly characterized by falling pitch contour type, comfort state – by rising and neutral – by flat pitch contour.

For all children with ASD voice and speech is characterized by high values of the pitch, abnormal spectrum, and well-marked high-frequency. Discomfort state in the vocalizations and speech of ASD children, adults recognized better ($p<0.01$ Mann-Whitney test) than comfort and neutral state. Discomfort ASD children's speech samples are characterized by vowels' highest average pitch values, pitch range, and third formant frequency of vocalizations and words ($p<0.001$) than comfort and neutral speech samples.

Pitch average values (figure 2), pitch variation values (F0max-F0min) in ASD-1 child's discomfort, neutral and comfort speech significantly higher ($p<0.001$) than in ASD-2 child's speech. Pitch contour type does not change depending on the emotional state of ASD children. The F3 values in discomfort speech of ASD-1 children significantly higher than in corresponding voice features in ASD-2 children ($p<0.01$) and TD peers ($p<0.01$).

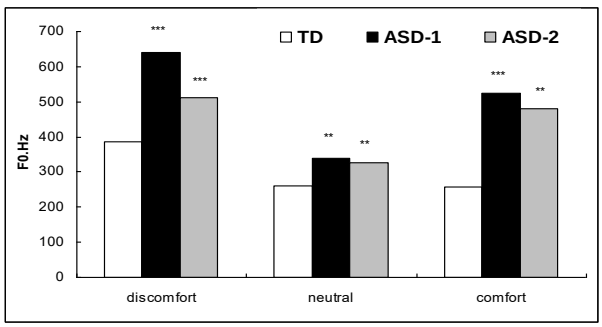


Figure 2. Vowel's pitch average value in discomfort, neutral and comfort state. **- $p<0.01$, *** - $p<0.001$ Mann-Whitney test.

The heaver child disease, the higher pitch values and third formant values, the lower speech level was revealed. Spearman correlation ($p<0.05$) was revealed between child's group and pitch values, third formant values.

Conclusion and discussion

We described the set of acoustical features that can be considered as one of the diagnostic sign of neurological disease and its severity. This result is amplifying with the findings of other studies on the early diagnosis of the infant's state on the voice features (Wasz-Hockert, et al., 1968). We present the first data for Russian ASD children of acoustic measures of participant's speech. Differences between children with ASD and TD on the basis of higher values of pitch, pitch variability and formant characteristics of ASD children were revealed. Our data confirm other studies with similar results [e.g. Paul, et al., 2005]. We believe that the acoustic features of speech of children with different neurological state are perspective for early diagnosis of developmental risk.

Acknowledgements

The work was supported by Russian Foundation for Basic Research (grants 15-06-07852a, 16-06-00024a).

References

Bonneh, Y.S., Levanov, Y., Dean-Padro, O., Lossos, L., Adini, Y. 2011. Abnormal speech spectrum and increased pitch variability in young autistic children. Front. Hum. Neurosci., 4. doi: 10.3389/fnhum.2010.00237

Paul, R., Augustyn, A., Klin, A., Volkmar, F. 2005. Perception and production of prosody by speakers with autism spectrum disorders. Journ. Autism Dev. Disord. 35, 205–220.

Nakai, Y., Takashima, R., Takiguchi, T., Takada, S. 2014. Speech intonation in children with autism spectrum disorder. Brain and Devel., vol. 36, 6, 516-522.

Wasz-Hockert, O., Lind, J., Vuorenkoski, V, Partanen T, Valanne E. 1968. The infant cry, a spectrographic and auditory analysis. London: Heineman Medical Books.

Xie, Q., Ward R.K , Laszlo, C.A. 1996. Automatic assessment of infants' levels-of-distress from the cry signals. IEEE Trans. on Speech and Audio Proc. vol. 4, 253 - 265.

Features of written texts of people with different profiles of Lateral Brain Organization of Functions (on the Basis of *RusNeuroPsych* Corpus)

Tatiana Litvinova[1], Ekaterina Ryzhkova[2], Olga Litvinova[3]
[1]Regional Centre for Russian Language, Voronezh State Pedagogical University, Russia
[2]Department of Russian Language, Voronezh State University of Engineering Technologies, Russia
[3]Department of English Language, Voronezh State Pedagogical University, Russia

Abstract

The aim of the study is detection of typological characteristics of written texts created by people with different profiles of the lateral brain organization of functions (LBOF). The material of the study is a special Russian text corpus *RusNeuroPsych* containing metada about LBOF (motor, sensory, cognitive) of their authors. Numerical values of a range of formal language parameters (index of lexical diversity, frequencies of parts of speech, etc.) were extracted from 242 texts and statistically significant ($p<0.05$) correlations between numerical values of a range of parameters of written texts and LBOF of their authors were identified for the first time for Russian texts.

Key words: written text, Russian, neuropsychology, brain lateralization, text corpus.

Background

One of the most important neuropsychological characteristics reflecting individual differences in the joint operation of the human brain hemispheres (asymmetry) is the lateral brain organization of functions (LBOF, Khomskaya et al. 1997). It is considered the foundation for the typology of individual differences of the mental condition of healthy individuals as part of a study in neuropsychology of individual differences. Neuropsychology of individual difference is an application of neuropsychological concepts and methods to the assessment of healthy subjects that tries to explain normal functioning by using principal of cerebral organization particularly characteristics of interhemispheric asymmetry and interaction (Glozman 2004, 838). The studies by Khomskaya et al. (1997) showed a stable correlation between the types of LBOF and different aspects of cognitive, motor and emotional activity of the normal subjects, which means that we have a correct foundation for the norm typology.

LBOF has an influence on the characteristics of the speech production as well (Shubin 2007) but this problem has not been sufficiently studied. There are mostly studies into the connection between the lateral brain organization and types of speech disorders (e.g., see Gudkova 2010), ways of formation of speech, acquisition of reading and writing skills (Litvinova 2013).

According to the literature, the connection between the lateral brain organization and characteristics of a written discourse of healthy individuals has not been dealt with and this is why the ongoing research project is of significance. We hypothesize that the lateral brain organization of functions impacts the characteristics of a produced written discourse and the classification of individuals according to their lateral brain organization can be used as the basis for a classification of language personalities.

Aim of the study

The aim of the study is to detect typological characteristics of coherent written texts created by people with different profiles of the lateral brain organization of functions using methods of statistical analysis and corpus linguistics.

Experimental study

Material

In order to address this problem it is necessary to create the corpus of written texts containing information about the type of LBOF of their authors. The text corpus *RusNeuroPsych* created under the guidance of the authors currently contains 643 Russian-language written texts by 447 authors (native Russian speakers) from 12 to 35 years of age. *RusNeuroPsych* corpus contains metadata in the form of information about their authors: year of birth, gender, native language, education, the results of psychological testing and survey for identifying their motor, sensory and cognitive lateral profile using the most indicative and simple tests (see Sirotyuk 2003, Semago 2005, Balonov 1985). The index of the lateral brain organization (motor, cognitive, sensory as well as individually for hands, legs, eyes, ears) was calculated as the difference between the number of "right", "left" and "symmetrical" answers divided into the number of tests. An integral index of LBOF was also computed as the difference between the number of "right", "left" and "symmetrical" answers divided into the number of tests.

For the present study 242 texts by 121 respondents (each respondent wrote two texts – letter to a friend and description of a picture) aged from 24 to 35, 17 men, 104 women, were selected. The average length of text is 165 words.

Methods

The texts were marked with the help of a morphological analyzer polymorpy2 and online service istio.com and the numerical values of the formal-grammar parameters of texts were obtained (indices of lexical diversity of texts, frequencies of different parts of speech and their ratios and other frequent parameters that occur in texts regardless of their topic and genre, 22 in total). SPSS Statistics software was used to calculate the Pearson coefficient between the text parameters and indices of LBOF. Two series of experiments were conducted: in the first one both texts by the same author were considered as one ("a sum corpus") and in the second one two texts were considered individually ("an individual corpus").

Results

Significant correlations ($p < 0.05$) between the formal-grammar parameters of written texts and the type of LBOF of their authors which were observed in two series of the experiments were revealed. The largest number of correlations of the parameters ($r = 0.27$-0.41) of texts were found with $LBOF_{motor}$ (8), $LBOF_{hands}$ (8), $LBOF_{integral}$ (7). There were much fewer significant correlations found with the indices of sensory and cognitive asymmetry except $LBOF_{eyes}$ (5). A positive correlation of the indices of $LBOF_{hands}$, $LBOF_{motor}$ and $LBOF_{integral}$ with the index of lexical diversity TTR was identified and a negative one with a proportion of function words + pronouns; proportion of function words; proportion of cognitive words; proportion of full stops; proportion of 100 most frequent Russian words, i.e. the more right properties there are in the human LBOF, the higher is the lexical diversity of their texts and the fewer function words, pronouns, full stops, most frequent words they have.

Conclusions and future work

Our pilot research proved studies of the connection between parameters of texts and LBOF of their authors promising. There will have to be more respondents considering their gender distribution as well as more text parameters. One of future studies will be looking at the causes of the identified correlations. It is also planned to conduct a correlation-regression analysis to construct mathematical models allowing one based on the formal-grammar text parameters to predict the type of LBOF of their authors (cf. Juola 2013), as well as to search for correlations between the type of LBOF and other characteristics of the authors of texts (gender, age, data of psychological testing).

Acknowledgements

The study is financially supported by the grant of RFBR "Linguistic Parameters of a Written Text and Neuropsychological Characteristics of its Author: A Corpus Study", project number 16-36-00036.

References

Balonov L. Ya., Deglin, V. L. and Chernigovskaya T. V. 1985. Functional Brain Asymmetry in Speech Organization. In Sensory Systems. Sensory Processes in Hemisphere Asymmetry. Leningrad, Science.

Glozman, J. 2004. Russian neuropsychology after Luria. In Craighead, W., Nemeroff Ch. (eds.). The Concise Corsini Encyclopedia of Psychology and Behavioral Science. NY, Wiley & Sons.

Gudkova, T. V. 2010. Features of Functional Sensomotor Asymmetry in Preschool Children with a General Speech Disorder. PhD thesis. Saint Petersburg, Herzen State Pedagogical University of Russia.

IBM SPSS Statistics 22 Documentation. http://www-01.ibm.com/support/docview.wss?uid=swg27038407#ru

Juola, P., Neocker, Jr. J., Stolerman, A., Ryan, M., Brennan, P. and Greenstadt, R. 2013. Keyboard Behavior Based Authentication for Security. IEEE IT Professional, 15, 4, 8-11, July-Aug.

Khomskaya, Ye.D., Yefimova, I.V., Budyka, Ye.V. and Yenikolopova, Ye.V. 1997. Neuropsychology of Individual Differences (Left-Right Brain and Mental Condition). Moscow, Russian Pedagogical Agency.

Litvinova, G. V. 2013. Effect of Lateral Organization on the Formation of Speech in Children. Petropavolvsk-Kamchatskiy, Vitus Bering Kamchatka State University.

Semago, N. Ya. and Semago, M. 2005. Theory and Practice of the Evaluation of Mental Development of a Child. Preschool and Junior School Age. Saint Petersburg, Rech.

Shubin, A.V. and Serpionova, Ye.I. 2007. Brain Asymmetry and Features of Verbal Creativity. Voprosy psikhologii 4, 89-97.

Sirotyuk, A.L. 2003. Neuropsychological and Psychophysiological Learning Component. Moscow, Sfera.

Semantic differential as a method in empirical investigation of Self-Image as father

Robert Manerov[1], Kristina Manerova[2]
[1]Department of Psychology and Pedagogy, University of Emercom, Russia
[2]Department of German Linguistics, Saint Petersburg University, Russia

Abstract

In our current study psychosemantics principles are used in the development of own method "Father Image", based on the method of semantic differential of Ch.Osgood. Following images are conceptual constructs in the study: Father Image, Self-Image as a father, Real Self-image as a father, Constructive Self-image as a father. The semantic differential provides a measure of 47 signs of the "Father Image", expressed by bipolar seven-point scale. The 47-sign scales are named with antonym pairs of Russian adjectives and contradictional propositions, which were composed by the modified method of M. Kuhn and T. McPartland and then evaluated by groups of single and married male probationers with and without kids

Key words: semantic differential, self-Image, father Image

Introduction

The study was based on the psychosemantic approach (cf. V. Petrenko, A. Shmelev, Ch. Osgood, J. Kelly et al.). We used psychosemantic principles to develop our own method called "'Father image' semantic differential" and to interpret the findings. Currently, the main goals of the psychosemantic approach include building and reconstruction of the individual value system through which the subject perceives the world, other people and himself. Our own "'Father image' semantic differential" method was based on Osgood's semantic differential, which is a part of experimental psychosemantics.

Methodology and results

The object of study was evaluated by subjects based on 47 bi-polar seven-point scales (Table 1). Each scale was built on the principle of opposition of a pair of antonyms (we used antonyms and antonymous phrases with an evaluative component in Russian). Let us define the basic concepts of our study. *Father image* – a set of views on paternal roles, functions and qualities that reflect the sociocultural, gender, age-related attitudes, traditions and stereotypes, as well as one's personal experiences with one's father.

Self-Image as a father – a set of man's views on his own paternal needs, roles, functions and qualities that reflect the sociocultural, gender, age-related attitudes, traditions and stereotypes towards a man as a father.

Table 1. “Father image” semantic differential.

1	Intelligent	3210123	Unintelligent
2	Hostile	3210123	Friendly
3	Possessing authority	3210123	Lacking authority
4	Unsympathetic	3210123	Sympathetic
5	Caring	3210123	Not caring
6	Repulsive	3210123	Charming
7	Can be trusted	3210123	Cannot be trusted
8	Apathetic	3210123	Energetic
9	Teaches the child a lot of things	3210123	Does not teach the child much
10	Strict	3210123	Gentle
11	Decent	3210123	Indecent
12	Able to make sacrifices	3210123	Not able to make sacrifices
13	Insincere	3210123	Sincere
14	Supports the family	3210123	Does not support the family
15	Unfair	3210123	Fair
16	Worthy of emulation	3210123	Not worthy of emulation
17	Indifferent	3210123	Compassionate
18	Faithful	3210123	Unfaithful
19	Pessimistic	3210123	Optimistic
20	Loved	3210123	Hated
21	Insecure	3210123	Confident
22	Makes the child proud	3210123	Does not make the child proud
23	Rude	3210123	Gentle
24	Raises the child	3210123	Does not raise the child
25	Unhappy	3210123	Happy
26	Close	3210123	Distant
27	Vicious	3210123	Virtuous
28	Children love him	3210123	Children don’t love him
29	Irritable	3210123	Cool-headed
30	A family man	3210123	Not a family man
31	Passive	3210123	Active
32	Able to protect	3210123	Not able to protect
33	Selfish	3210123	Altruistic
34	Considerate	3210123	Inconsiderate
35	Weak	3210123	Strong
36	Irresponsible	3210123	Responsible
37	His love has to be earned	3210123	His love does not have to be earned
38	Lazy	3210123	Hardworking
39	Educated	3210123	Uneducated
40	Despised	3210123	Respected
41	Reliable	3210123	Unreliable
42	Authoritarian	3210123	Democratic
43	Loving	3210123	Not loving
44	Sad	3210123	Cheerful
45	Determined	3210123	Free-floating
46	Angry	3210123	Kind
47	Spends a lot of time with the child	3210123	Does not spend much time with the child

Real Self-Image as a father – a set of man's views on his actual paternal needs, roles, functions and qualities, on what kind of father he actually is.

Constructive Self-Image as a father – a set of man's views on paternal needs, roles, functions and qualities which reflects what kind of father the man aspires to become and sees himself becoming in the future, as well as the obstacles to the achievement of the said.

Table 2. Factors of the real and constructive Self-Image as a father in men sith different paternal and marital status.

SELF-IMAGE AS FATHER	#	Groups of men		
		Married fathers	Married men without children	Unmarried men without children
		Factors		
REAL	1	Morality (factor power – 0,178)	Syncretism (factor power – 0,180)	Syncretism (factor power – 0,181)
	2	The object of love (factor power – 0,139)	Caring and trustworthy teacher (factor power – 0,145)	The object of pride and love (factor power – 0,158)
	3	Social Activity (factor power – 0,097)	The object of love (factor power – 0,124)	Strong Personality (factor power – 0,137)
	4	Caring and trustworthy teacher (factor power – 0,071)	Social Activity (factor power – 0,089)	Social Activity (factor power – 0,086)
	5	Kindness (factor power – 0,057)	Mentor (factor power – 0,065)	Democratic (factor power – 0,055)
CONSTRUCTIVE	1	Syncretism (factor power – 0,220)	Syncretism (factor power – 0,213)	Trustworthy Teacher (factor power – 0,168)
	2	Morality (factor power – 0,107)	Trustworthy Teacher (factor power – 0,125)	Empathetic (factor power – 0,166)
	3	The object of children's love (factor power – 0,092)	The subject of love (factor power – 0,114)	Defender (factor power – 0,096)
	4	Social Activity (factor power – 0,057)	Material support (factor power – 0,102)	The object of pride (factor power – 0,095)
	5	Integrity (factor power – 0,047)	Morality (factor power – 0,096)	Social Activity (factor power – 0,081)
	6		The object of pride (factor power – 0,079)	Material support (factor power – 0,066)

In order to identify the structure of the real and constructive Self-Image as a father in men with different paternal and marital status, we carried out factorization of semantic differential scales. The scales were arranged in semantic factor groups. The factorization was carried out separately for all three groups of subjects, consecutively for the real and constructive images (Table 2).

Conclusions

A comparative analysis of the six factor structures of real and constructive Self-Image as a father in three groups of men led us to the following conclusions.

In both groups of men without children the constructive image is more geared towards the child and the family than the real one. The constructive image, unlike the real image, includes the "Supports the family" characteristic, which is not pronounced in the real fathers' group. It is likely that for most men who have not become fathers yet the issue of providing for the family becomes the central one in whether to have a child or not.

The real Self-Image as a father in married fathers is much more realistic and moderate than in the two groups of men without children, while the constructive (in many ways, ideal) father for them is a tentative model, distant from the real requirements and only partially realized in practice, as evidenced by the fathers' real experience.

The application of linguistically determined semantic differential method with subsequent factorization and quantitative assessment is justified in psychological research.

References

Manerov, R.V. 2013. The Self-Image as a father in the Men-Self-Concept. Thesis of Kandidate of Psychology Science, Herzen State Pedagogical University of Russia, URL: http://elibrary.ru/item.asp?id=22372944 (in Russian).

Manerov, R.V., Posokhova, S.T. 2012. The Factor analysis structure of the Self-Image as a father by men with different paternal and marital status. In: A young scientist in the modern science world: new aspects of the scientific search. – L&L Publishing, 189-197 (in Russian).

Manerov, R.V., Posokhova, S.T., Lippo, S.V. 2008. The father image and the personal self-actualization In: Herald of St. Petersburg University. Psychology, Sociology, Pedagogy, Nr. 3, 23-30 (in Russian).

Manerov, V. Kh. 2012. The experience of semantic differential approach in the research of the audio perception of verbal message. In: Traditions and innovations in Psychology in Russia. Proceedings of the International conference, dedicated to 215th Anniversary of the Herzen State Pedagogical University of Russia, 450-455 (in Russian).

Automatic assignment of labels in Topic Modelling for Russian Corpora

Aliya Mirzagitova, Olga Mitrofanova
Department of Mathematical Linguistics, St. Petersburg State University, Russia

Abstract

The main goal of this paper was to improve topic modelling algorithms by introducing automatic topic labelling, a procedure which chooses a label for a cluster of words in a topic. Topic modelling is a widely used statistical technique which allows to reveal internal conceptual organization of text corpora. We have chosen an unsupervised graph-based method and elaborated it with regard to Russian. The proposed algorithm consists of two stages: candidate generation by means of PageRank and morphological filters, and candidate ranking. Our topic labelling experiments on a corpus of encyclopedic texts on linguistics has shown the advantages of labelled topic models for NLP applications.

Key words: topic modelling, topic labelling, Russian corpora.

Introduction

In recent years, topic modelling has become one of the most fruitful statistical NLP procedures which allows to reveal internal conceptual organization of text corpora. A topic model is constituted by a family of probability distributions over a set of topics extracted from a corpus, a set of words occurring in a corpus and a set of texts forming a corpus. Various algorithms of topic modelling (LSA, pLSA, LDA etc.) have been successfully applied to English corpora (Daud et al. 2010) in research dealing with information retrieval, content analysis, WSD, machine translation, etc. However, Russian corpora are seldom involved in topic modelling procedures. Certain positive results have been described in (Mitrofanova 2015). Our project tries to fill in this gap.

Resulting topics are commonly represented as the top *n* terms with the highest probabilities, which often poses a great challenge in their proper and accurate interpretation. Assignment of a topic labels, i.e. a single word or a phrase able to describe the semantics of a given topic, significantly assists in this task. In most of the works on topic modelling, topic labelling is conducted manually, which is a tedious process prone to subjectivity.

There have been proposed numerous techniques of automatic topic labelling for English texts, including those relying only on the content of a given corpus (Mei et al. 2007), and those requiring external resources like Wikipedia (Lau et al. 2011) or various ready-made ontologies. All of them are two-stage methods varying in the means of generating and ranking

candidate labels. In this paper, we adopt the unsupervised graph-based approach as described in (Aletras, Stevenson 2014), where promising results for English were reported. We elaborate it to make it applicable for Russian corpora of specialised texts by modifying it at both stages.

Methodology

Candidate Generation

In order to generate candidate labels, the first 10 topic words are used to query a search engine. After that, the titles of the top 30 search results are combined into a text, which is then tokenised and lemmatised. Subsequently, an oriented text graph $G = \{V, E\}$ is created, where V is a set of nodes containing lemmas, E is a set of edges. Two nodes are connected if the respective lemmas occur in the window of ±2 words. We experimented with three approaches to the weighting of the graph.

I. All of the edges are equal to 1 (unweighted graph).
II. The edges are weighted according to the co-occurrence frequency for corresponding lemmas calculated inside the given text.
III. The edges are weighted with PMI values computed using the Russian Wikipedia as a referential corpus (228 million tokens).

Next, the PageRank value (Mihalcea 2004) is computed for each node. The obtained text graph now takes the following form: more important words have larger nodes with higher PageRank values, while more semantically related bigrams have thicker edges with bigger weight.

Since Wikipedia does not have individual articles for most technical terms, we cannot verify the validity of a candidate label by checking whether it is a title, as it was proposed in the previous approaches. Therefore, appropriate n-grams are filtered from the text graph according to the following morphological patterns: Adj + N, N + N in genitive case, N + Prep + N, N + Conj + N, etc. The contact phrases are concatenated into a single group and added as a supplementary candidate label.

Candidate Ranking

The second stage includes ranking of the extracted candidates. We examine the next three possible ranking metrics for each phrase label.

A. Simply summing the scores of the constituent words.
B. Normalizing the sum of the scores with regard to the phrase length.
C. Multiplying the sum by the coefficient calculated as $1 + 1/i$, where i is the position of the topic word in the original query. Thus we use the information about the probability of a constituent word belonging to the topic.

Experimental Evaluation

For experiments, we collected a corpus of Russian encyclopaedic texts on linguistics containing of 1,900 documents with a total of 1,3 million tokens. After pre-processing, that is lemmatising with an open-source tool *pymorphy2* and removing stop words, the size of the experimental corpus reduced to 800,501 tokens.

We performed a series of experiments on topic modelling with LDA algorithm implemented within a *scikit-learn* package for Python and obtained 20 topics, i.e. non-structured clusters of semantically related words. Finally, we automatically assigned a label to each topic.

Evaluation and Results

To evaluate the quality of the automatic assignment, we asked experts to manually assess the extracted labels according to the following ordinal scale from 0 to 3 as suggested in (Lau et al. 2011):

0 Label is completely irrelevant for the topic.
1 Label is hardly related to the topic and/or it is ungrammatical.
2 Label is semantically related to the topic, but covers its content only partially and/or has grammatical mistakes.
3 Label perfectly describes the topic and it is grammatically correct.

In addition, we had to consider the grammaticality of the labels in case of erroneously extracted phrases by means of morphological patterns discussed earlier.

We used choosing the first topic word with the highest marginal probability as a baseline method. The results for each experimental configuration and the baseline are reported in Table 1.

Table 1. Evaluation results for each experimental configuration.

Label ranking	Graph weighting			Baseline
	I	II	**III**	
A	2.01	2.03	**2.07**	1.03
B	1.63	1.70	1.87	
C	1.70	1.73	1.77	

Discussion

In this study, we address the gap in topic modelling for Russian corpora and present an algorithm for automatic assignment of topic labels adapted for Russian. It is based on the method described in (Aletras, Stevenson 2014),

but differs from it in several respects. In particular, we introduced a step of identifying valid phrases using morphological patterns. Moreover, we conducted a number of experiments with various combinations of procedures for weighting a text graph and ranking the candidate labels.

The expert evaluation of results indicates that building a graph weighted with PMI values and ranking the candidates by simply summing the PageRank scores of the constituent words (A) performs best (2.07 out of 3). However, using inner co-occurrence frequency instead of PMI has also shown acceptable results (2.03), which means that it is not necessary to perform the heavy computations of association scores on a referential corpus.

The lower results for candidate ranking using normalised sum (B) and a special coefficient reflecting the importance of a topic word from a query (C) can be explained by the bias of these metrics. Phrases extracted by the B methods tend to be short and too general, whereas C favors all the candidates with a topic word while ignoring other relevant labels.

Future work could include improvement of the phrase extraction algorithm, e.g. instead of filtering n-grams with part-of-speech patterns we could apply shallow parsing and consider syntactic chunks as candidate labels.

References

Aletras N., Stevenson M., Court R. 2014. Labelling Topics using Unsupervised Graph-based Methods. In Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics, vol. 2, 631-636, Baltimore, USA.

Daud A., Li J., Zhou L., Muhammad F. 2010. Knowledge discovery through directed probabilistic topic models: a survey. Frontiers of Computer Science in China 4, 280–301.

Lau J., Grieser K., Newman D., Baldwin T. 2011. Automatic Labelling of Topic Models. In Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, vol. 1, 1536–1545, Stroudsburg, USA.

Mei Q., Shen X., Zhai C. 2007. Automatic labeling of multinomial topic models. In Proc. of the 13th Intern. Conference on Knowledge discovery and data mining, 490, New York, USA.

Mihalcea R. 2004. TextRank: Bringing Order into Texts. In Proc. of EMNLP 2004, 404-411, Barcelona, Spain.

Mitrofanova, O.A. 2015. Verojatnostnoje modelirovanije tematiki russkojazychnyh korpusov tekstov s ispol'zovanijem kompjuternogo instrumenta GenSim. [Probabilistic topic modeling of the Russian text corpora by means of GenSim toolkit]. In Trudy mezhdunarodnoj konferencii «Korpusnaja lingvistika – 2015», St.-Petersburg, Russia.

The time course of sociolinguistic influences on wordlikeness judgments

James Myers, Tsung-Ying Chen
Graduate Institute of Linguistics, National Chung Cheng University, Taiwan

Abstract

This study examined how and when sociolinguistic factors affect wordlikeness judgments by near-native bilinguals of Mandarin, the prestige language of Taiwan, and Southern Min (Taiwanese). Auditory syllables nonlexical in both languages were recorded by two bilingual speakers, one with a S. Min accent and one with a Mandarin accent. Accent and target language (judging the syllables as Mandarin-like or as S. Min-like) were crossed across participant groups. Binary judgments collected via the Worldlikeness Web app were analyzed in terms of target language, accent, participant gender, Mandarin and S. Min neighbourhood density, and reaction time. Response patterns were affected by all of these variables, including reaction time, in ways consistent with the differing social status of the two languages.

Key words: wordlikeness, neighbourhood density, bilingualism, gender, time course

Introduction

Mandarin is the prestige language in Taiwan, though many speakers are also native speakers of Southern Min (Taiwanese), another Sinitic language, even if, as adults, they may be more fluent in Mandarin. This social situation raises psycholinguistic questions: how and when is the phonological processing of near-native bilinguals affected by sociolinguistic variables like language status, gender (given that women are expected to favour the prestige norm; Labov, 2001) and accent (given that S. Min-accented Mandarin is expected to be disfavoured; Chung, 2006)?

To find out, we conducted a wordlikeness judgment task in which speakers rated the acceptability of nonlexical items as possible words in Mandarin or in Southern Min. Since this task is sensitive to neighbourhood density (the number of lexical items minimally different from a test item; Bailey and Hahn, 2001), and the influence of neighbourhood density increases over time (Stockall, Stringfellow, and Marantz 2004), we were also interested to see how the social variables interacted with neighbourhood density (including in the non-target language: Frisch and Brea-Spahn 2010), as modulated by reaction time (since slower responses may be sensitive to later processes).

The lexicons of Mandarin and S. Min share crucial similarities: most morphemes are cognates across these languages, morphemes are virtually

always monosyllabic, and syllable structure is very simple. However, the phonotactics of S. Min is less restricted (e.g., licensing an oral/nasal vowel contrast), making its syllable inventory (around 2400) larger than that for Mandarin (around 1400). These similarities and differences make it possible to test bilinguals with a single set of nonlexical syllables that vary in wordlikeness relative to Mandarin, S. Min, or both.

Methods

We used an auditory wordlikeness judgment task.

Participants. 80 bilingual speakers of Mandarin and Southern Min (mean age 22 years, 42 female) were paid for their participation, with 20 in each of four groups defined by crossing target language and stimulus accent (explained below).

Materials. An initial set of 5116 syllables was generated by randomly combining Mandarin and S. Min onsets and rimes and removing items lexical in either language. IPA transcriptions of these syllables were then presented in random order to two female bilingual speakers for recording, one raised in a Mandarin-speaking home and the other in a S. Min-speaking home, which affected their accents. In a pretest, the sound files were presented to 12 bilingual listeners. Items that were misperceived as lexical by more than one listener were reviewed by another two bilingual speakers and removed if the speakers agreed on the judgment. This screening procedure left 129 syllable types for the main experiment. Mandarin and S. Min neighbourhood densities were computed for each item; to further reduce the influence of acoustic ambiguity and to aid cross-linguistic comparisons, these computations ignored tone and vowel nasality. Neighbourhood densities were correlated ($r^2(127) = .1$, $p < .001$), but not enough to pose collinearity problems.

Procedure. Depending on which of the four groups they were assigned to, participants were asked to judge syllables that were or were not S. Min-accented as being like Mandarin or like S. Min. The Web app Worldlikeness (Chen and Myers forthcoming; http://lngproc-4083.nitrouspro.com:3000/) was used to present the stimuli in a different random order for each participant. Responses were made by pressing either the ‘L’ key (like the target language) or the ‘S’ key (not like it). Trials ended if a response was received, or else after 4,000 ms. Both responses and reaction times (RT) from stimulus onset were recorded. Experimental parameters and results are available for download from the Worldlikeness website.

Results and discussion

The data were analyzed using mixed-effects logistic regression with participant and item as random variables, and with participant gender, target language, stimulus accent, log Mandarin and S. Min neighbourhood density *z* scores, log trial RT *z* scores, and all interactions (except between the two neighbourhood densities) as fixed variables. All of these factors influenced responses in ways consistent with the greater social status of Mandarin (all effects and interactions reported below were significant at $p < .05$). The overall acceptance rate for Mandarin (.27) was lower than for S. Min (.43), suggesting a greater resistance to nonwords in the more prestigious language. Mandarin neighbours improved Mandarin-likeness judgments, but this factor had no effect on S. Min-likeness. By contrast, S. Min neighbours increased S. Min-likeness but also lowered Mandarin-likeness, as if items were "tainted" by an affinity with the non-prestige language. This negative effect was particularly strong for female participants, reflecting the common finding that women favour prestige norms. S. Min accent also enhanced the positive effect of S. Min neighbours on judgments, but had no effect on the influence of Mandarin neighbours, suggesting that the S. Min lexicon may be encoded less abstractly than the prestige language, making phonetic detail (accent) matter when activating neighbours.

Particularly intriguing were interactions with RT. As shown in Figure 1, when judging Mandarin, slower responses were more accepting (thin lines), while the reverse was true when judging S. Min (thick lines). The left plot in Figure 1 shows that the rise in acceptance was particularly steep when judging the Mandarin-likeness of items with fewer Mandarin neighbours; S. Min-likeness and S. Min neighbours (right plot) showed no such interactions with RT. One interpretation of these results is that Mandarin neighbours are activated quickly but undergo further processing, compared with a more uniform process for S. Min. This may explain why Mandarin neighbours enhance the fastest judgments, yet reverse their influence for slower judgments, perhaps as more distant neighbours become activated. By contrast, the only temporal change for S. Min seems to be a reduction in an initial positive response bias.

Conclusions

Our bilingual wordlikeness judgment study confirms that gender, accent, and the social status of languages all influence real-time phonological processing. In particular, judgments for the more prestigious language were more critical, were hurt by neighbours in the less prestigious language (especially for women), and may have been processed more deeply (perhaps an indirect effect of the participants' lower fluency in the less prestigious

language). Many speakers across the world are near-native bilinguals in languages differing in social prestige. Since our experiment was run using Worldlikeness, a free web app for collecting and sharing wordlikeness judgments, we hope that interested scholars will use it to extend our findings across a much wider variety of languages.

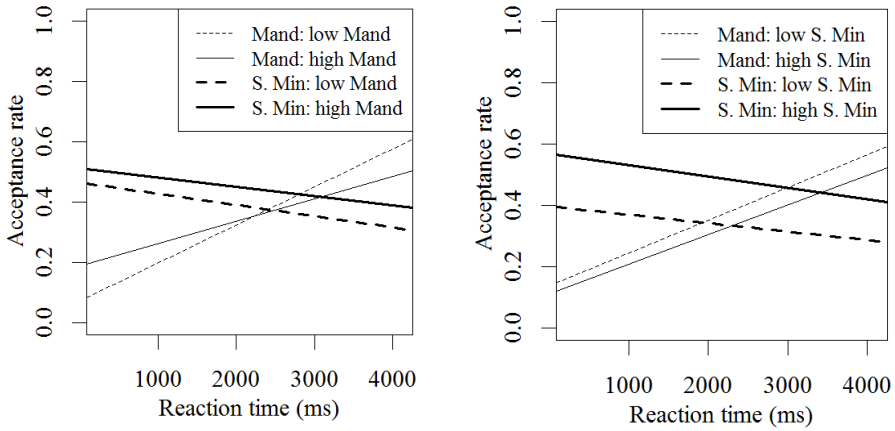


Figure 1. The effects of target language, reaction time, and neighbourhood density (left: Mandarin, right: S. Min) on wordlikeness judgments.

Acknowledgements

This study was supported by Ministry of Science and Technology (Taiwan) grant MOST 103-2410-H-194-119-MY3.

References

Bailey, T. M. and Hahn, U. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? Journal of Memory and Language 44, 568-591.

Chen, T.-Y. and Myers, J. Forthcoming. Worldlikeness: A Web-based tool for typological psycholinguistic research. Proc. of the 40th Annual Penn Ling. Conf., Philadelphia, USA.

Chung, K. S. 2006. Hypercorrection in Taiwan Mandarin. Journal of Asian Pacific Communication 16, 197-214.

Frisch, S. A. and Brea - Spahn, M. R. 2010. Metalinguistic judgments of phonotactics by monolinguals and bilinguals. Laboratory Phonology 1, 345 - 360.

Labov, W. 2001. Principles of Linguistic Change, vol. 2: Social Factors. Oxford, Blackwell.

Stockall, L., Stringfellow, A., and Marantz, A. 2004. The precise time course of lexical activation: MEG measurements of the effects of frequency, probability, and density in lexical decision. Brain & Language 90, 88-94.

The function of olfactory experience in reasoning: An empirical study

Katalin Nagy
Department of Languages, University of Jyväskylä, Finland

Abstract

This study reports the role of olfactory experience (i.e. smell of medication) in a nine-year old girl's reasoning in pair-work situation where the children were asked to choose items useful on a desert island. The extract analysed here is part of the larger data set of my dissertation, in which I investigate how sensory-motor activities involved in reasoning. I video-recorded an experimental task, in which the participants (N=27; age=9; Hungarian L1) have been asked to choose 7 items out of 14 to take those to an imaginary uninhabited island. The multimodal analysis shows that children did not choose the vitamin pills due to its unpleasant smell. The findings suggests that crossmodal experiences can be structural elements of reasoning.

Key words: multimodal analysis, sensory-motor activities, children's reasoning

Introduction

The distributed view of language has became a widely used term in applied linguistics. Most often it used to refer to the bodily, ecologically, socially or situationally distributed nature of language (Streeck, Goodwin & LeBaron 2011). During the last two decades, a great deal of research has been conducted on the embodied, visible aspect of interaction. In the last five years, kinetic behaviour, especially the use of gestures has been studied in a variety of contexts, including children's reasoning (e. g. Alibali et. al. 2011, 2014; Ehrlich et. al 2006). However, the function of sensory perception and motor activity during reasoning has been under-researched so far. Current investigations suggest that the cross-sensory experience of the world is created on the basis of interrelation between different sensual perceptions (Fulkerson 2014; Calvert & Thesen 2004; Ernst et al. 2007). Nevertheless, we have little information about how olfactory experiences are connected to body movements and verbal utterances when people interact. To fill this gap in the research I explore how the experience of smell were integrated into children's reasoning about the possible need of vitamin pills in a desert island.

Method

Data collection

Data collection took place in the hobby room of a Hungarian elementary school in the period of 3 weeks, during the afternoon day-care service. The multimodal data includes video-recordings of children completing a desert island task. In this activity the students were asked to choose 7 objects out of 14 to take with themselves to an imaginary uninhabited island. The task was completed in pairs were children were asked to make a shared choice. Furthermore, participants were asked individually and in pairs to justify their choices in an interview conducted by the researcher. In this paper I analyse a unique extract of a pair-work where children smelled the vitamin pills while they were reasoning about its' necessity.

The children and their parents were informed about the research task and the use of data in advance and their permissions were collected according to the Ethical Regulation of the University of Jyväskylä[1]. Further, I used pseudonyms and I blurred the video extracts in order to ensure the participants' privacy.

Participants

All together 27 fourth-grade students of two classes completed the desert island activity. In this study I analyse a pair-work of Janka and Orsi, since they smelled one of the task objects (vitamin pills) while they were solving the task. The children recreated their olfactory experience at the verbal, visual and kinetic levels of reasoning while they negotiated and made their decision whether they should or should not take the pills.

Data presentation and analysis

I applied multimodal interaction analysis to examine the integration of olfactory experience, gestures and verbal utterances. I annotated verbal and body actions in the Elan software. This kind of separate annotation of auditive and visible modalities was the most suitable strategy of data presentation I found for the purposes of my study. However, verbal and body activities have been viewed here as overlapping modalities, since speech is embodied by its nature (Levinson and Holler 2014). The transcript of the extract analysed in this paper covers an approximately 10-second-interval in the video-data. Here I provide a transcription of the utterances where an English translation appears below the original Hungarian utterances in

[1] Principles of research data management at the University of Jyväskylä, 2014. https://www.jyu.fi/tutkimus/tutkimusaineistot/rdmenpdf (accessed on 21 May 2016).

italics, followed by the annotation of bodily actions in double brackets. Overlapping actions are annotated in rectangles (see: Table 1).

Table 1. Transcript.

1	Janka:	Melyik legyen?
		Which one should it be?
		((picks up vitamine box))
2	Orsi:	Szerintem
		I think
		((moves Rh towards the VB))
3	Janka:	((opens VB))
4	Orsi:	((slightly pulls Rh backwards))
5	Janka:	Ebben igazi vitamin van!
		There is real vitamin in this!
		((looks at Orsi))
6	Orsi:	((looks at Janka (.) nods))
7	Janka:	((looks at VB, pulls it under her nose, smells))
8	Orsi:	((looks and moves her body and head towards VB))
9	Janka:	((pulls VB away of her nose))
10	Orsi:	((stops))
11	Janka:	((pushes VB under Orsi's nose))
12	Orsi:	((pulls head and torso backwards but moves her head towards VB and smells))
13	Janka:	Büdös.
		Stinky.
14	Orsi:	((gazes at Janka))
15	Janka:	((pulls back VB, gazes downwards, closes VB))

Multimodal data is used to analyse auditive and visual modalities of speech in linguistics. However, due to the crossmodal and multisensory nature of interaction it includes information about the integration of olfactory experiences to speech as well (Fulkerson, 2013). The data shows that involving movements and verbal utterances, sharing the experience of smell was in the focus of reasoning. Janka made her decision mainly on the basis on the perceived olfactory information. Orsi remained passive, but accepted the sensory reason (smell) and it's verbal conceptualization ('stinky') provided by her partner. Although Janka integrated multisensory experiences with speech, latter modality was only used to comment her decision. Based on this data I argue that reasoning is a multisensory activity. Children's justification included only one verbal comment of Janka ('büdös'/'stinky') and the rest of the reasoning elements came from

integrated cross-sensory perceptions (visual, auditory and olfactory) and kinetic actions.

Results

The micro-level observation of the data indicated that smell and vision in connection to the synchronised movements of heads, upper bodies, limbs and verbal processes were integrated while children were negotiating about the necessity of vitamin pills. Janka recycled the experience of smell to make her justification meaningful when she pushed the pills under Orsi's nose. Her decision was indicated bodily when she put the pills among the unnecessary objects. Finally, she summarised the action in a verbal utterance ('stinky'/ 'büdös').

Although there is a constant seek of underlying mental processes which may regulate human argumentation (e. g. Johnson-Laird, Khemlani and Goodwin, 2015) the findings of this paper suggest that a wide scale of cross-sensory experiences have meaningful functions in reasoning. Nevertheless, linguistic research on the connection between smell and meaning-making has just started (Pennycook and Otsuji 2015) and the findings of my case study are also limited. Therefore further studies are needed to explore how olfactory experiences are contribute in reasoning.

References

Fulkerson, M. 2013. Explaining multisensory experience. In Brown, R. (ed.) 2013, Consciousness inside and out: Phenomenology, neuroscience and the nature of experience, 365-373. Dordrecht: Springer.

Johnson-Laird, P. N., Khemlani, S. S. and Goodwin, G. P. 2015. Logic, probability, and human reasoning. Trends in Cognitive Sciences, 19(4), 201-214.

Lewinson, S. C. and Holler, J. 2014. The origin of human multi-modal communication. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 369 (1651), 2013030.

Pennycook, A. and Otsuji, E. 2015. Making scents of the landscape. Linguistic Landscape 1(3), 191–212.

Gender features in German: Evidence for underspecification

Andreas Opitz, Thomas Pechmann
Institut für Linguistik, Leipzig University, Germany

Abstract

A series of behavioural experiments is reported that investigate the processing of grammatical gender of nouns in German. Results consistently indicate processing differences between nouns of different genders. Masculine nouns show indications of increased processing cost compared to feminine nouns. We assume that the lexical representation of nouns is characterized by underspecified gender information. This assumption is in contrast to more traditional views stating that only inflected forms are underspecified with respect to grammatical features. However, the presented account supports the idea that underspecification as a general characteristic of the mental lexicon is mainly driven by economical reasons: a feature that is never used for grammatical operations (e.g., evaluation of agreement) is not needed in the language system at all.

Key words: grammatical gender, underspecification, German, mental lexicon

Background

In models of language processing, grammatical categories (e.g., gender or case) are traditionally split into distinct classes. For example, grammatical gender in German classifies into masculine, feminine or neuter. Current morphological theories however propose more differentiated analyses of these categories. Almost all frameworks rely on abstract feature decomposition and the concept of underspecification (see, e.g., Distributed Morphology (cf. Halle & Marantz, 1993), Paradigm Function Morphology (Stump, 2001), Minimalist Morphology (Wunderlich, 1996), and many others). The overall idea behind these two concepts is a decomposition of traditional labels into more abstract, binary features, thus allowing to refer to natural classes of such categories. Accordingly, the three instances of grammatical gender in German can be described by the following two abstract binary features [±f] and [±m]: 'feminine' [+f, −m], 'masculine' [−f, +m], 'neuter' [−f, −m]. In contrast, psycholinguistic models of inflection consistently lack such more differentiated morphological analyses. This holds, for such diverse models as schema-based models (Bybee, 1995), variants of connectionist models (cf. Rumelhart & McClelland, 1982), serial modular models (Levelt, Roelofs, & Meyer, 1999), the Augmented Addressed Morphology Model (Caramazza, Laudanna, & Romani, 1988), and others. However, relevant reason to implement the notions of

decomposed features and underspecification into a cognitive model of language would only be given if traditional and underspecification-based approaches make different, empirically testable predictions. Although there is first evidence in favour of such an account (e.g., Penke, 2006; Clahsen, Eisenbeiss, Hadler, & Sonnenstuhl, 2001), these findings have not yet been incorporated into existing models. The present study addresses the question whether there are inherent processing differences between nouns of different grammatical genders in German that are not due to a syntactic process of agreement checking (as suggested in Opitz et al. 2013).

Experiment 1 − Gender & Agreement / Grammaticality Judgment

Method & Procedure: 180 German nouns (60 of each gender), each of which was embedded in a syntactic structure of the type preposition + adjective + noun. For each phrase two illicit versions were created by marking incorrect gender agreement on the adjective. Stimuli were presented visually word-by-word centred on a computer screen with a fixed duration. 24 Participants performed a grammaticality task after each trial.

Results: Main effect of Gender for accuracy of responses: $F_1(2,46) = 7.7$, $p < .01$; $F_2(2,177) = 6.87$, $p < .01$. Feminine phrases were rated with higher accuracy (97.9%) than masculine (93.3%) and neuter phrases (93.1%). Main effect of Gender for reaction times: $F_1(2, 46) = 7.25$, $p < .01$; $F_2(2, 177) = 3.24$, $p < .05$. Responses to feminine phrases were faster (720ms) than to masculine phrases (758ms).

Experiment 2a − Morphological Marking / Gender Decision

Method & Procedure: A total of 252 nouns were chosen and distributed over 3 lists. Each list consisted of 84 nouns, 42 feminine and 42 masculine. In each of these two groups there were 21 nouns with derivational affixes clearly indicating their gender and 21 mono-morphemic nouns without gender cues. Items were presented visually and in a pseudo-randomized order. The task for 18 Participants was to decide whether the presented word was masculine or feminine by pressing a corresponding button.

Results: Main effect for Gender ($F_1(1, 17) = 22.01$, $p < .001$, $F_2(1, 83) = 6.78$, $p < .05$) but no effect for Morphological Marking ($F_1(1, 17) = 1.79$, $p = .19$, $F_2(1, 83) = 1.23$, $p = .27$). Again, there were significantly longer reaction times to masculine (769ms) than to feminine nouns (715ms).

Experiment 2b – Gender Verification

Method & Procedure: A total of 90 nouns was used, 30 of each gender. The task for 30 participants was to decide whether the presented word belonged to the gender category asked for in the particular block by pressing a corresponding Yes or No button.

Results: Main effect for Gender (F_1 (2, 58) = 3.55, $p < .05$; F_2 (2,83) = 3.90, $p < .05$). Decisions for feminine nouns were faster (686 ms) than for masculine nouns (720 ms). Neuter nouns scored numerically in between (703 ms) and did not differ statistically from either feminine or masculine nouns.

Experiment 3 - Word Class Decision

Method & Procedure: 60 nouns were used as experimental items, 20 of each of the three genders. 60 additional words (30 adjectives and 30 inflected verbs) were used as fillers. 30 Participants performed a word class decision.
Results: Main effect for Gender (F_1 (2, 58) = 17.7, $p < .001$; F_2 (2, 57) = 3.5, $p < .05$). Responses to feminine nouns (621ms) were shorter compared to responses to masculine (656ms) and neuter nouns (652ms). Latencies did not differ between neuter and masculine nouns.

Discussion

In all reported experiments we obtained evidence that gender features of nouns have an impact on language processing in German. Consistently, masculine nouns induced longer reaction times and partially lower accuracy rates, both indicating increased processing demands for masculine nouns, compared to members of the feminine category. We assume that the observed effects are grounded in an underspecified representation of grammatical features. In contrast to previous accounts, both in theoretical linguistics and psycholinguistics, we propose that the notion of underspecification extends to the representation of gender features of nouns in the mental lexicon. More precisely, we assume gender features of German nouns to be lexically specified as follows: masculine nouns:[−f, +m]; neuter nouns: [−f]; feminine nouns: [ ]. Moreover, the proposed specifications not only match the present data, but also agree with existing accounts of inflectional morphology (Blevins, 1995). These specifications can alternatively be modelled as generic gender nodes in an activation based model (cf. Levelt et al. 1999). Traditionally, generic gender nodes are viewed as categorical instances of grammatical gender. Each noun is associated with one of these nodes. In contrast, an underspecification-based account predicts that nouns in the mental lexicon differ in the number of associations to feature nodes (see Figure 1). Thus, the number of these associations corresponds to processing costs as mirrored in reaction times and error rates in behavioural experiments. This assumptions straightforwardly leads to further predictions concerning , e.g., priming experiments by providing a possible explanation why the so called gender

congruency effect occurs rather unsystematically across experiments and is notoriously hard to replicate (cf. Friederici & Jacobson, 1999).

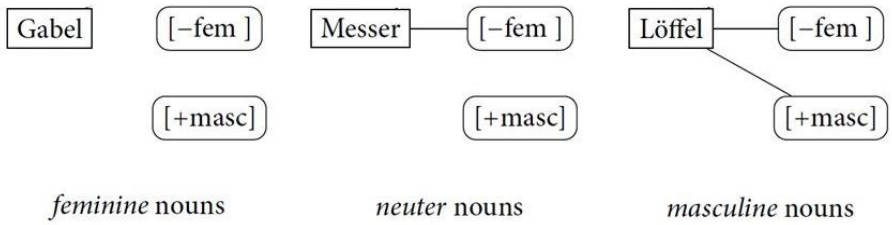


Figure 1. Lexical specification for the three genders of nouns in German.

References

Blevins, J. 1995. Syncretism and paradigmatic opposition. *Linguistics and Philosophy, 18*, 113–152.

Bybee, J. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes, 10(5)*, 425–455.

Caramazza, A., Laudanna, A., & Romani, C. 1988. Lexical access and inflectional morphology. *Cognition, 28*, 297–332.

Clahsen, H., Eisenbeiss, S., Hadler, M., & Sonnenstuhl, I. 2001. The mental representation of inflected words: An experimental study of adjectives and verbs in German. *Language, 77*, 510–543.

Friederici, A. D., & Jacobson, Th. 1999. Processing grammatical gender during language comprehension. *Journal of Psycholinguistic Research, 28*, 467–484.

Halle, M., & Marantz, A. 1993. Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The View from Building 20. Essays in Linguistics in Honor of Sylvain Bromberger. Vol. 24 of Current Studies in Linguistics* (pp. 111–176). Cambridge, Mass.: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22(1)*, 1–75.

Opitz, A., Regel, St., Müller, G., & Friederici, A. D. 2013. Neurophysiological evidence for morphological underspecification in German strong adjective inflection. *Language, 89(2)*, 231–264.

Penke, M. 2006. *Flexion im mentalen Lexikon*. Tübingen: Max Niemeyer.

Rumelhart, D. E., & McClelland, J. L. 1982. An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89*, 60–94.

Stump, G. 2001. *Inflectional Morphology*. Cambridge: Cambridge University Press.

Wunderlich, D. 1996. Minimalist morphology: The role of paradigms. In: G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1995* (pp. 93–114). , Dordrecht: Kluwer.

Distributional analysis of Russian lexical errors

Polina Panicheva
Department of Mathematical Linguistics, Saint Petersburg State University, Russia

Abstract

An algorithm of analyzing obscure lexical collocations is proposed. It is based on a co-occurrence model and distributional semantic filtering. We apply the proposed technique to lexical errors of construction blending, as annotated in the Corpus of Russian Student Texts. Results of error processing are analyzed and classified; reasons for different results in the paraphrasing experiment are discussed.

Keywords: Distributional Semantics, lexical errors, construction blending, Russian.

Introduction

We propose a framework for analyzing violation of syntagmatic relations resulting in construction blending [Puzhaeva et al. 2015]. Our toolkit includes models of meaning and selectional restrictions, applied to analyzing different types of abnormal collocations: native speakers' and learners' errors, metaphorical expressions, peculiarities in clinical texts, etc. The algorithm allows to identify and correct obscure collocations. We discuss the application of our approach to a corpus of native speaker errors.

Datasets

As a training corpus we use the RNC-Sketches syntactic bigram statistics. It provides statistics on syntactic relations in the Russian National Corpus (RNC), where every keyword is associated with a list of its relations and their frequencies in terms of MaltParser and TreeTagger; the latter are used to create RNC Sketches [Sharoff 2008, Sharov 2011] to the testing data. Total word frequencies were obtained from the Russian Frequency Dictionary [Lyashevskaya, Sharov 2009]. We supply our algorithm with an RNC-based Word2Vec semantic model [Kutuzov, Andreev 2015].

The data used for automatic error analysis is provided by the Corpus of Russian Student Texts (CoRST). It contains educational texts by native speakers of Russian and is annotated with different types of errors. The errors caused by construction blending [Puzhaeva et al. 2015] are especially relevant to our task, as they present subtle violations of selectional restrictions.

Statistical models

We use the RNC-Sketches syntactic bigrams as the syntactic model and apply automatic ranking of the erroneous keywords based on their context. The list of possible substitutes for a particular keyword is the intersection of the words occurring with every syntactic relation in the keyword context. The substitutes are ranked using the association measure scores: context-based paraphrasing (CBP) [Shutova 2010], and Word2Vec-based semantic scoring [Kutuzov, Andreev 2015].

Context-based paraphrasing

The context-based paraphrasing (CBP) likelihood estimation is based on the same grounds of syntactic co-occurrence, but is not symmetric and does not account for context word frequencies:

$$(1)\ L_i(CBP) = \frac{\prod_{n=1}^{N} f(w_n, r_n, i)}{(f(i))^{N-1}}$$

Word2Vec semantic scoring

In order to account for purely semantic word properties, i.e. restrict the list of substitutes to words semantically similar to the keyword, we apply the Word2Vec model trained with RNC data. Semantic similarity between a keyword *kw* and it's substitute *i* is calculated as the cosine distance between the corresponding vectors in the Word2Vec semantic space:

$$(5)\ Sim(kw, i) = \cos(kw, i)$$

The similarity threshold for the candidates with the initial erroneous word is experimentally set to 0.1.

Experiment setting

We perform a proof-of-concept experiment by analyzing the errors caused by construction blending in CoRST with context-based paraphrasing and additional Word2Vec semantic scoring. The errors are made by native speakers and represent violations of selectional restrictions. There are 27 sentences in the corpus annotated with a noun presenting a lexical construction blending error. We set out to automatically suggest a list of substitutes for the erroneous nouns and score them according to the CBP procedure with Word2Vec semantic filtering.

The results are manually analyzed, and the errors are grouped according to their proposed substitution candidates. The first group contains errors for which the distributional algorithm proposed no relevant candidates. For the

second group we calculate the Accuracy of the results by applying manual evaluation. A candidate is marked correct if it fits the context at least as well as the erroneous keyword and leaves the meaning of the sentence unchanged. Evaluation is performed in two settings:

1. The **strict mode** implies that the substitutes provided by the algorithm are correct if the candidate with the highest rank is correct.
2. The **loose mode** renders the substitutes list correct if there is a correct candidate among the four highest ranked candidates.

Results and analysis

Errors with no substitution candidates

There are 12 errors with no relevant candidates proposed. Eight of them obtain candidates by CBP, but the candidates are correctly filtered out by Word2Vec. Four errors get no proposed candidates, as their syntactic context is so obscure that there are no words attested in the corpus occurring with all the relevant syntactic distribution. The errors are exemplified in Table 1. Manual analysis shows that all of these cases appear to contain no error, or the error is annotated with a mistake, e.g. for a wrong word (Ex.1). A few of the 11 cases contain morphosyntactic analysis errors (Ex.1) or obscure syntactic relation names (Ex.2) immediately affecting the CBP candidate choice.

Errors with relevant substitution candidates

15 errors obtain substitution candidates with CBP which pass the semantic filtering. Examples are presented in Table 2. Nine errors are correctly analyzed in the strict mode (Ex.1), 4 errors are correctly analyzed in the loose mode (Ex.2). There are 2 errors left which only get incorrect candidates (Ex.3).

Conclusions

The distributional approach to lexical errors is an adequate measure of the distributional specificity of a construction in text; it also presents a useful tool which automatically suggests lexical substitutes for unusual lexical co-occurrences. Where lexical substitution is impossible, manual analysis confirms no lexical error in the sentence (44%). Proposed lexical substitutes (56%) are correct in 60% and 87% in strict and loose mode respectively.

Future work includes modifying the morphosyntactic analysis to minimize parsing errors. Future applications of the approach include specific error collections, i.e. language acquisition and learner errors, clinical texts, in order to shed light on their distributional nature.

Table 1. Errors with empty candidate lists.

№	Example sentence	Syntactic context	
		Relation	Word
1	…всех тех, кто взял на себя **роль** донесения фактов до массового сознания / … those who took the **role** of informing the masses	1-компл	взять / take
		1-компл	донесение / informing
		до_Gen	сознание / -
2	находит **себе** применение третий ход по реализации стратегии дискредитации… / the third approach to discredit applies **itself** …	неакт-компл	находить / -

Table 2. Errors with relevant substitution candidates.

№	Example sentence	Candidates	Result
1	Обязательно попробуйте национальный **окорок** – хамон... / You have to try the national **ham** – jamon…	блюдо / meal напиток / drink продукт / product	Strict correct
2	Если следовать **взглядам** Мари Биша … / following the **views** of Marie Bichat …	тенденция / trend правило / rule	Loose correct
3	Путешествие в Санкт-Петербург не нанесет **ущерба** вашему кошельк / A trip to St. Petersburg will not bring **damage** to your purse	потеря / loss	Incorrect

Acknowledgements

The reported study is supported by RFBR grant 16-06-00529.

References

Kutuzov, A., Andreev, I. 2015, 'Texts in, meaning out: neural language models in semantic similarity task for Russian', arXiv preprint arXiv:1504.08183.

Lyashevskaya, O., Sharov, S. 2009, The Frequency Dictionary of Modern Russian (on the materials of the Russian National Corpus), Moscow. (in Russian)

Puzhaeva, S.; Zevakhina, N., Dzhakupova, S. 2015, Construction blending in non-standard variants of Russian in the Corpus of Russian Student Texts. Proc. 6th Intern. Conf. “Corpus Linguistics-2015”, 390-397. St. Petersburg. (in Russian)

Sharoff, S.; Kopotev, M.; Erjavec, T.; Feldman, A., Divjak, D. 2008, Designing and Evaluating a Russian Tagset., in 'LREC'.

Sharov, S., Nivre, J. 2011, The proper place of men and machines in language technology. Processing Russian without any linguistic knowledge. Proc. Annual Intern. Conf. Dialogue, Computational Linguistics & Intellectual Technologies', pp. 657.

Shutova, E. 2010, Automatic metaphor interpretation as a paraphrasing task, in 'Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL', pp. 1029--1037.

Serbian pitch accents in tri-syllables produced by Serbian and Russian speakers

Ekaterina Panova
Dept of history and theory of language, St. Tikhon's Orthodox University, Russia

Abstract

This study is based on the analysis of tri-syllables in initial, medial and final position of statements. For each syllable of the tri-syllables the set of pitch parameters was calculated, as well as F0 inter-syllable intervals. In FA pitch parameters reach maximum values on first syllables and in RA – on second ones. FA and RA more differ in initial than medial position and tend to neutralization in final position. In initial and medial position Russian speakers realize a "type of accent" that is similar to Serbian RA and in final position a "type of accent" that similar to Serbian FA.

Key words: pitch parameters, pitch accent, Russian, Serbian, tri-syllable.

Introduction

Traditionally, Serbian stress is characterised by two contrasts – pitch (falling/rising) and duration (long/short) that make four combinations: long rising (LR), long falling (LF), short rising (SR) and short falling (SF). Nevertheless, such clear classification of Serbian pitch accents, formed by the end of XIX century, has been revealing many discussions (see Lehiste, Ivic 1986, Keijsper 1987, Jokanovic-Mihajlov 2006). The main problems are concerned the distinctive parameters of falling (FA) and rising accents (RA). Recent investigations confirmed that standard Serbian pitch contrasts realized on the sequence of stress and post-tonic syllable(s): negative intervals between stress and post-tonic syllable are typical for FA, while positive intervals for RA; FA have early peak locations, while RA late ones. Our studies (Panova 2015, Panova 2016) supported these previous investigations and revealed that in di-syllables post-tonic syllable provided better FA/RA distinction than stressed one. The parameter of peak location (i.e. timing of F0 maximum) can provide FA/RA distinction only with respect to the pitch contour of the whole word, but not only with respect to stressed syllable. Russian speakers had difficulties in the production of FA/RA Serbian contrast: in non-final position of the statements they produced "types of accents" that were similar to Serbian RA.

Method

For the present study 42 words of tri-syllables with stress on the first syllable and different types of accents were selected. Each target tri-syllable word

was embedded in frame statements so as to occur in initial and medial position (42*3). Two native speakers of Serbian (S1, S2, females) and four Russian speakers (R1, R2, R3, R4, females) read the sentences in neutral style and normal tempo.

For both Serbian and Russian samples we calculated F0 contour and obtained following pitch parameters (Smirnova et al. 2007) for each syllable of the tri-syllables: F0 start value, F0 end value, F0 maximum, F0 minimum, F0 mean value, F0 range and timing of F0 maximum (time point of F0 maximum measured in % of the total syllable duration). For each tri-syllable word we also obtained values of two F0 inter-syllable intervals between first and second syllable and between second and third one.

For the statistical analysis an ANOVA repeated measures was performed separately for Serbian and Russian speakers and for initial, medial and final position of the statements. For the first six pitch parameters we investigated effects between independent variables ACCENT (LF, LR, SF and SR), and SYLLABLE (1, 2, 3). For the F0 inter-syllable interval we investigated INTERVAL (1 vs. 2) and ACCENT (LF, LR, SF and SR). More detailed statistical analysis was provided by post Tukey HSD tests. For the analysis of timing of F0 maximum we used Survival Analysis with ACCENT (LF, LR, SF and SR) as a grouping variable.

Results

The results for Serbian speakers showed that the main effects of SYLLABLE and ACCENT as well as interaction between SYLLABLE and ACCENT were highly significant ($p<0.0001$) for F0 start value, F0 end value, F0 maximum, F0 minimum, F0 mean value in initial and medial position (in Figure 1 we give an example of F0 start values). For all these pitch parameters we can obtain the same tendencies: FA reach maximum values on first syllables, second and third syllables demonstrate gradual decrease, while RA reach minimum values on first syllable (for F0 end value on third syllable) and maximum values on second syllable. Post hoc test showed that regarding these pitch parameters Serbian four accents divided mostly on two types: falling (LF and SF) and rising (LR and SR), within these types there is not any significant difference. At the same time Serbian accents differ more clearly in initial than in medial position, where the distinction between FA and RA is broken, because SF values in second syllable approach to the values of RA. Regarding syllables Serbian accents differ more in first and second syllables than in third ones.

In final position the results of these pitch parameters didn't show any significance regarding the main effect ACCENT, although the effect of SYLLABLE and interaction between ACCENT and SYLLABLE were significant ($p<0.001$).

The results for F0 range and timing of F0 maximum for Serbian speakers were not significant in all positions.

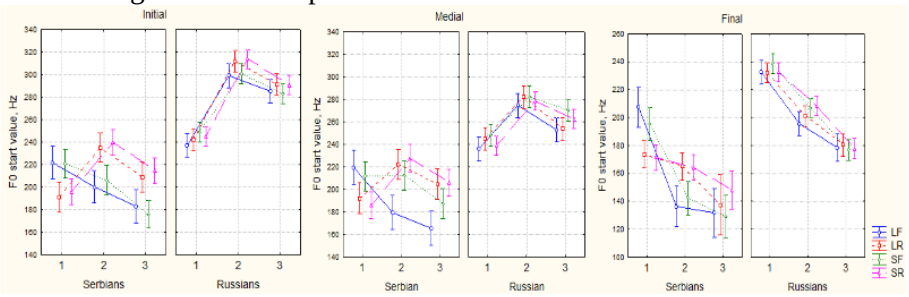


Figure 1. F0 start value scores of the first (1), second (2) and third (3) syllables with LF, LR, SF and SR for Serbian and Russian speakers in initial, medial and final position.

For Russian speakers the main effect of ACCENT was not significant regarding all pitch parameters except for marginally significant results for F0 start value ($p=0.043$) and F0 maximum ($p=0.047$) in final position. However, for Russian speakers the main effect of SYLLABLE was highly significant ($p<0.0001$) for all the parameters in all positions except for F0 range. As we can see from Figures 1, in initial and medial position the values of pitch parameters for Russian speakers are similar to Serbian RA: maximum values are reached on second syllable, while minimum values are on first one. In final position, on the contrary, the values of pitch parameters for Russian speakers are similar to FA.

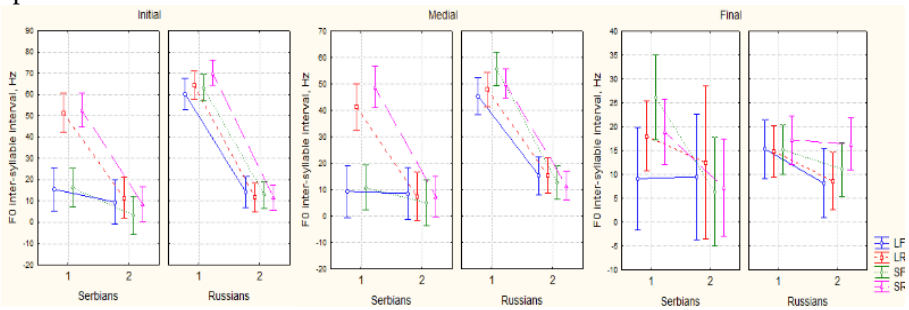


Figure 2. F0 inter-syllable interval scores between first and second (1) and second and third (2) syllables with LF, LR, SF and SR for Serbian and Russian speakers in initial, medial and final position.

The analysis of F0 inter-syllable interval revealed that for Serbian speakers in initial and medial position there is a significant difference between FA and RA only in interval between first and second syllable, while

interval between second and third syllable is not significant for FA/RA distinction (see Figure 2). FA have smaller intervals between first and second syllable than RA. For Russian speakers the values of interval between first and second syllable are similar to Serbian RA.

Conclusion

The results of pitch parameters of the tri-syllables produced by Serbian speakers showed that FA/RA distinction is provided on all three syllables, although first and second syllables are more significant than third one. For FA the pitch parameters (F0 start value, F0 end value, F0 maximum, F0 minimum, F0 mean value) reach maximum values on first syllables and minimum values on third one, while for RA pitch parameters reach maximum values on second syllables and minimum on first ones (except for F0 end value). The FA/RA contrast realizes more clear in initial, than medial position and in final position tends to FA/RA neutralization. In medial position the values of pitch parameters of the second syllable for SF approach to the values for RA, that correspond with the fact about tonal prominence of SF (Lehiste, Ivic 1986). FA/RA contrast can also be observed on the different F0 inter-syllable intervals between first and second syllable: RA have larger intervals than FA. The values of F0 range and timing of F0 maximum didn't demonstrate any FA/RA distinctive ability.

Regarding analyzed pitch parameters Russian speakers realize a "type of accent" that is similar to Serbian RA in initial and medial position and a "type of accent" that similar to Serbian FA in final position.

References

Jokanovic-Mihajlov J. 2006. Akcenat i intonacija govora na radiju i televiziji. Beograd.

Keijsper, C.E 1987. Studing Neoštokavian Serbocroation Prosody. Dutch Studies in South Slavic and Balkan Linguistics. SSGI 10, 101-193.

Lehiste I., Ivic P. 1986. Word and sentence prosody in Serbocroatian. Cambridge, Mass., MIT Press.

Panova E. Realization of Serbian accents by Serbian and Russian speakers (analysis of pitch parameters). Proc. International Conference of Experimental Linguistics ExLing 2015, 26-27 June 2015, Athens, Greece, 58–61

Panova E. L1 and L2 Serbian accents: Analysis of Pitch Parameters. Proceedings of the Speech Prosody 2016, May 31 - June 3, 2016, Boston, MA, USA, 474–478.

Smirnova, N., Starshinov A., Oparin I. & Goloshchapova T. 2007. Speaker Identification Using selective Comparison of Pitch Contour Parameters. Proc. 16th ICPhS, Saarbrucken, 203–206.

Effect of saliency and L1-L2 similarity on the processing of English past tense by French learners: an ERP study

Maud Pélissier[1], Jennifer Krzonowski[2], Emmanuel Ferragne[1]
[1]Laboratoire CLILLAC-ARP, EA 3967, Université Paris Diderot, France
[2]Laboratoire DDL, UMR 5596, CNRS – Université Lyon 2, France

Abstract

This study explored the effect of saliency and L1-L2 similarity on the processing of second language morphosyntax. ERP responses to violations of past tense morphology were obtained from adult intermediate French learners of English. Results show that participants processed L2-specific violations as salient events and not as morphosyntactic incongruities.

Key words: ERPs, L2 processing, syntax, L1-L2 similarity, saliency

Introduction

The way the syntax of our first language (L1) interacts with the syntax of a language we are trying to learn (L2) remains a much debated issue in the field of SLA. Some of the possible facilitating factors include the presence of similar structures in the L1 and the saliency of the morphosyntactic structure under scrutiny in the L2 (MacWhinney, 2005). In this study, we focused on a structure that contrasts these two factors: ERP responses to morphosyntactic violations of the past tense in polar questions in French learners of English with the auxiliaries DID and HAD. Polar questions using HAD followed by a past participle work in a way that is similar to French, where the past tense is marked both on the auxiliary and the main verb. On the contrary, questions with DID are specific to English in that the past tense is marked only on the auxiliary. However, violations of past-tense inflection are phonologically more salient with DID, where a past morpheme is added to the main verb, than with HAD.

Methods

Participants

26 intermediate French learners of English (5 male, aged 18.5 ± 1) took part in the experiment. They were first year University students of English having spent less than a month in an English-speaking country.

Materials and Procedure

The material consisted of 192 simple polar questions, half of them containing the auxiliary DID (DID Condition) and half HAD (HAD condition). Half of the sentences in each condition were made incorrect by varying the presence of the past morpheme. 120 sentences containing other agreement violations and 120 sentences containing a semantic violation were added as fillers.

Participants were asked to focus on the meaning of the sentence and evaluate its semantic acceptability while EEG data were recorded. A fixation cross appeared first for 500 ms and remained on the screen during the auditory presentation of the stimulus and for 1000 ms afterwards. A screen then prompted the participant to evaluate the semantic acceptability of the sentence by pressing a coloured button. As soon as the participant answered or after 2000 ms, the fixation cross appeared again and the next stimulus was presented.

Participants also completed a timed Grammaticality Judgment Task (GJT) with similar stimuli and additional fillers.

EEG data acquisition and analysis

EEGs were recorded with a Biosemi ActiveTwo system with 32 active electrodes, referenced on-line to the two mastoids and re-referenced off-line to the average of the two mastoids. Data were filtered on-line between 0.1 and 100 Hz. Electrode impedance was maintained below 20 Ohms and the signal was sampled at a rate of 512 Hz. Epochs from -200 ms to 1000 ms around the critical point (beginning of the critical past morpheme) were extracted from continuous data. After baseline correction (-200-0 ms) and low-pass filtering at 30 Hz, trials for which peak-to-peak amplitude exceeded 70 μV on the EOG channel or 100 μV on the other channels were automatically rejected. Electrodes were divided into central and lateral sites, the latter also divided into anterior/posterior region and left/right hemisphere. The following temporal windows were selected: 600-900 ms for the P600 and 300-500 ms for the LAN or N400.

Results

Behavioural measures: the GJT

A sensitivity index (*d'*) was computed for each participant and each auxiliary. Analyses showed that the participants' *d'* was marginally better in the Had condition ($F(1,25)=3.48$, $p=.07$) but their response time was shorter with DID ($F(1,25)=7.98$, $p<.01$) : on average, it took them 562 ms to respond to sentences containing DID and 634 ms for sentences containing HAD.

EEG results

A repeated-measures ANOVA with mean amplitude in the P600 window as dependent variable and Condition (Correct / Incorrect), Auxiliary (DID / HAD), Hemisphere (Left / Right) and Region (Anterior / Posterior) as within-subject variables showed an effect of the interaction between Condition and Auxiliary ($F(1,28)=9.15$, $p<.01$). Post-hoc analyses revealed that the effect of Condition in this time window was limited to sentences with DID ($p<.001$). A similar ANOVA was conducted on the mean amplitude in the 300-500 ms window and an effect of the Condition × Auxiliary interaction ($F(1,28)=25.68$, $p<.001$) was found. Post-hoc analyses revealed that with DID, the amplitude was greater in the Incorrect than in the Correct condition ($p<.001$) but that with HAD, the amplitude was more negative in the Incorrect than in the Correct Condition ($p<.001$).

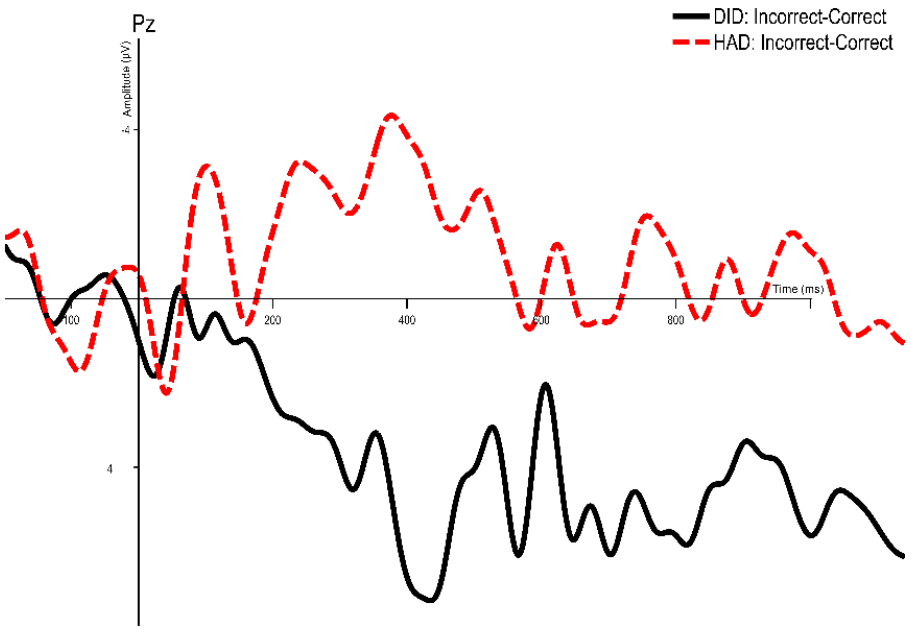


Figure 1. Difference wave (Incorrect – Correct) for each Auxiliary at Pz.

Discussion

Violations in the DID condition thus elicited a P600 as well as a positive peak in the 300-500ms window, resembling a P3 component. These violations involve the presence of the past morpheme in a context where it should be absent. They are therefore more phonetically salient than

violations with HAD, which are due to the absence of this same morpheme. These results are therefore consistent with the hypothesis that the P600 reflects, as the P3 does, the subjective salience of the stimulus (Sassenhagen, Schlesewsky, & Bornkessel-Schlesewsky, 2014). Besides, polar questions with DID represent a complex L2-specific structure, since they involve the movement of the inflectional morpheme from the main verb (where it would be in a declarative sentence) to the auxiliary. This represents an additional processing cost; yet participants were faster to decide for these sentences. This apparent discrepancy, as well as the presence of the P3, suggests that the P600 effect observed here in the DID condition is not a reflection of a better perception of the morphosyntactic error at hand but of an explicit reaction to the superior saliency of this violation.

Violations in the HAD condition elicited a negativity in the 300-500ms window that was not limited to anterior sites, thus more reminiscent of an N400 than a LAN. N400 effects have been found to be elicited by morphosyntactic violations even in native speakers (Tanner & Van Hell, 2014), possibly because those speakers rely more on lexico-semantic information to process their native language. It thus seems that these violations with HAD were not perceived as subjectively salient events but as lexical violations.

These results suggest that when the processed structure does not exist in the L1, other cues such as the phonological salience of the violation are used to process morphosyntactic violations. These findings also have theoretical relevance since they strongly support the P600-as-P3 hypothesis.

Acknowledgements

This research was supported by an IUF grant awarded to Dr. Emmanuel Ferragne.

References

MacWhinney, B. 2005. Extending the Competition Model. *International Journal of Bilingualism*, *9*(1), 69–84.

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. 2014. The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, *137*, 29–39.

Tanner, D., & Van Hell, J. G. 2014. ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, *56*(1), 289–301.

Phonostylistic study of Spanish-speaking politicians: Populist vs. conservative

Carmen Patricia Pérez
CLILLAC-ARP, Université Paris Diderot – Paris 7, France

Abstract

Conservative and Populist politicians can be easily recognized thanks to their phonostyle characterized by specific prosodic patterns. In this study, I analyzed four politicians' phonostyle in public 'spontaneous' speeches: Hugo Chavez (HC), José D. Ortega (JO), José R. Zapatero (Z) and Enrique Peña (EP). The acoustic analysis suggests that two main types of phonostyles can be found: a populist's phonostyles (HC and JO) and a conservative one (Z and EP).

Introduction

Conservative and populist politicians have a particular and typical way of speaking, their own 'phonostyle(s)', varying according to the different 'phonogenres' (specific conditions of productions such as interview, public speech, etc.). They are easily recognizable by the public. Studies on French politicians show that it is thanks to prosodic features such as prominence, acceleration, register change, breaks, etc. (Fónagy 1983; Duez 1997; Touati 1995; Léon 1993; Martin 2012). I will describe the prosodic features used by 4 Spanish-speaking politicians in public 'spontaneous' speeches: H. Chávez (Venezuela), J. Ortega (Nicaragua), J. Zapatero (Spain) and E. Peña (Mexico). This study is purely phonostylistic; I consider that the differences observed are due to the social and political backgrounds and not to the different varieties of spoken Spanish (Sosa 1999; Hualde & Prieto 2015).

Methodology

Corpus

The 4 realizations illustrated below come from 'spontaneous' public speeches delivered by the 4 politicians. They may be considered as representative of each speaker.

Intonation model

The interpretation of the prosodic analysis is based on Ph. Martin's model "Incremental Prosodic Structure" (1975-2015), where rising and falling contours do contrast indicating a relation of dependency between them, triggered by the following contours, firstly the final one of the utterance. These contours are developed on prosodic words (aka accent phrases, group

of one or more words with only one stressed syllable). They are described as follow: **C0**: Fall (very low) on the last stressed syllable and eventually on the following unstressed syllables to signal the end of an utterance; **C1**: Rise, above the glissando threshold (see the glissando formula in Rossi 1971, correlated with the speed of the melodic change); **C2**: Non-final falling contour, above the glissando threshold; **Cn**: 'Neutralized', i.e. slightly rising or falling, with a shortened vowel, below glissando threshold; **Cc**: fall-rise, flat or slightly falling on the stressed syllable and rising on the following unstressed one(s). **Ch** is phonetic, used by HC; it falls very low ('high dive' and lengthening on the last syllable) at the end of each intonation phrase (IP).

Acoustic analysis

After an initial perceptual analysis (Pérez 2014), the four politicians were classified in two different groups: populist (HC and JO) and conservative (Z and EP).

Populist Phonostyle: Hugo Chávez and José Ortega

HC and JO's utterances are divided into short chunks separated by long pauses. The last one ends with a C0; the preceding ones finish with a C1 on the penultimate stressed syllable of a word (Spanish frequent word stress pattern), followed by a Ch, a spectacular 'high dive' of about sixteen semitones on the last unstressed syllable, which is nearly twice as long as the stressed vowel. This is the phonetic marker of HC's phonostyle, as he does it in a regular way. When 'Ch' contours do not fall very low they can be considered as continuation contours (-*).

To interpret the prosodic structure, we need several levels hierarchy. At the top level, it is an enumeration of C1+Ch all contrasting with C0. At lower levels, there may be C1, Cn and C2 contours contrasting between them, with restrained melodic movements, but not depending on the final C0. This structure does not seem to be congruent with the syntactic-semantic structure as contours are regularly similar. In this way, the IPs could be analyzed like autonomous utterances.

JO's realizations are less marked and less regular than HC's ones.

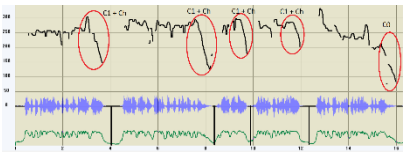


Figure 1. Chávez in public speech.

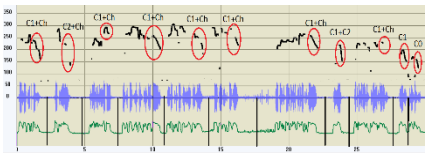


Figure 2. Ortega in public speech.

Conservative Phonostyle: J. Zapatero and E. Peña

Here also speeches are segmented into short chunks; for Z, almost all the contours end with a rising contour (C1) on the penultimate syllable, going most of the times higher on the last unstressed syllable but sometimes there are C2 or Cn contours. Z's utterance seems to be an enumeration with similar C1's (all pertaining to the top level) that contrast with the final C0, but inside the IP there are slope contrasts with less melodic movements.

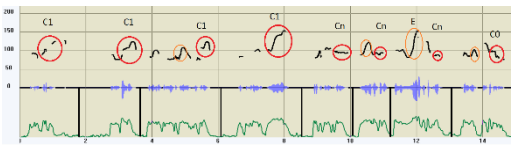


Figure 3. Zapatero in public speech.

For EP, the contour frequently employed is also C1 on the stressed syllable with a following unstressed syllable seldom rising but most of the time falling a little (but never like HC or JO). In the prosodic structure there is more contrast at the top and lower levels hierarchy.

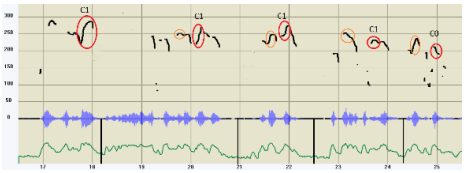


Figure 4. Peña in public speech.

Discussion and conclusion

The acoustic analysis suggests that two main types of phonostyles can be found: a populist phonostyles (HC and JO) and a conservative phonostyles (Z and EP). These phonostyles differ in similar speech situations ('phono genre'), mainly in (1) the realization of the final Intonation Phrase (IP) contours (see figure 5), (2) the F0 range and (3) the lengthening at the end of the IP. (1) In HC and JO, there is a F0 rising on the stressed syllable followed by a 'high dive' while Z's contours are most of the time made of a rise on the stressed syllable continuing with a rise to a higher F0 value. It could be noticed that EP is perceived as close to Z, but his contour is most of the time rising on the stressed syllable and falling a little on the following one(s). (2) the F0 range is wider for the 'populist' phonostyle than for the 'conservative' one (while the average F0 is similar: 250 Hz). (3) the

lengthening at the end of the IPs, is very frequent in 'populist' phonostyles and not at all present in 'conservative' phonostyles. Furthermore, the speech rate is similar and the IP construction is of the same type: speeches are generally segmented into small chunks (IP) and there is a resetting at the beginning of the IPs (not in Z pattern).

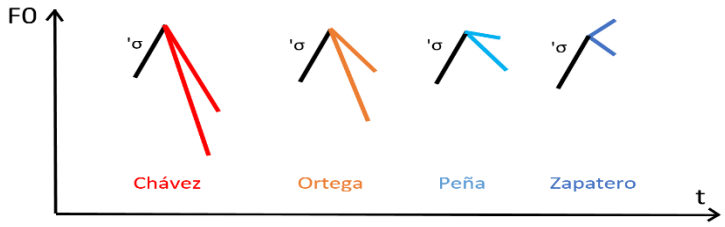


Figure 5. Common realization of the final IP contour (stressed 'σ + unstressed syllable).

References

Duez, D., 1997. Acoustic markers of political power. Journal of Psycolinguistic Research, 26(6), 641-654.

Fónagy, I., 1983. La vive voix. Essais de psycho-phonétique. Paris: Payot.

Frota, S. & al., 2007. The phonetics and phonologie of intonational phrasing in romance. Current issues in linguistics theory, pp. 131-154.

Léon, P., 1993. Précis de phonostylistique. Parole et expressivité: Nathan.

Martin, P., 1975. Analyse phonologique de la phrase française. Linguistics 146, pp. 35-68.

Martin, P., 2010. Intonation in Political Speech: Ségolène Royal vs. Nicolas Sarkosy. Rome, pp. 54-64.

Martin, P., 2015. The structure of spoken language. Intonation in Romance: Cambridge University Press.

Rossi, M., 1971. Le seuil de glissando ou le seuil de perception des variations tonales pour la parole. Phonetica, Volume 23, pp. 1-33.

Sosa, J. M., 1999. La entonación del Español: Su estructura fónica, variabilidad y dialectología. Madrid: Catedra.

Touati, P., 1995. Pitch range and register in french political speech. Proc. XIII International Congress of Phonetic Sciences, Volume 4, pp. 244-248.

Experimental L2 text production with WinPitch LTL

Darya Sandryhaila-Groth
LLF, UFR Linguistique, Paris-Diderot Paris 7, France

Abstract

Speech production of adults learning French as a second language in a non-francophone environment will be discussed in this paper. The focus is mostly on the prosody of French. Two groups of adult US native speakers used WinPitch Pro and its WinPitch LTL version for teaching and learning a foreign language. Their respective performances have been compared and evaluated.

Key words: Second language and prosody teaching, speech visualization.

Introduction

The oral performance in French as L2 has been ignored for a long time, especially suprasegmental but also their segmental aspects (Guimbretière 1994, 2000; Lauret 2007). Only recently, notable changes have occurred for learners of French, i.e., when authors of teaching methods began to be more interested in phonetics and included several exercises of repetition, discrimination, etc. in their textbooks of French (Abry 2009; Abry and Chalaron 2011; Kamoun and Ripaud 2016).

Methodology

In this study, two groups of learners of French were analyzed. All of them were American English native speakers and had an intermediate level in French. The first group of participants were university students at UCLA and the second one were adult students at the French language school *Alliance française*.

In a first step, individual comments were provided to each of the students, after the instructor has been listening to their individual recordings with Audacity software. The students had worked in groups and been listening to each other, and they were all interacting during the learning process. They were able to give their opinion about the quality of the repetition of a student, and his phonetic/prosodic errors. In addition, the instructor was listening and correcting the oral productions as well. To simplify the repetition task, models of the sentences were played to the students at reduced speed (70%), with the help of the WinPitch software. At the end of a training period, a final recording of each of the students in both groups was made with WinPitch LTL.

Hypothesis

The first hypothesis is that the first group of young university students at UCLA (on average 28 years old) has a better performance in their speech production than the second group of adult students (on average 65 year old); not only because the age difference, but also because of the first group learning French as a main subject in their university syllabus, while the second one is learning French mainly for pleasure and travel purpose.

The second hypothesis is that the real-time visualization during the prosodic training with WinPitch helps the students in improving their quality and 'natural sounding' of their speech productions in French.

Corpus

The corpus includes recordings from a model French speaker and the students from the two groups, all reading a short declarative text **"Dimanche en famille"**, a text coming from a short story written by P. Léon.

In this paper, only one sample sentence out of the whole corpus is analyzed: **Elle aimerait bien une petite friture de poissons.** *"She would like to eat some deep-fried fish."* Results from two male speakers of the first group and two female speakers of the second group are shown, see the Figs below.

WinPitch and L2 teaching

In this study we work with WinPitch LTL, a program developed for language teaching and learning by Philippe Martin, and WinPitch Pro. WinPitch LTL was first presented to potential users in Martin and Germain (2000), and is innovative in its real-time visualization. Designed as a traditional language lab with two tracks, the students first listen to the model speech and then try to reproduce it. The instructor can directly correct errors of the student's repetition (suprasegmental and segmental) or add comments for the next class. He can also manipulate the F0 curve and use different colorings to highlight, e.g., a rising/falling intonation or a final intonation.

WinPitch screenshots of student's oral productions

In this section we will consider WinPitch screenshots from the students of the two groups. First the comparison of the sentence production (two recordings of each speaker) without training and then after training, during the final recordings. For the results in some cases, I accepted almost correct oral production as good, see the tables below.

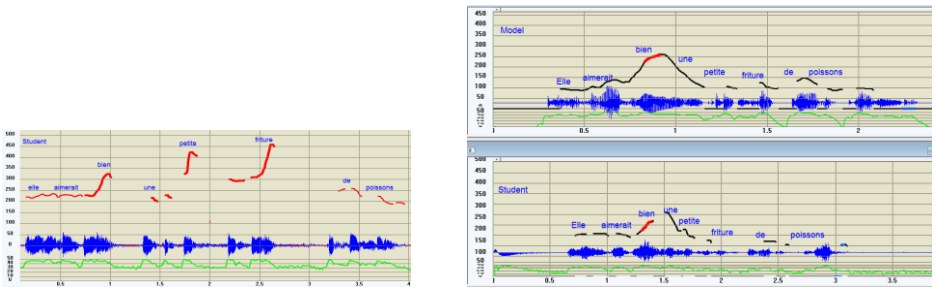


Figure 1.

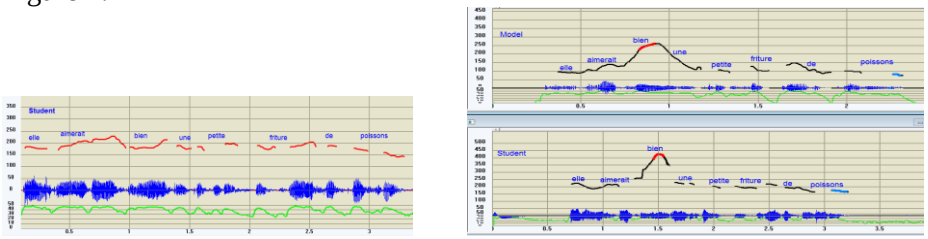


Figure 2.

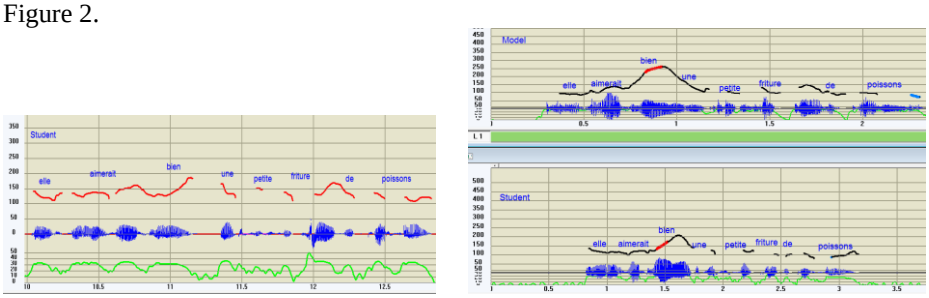


Figure 3.

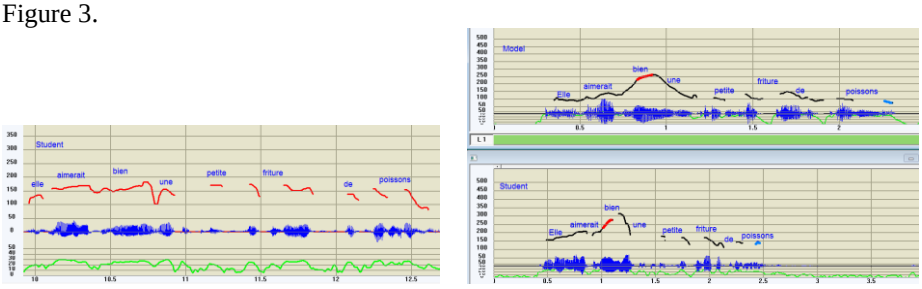


Figure 4.

Figures 1-4. Figure 1: Group 2 student 1, WP Pro first recording (left) and WP LTL final recording (right). Figure 2: Group 2 student 2, WP Pro first recording (left) and WP LTL final recording (right). Figure 3: Group 1 student 1, WP Pro first recording (left) and WP LTL final recording (right). Figure 4: Group 1 student 2, WP Pro first recording (left) and WP LTL final recording (right).

Table 1.

Elle aimerait bien une petite friture de poissons	First rec. AfSp.1gr.2	Final rec. AfSp.1gr.2	First rec. AfSp.2gr.2	Final rec. AfSp.2gr.2
Prosodic words	-	-	+	+
Declarative (C0d)	+	+	+	+
Prominence “bien”	+	+	+	+
Speech fluency&v.linking	-	-	+	+

Table 2.

Elle aimerait bien une petite friture de poissons	First rec. Sp.1gr.1	Final rec. Sp.1gr.1	First rec. Sp.2gr.1	Final rec. Sp.2gr.1
Prosodic words	-	+	-	+
Declarative (C0d)	+(!)	+	+	+
Prominence “bien”	+	+	+	+
Speech fluency&v.linking	-	+	-	-/+

Conclusions

The results of the presented sample sentence suggest a clear improvement in the speech production for the students in both groups after a training with WinPitch LTL. In a next step, we will continue analyze the full corpus, read by other speakers from the two groups, to confirm the hypothesis that the language proficiency depends on the pursued purpose, the wish to sound more natural, to be aware of the foreign language intonation while speaking.

References

Abry, D., Chalaron, M-L. 2009. Les 500 Exercices de phonétique A1/A2. Hachette.

Germain, A., Martin, P. 2000. Présentation d’un logiciel de visualisation pour l’apprentissage de l’oral en langue seconde. www.alsic.org, 3, No 1, 61–76.

Gumbretière, E. 1994. Phonétique et enseignement de l'oral. Paris, Didier-Hatier.

James, E. 1976. The acquisition of prosodic features of speech using a speech visualizer. IRAL 14 (3):227–243.

Kamoun, Ch., Ripaud, D. 2016. Phonétique essentielle du français. 100% FLE, Paris, Didier.

Lauret, B. 2007. Enseigner la prononciation du français: questions et outils. Paris, Hachette.

Martin, Ph. 1982. Utilisation d'un visualiseur de mélodie en vue d'une didactique. Options nouvelles en didactique du français langue étrangère. 181–186. Paris, Didier.

Martin, Ph. 1975. Analyse phonologique de la phrase française. Vol. 146, 35–68. Linguistics.

WinPitch LTL, 2015. www.winpitch.com.

Exploring prosodic convergence in Italian game dialogues

Michelina Savino[1], Loredana Lapertosa[1], Alessandro Caffò[1], Mario Refice[2]
[1]Dept. of Education, Psychology, Communication, University of Bari, Italy
[2]Dept. of Electrical and Information Engineering, Polytechnical Univ. of Bari, Italy

Abstract

In this study we explore the manifestation of prosodic convergence between pairs of Italian speakers involved in a non-competitive game. Results show evidence of prosodic convergence and/or divergence between partners, where prosodic parameters and coordination strategies involved can vary across dialogue pairs. Also, degree of asymmetry in prosodic convergerce appears to be related to speaker empathy.

Keywords: Prosodic convergence, game dialogues, Italian, Big Five Questionnaire

Introduction

Conversational partners have been observed to adapt each other's speech over the course of the interaction. This phenomenon, variously termed as convergence, entrainment, alignment, accomodation, or adaptation, is considered crucial for mutual understanding and successful communication, and influenced by linguistic, social, and cultural factors (e.g. Giles et al. 1973). A body of research has been devoted to measuring prosodic parameters involved in speech adaptation in a number of languages (for example, Levitan et al. 2015) not including Italian. This paper offers a preliminary investigation on prosodic convergence between Italian interactans, and explores the influence of speakers' personality traits on the convergence process.

Method

Corpus

Our corpus consists of five dialogues where pairs of players are involved in a modified version of the old Chinese Tangram Game, as developed within the PAGE project. Participants in a game round were given Tangram figures according to their role: the Director, who received a set of four Tangram figures, one of which marked by an arrow; the Matcher, who was given one of the figures belonging to the Director's set. Players could not see each other's figures, and goal of the game in each round was to establish – on the basis of common agreement – whether the figure given to the Matcher was

the same as the one marked by the arrow in the Director's set, or not. The game session consisted of 22 rounds, with an average duration of 30 min.

Speakers were selected according to a number of parameters which could influence adaptation, namely age, gender, and familiarity. They were all young adult females (aged 21-24), and MA student classmates. After the recording sessions, participants were administered the Big Five Questionnaire (BFQ-2, Caprara et al. 2007), a protocol used in psychology for assessing individual "Big Five Personality Factors" (Energy, Friendliness, Conscientiousness, Emotional Stability, Openness) along with their subdimensions.

Annotations and prosodic measurements

All dialogues were orthographically transcribed, and speech signal was manually annotated (by using Praat, Boersma 2001) along the following tiers: 1) Game rounds; 2) Inter-Pausal Units (speech bounded by silence longer than 100msec); 3) words; 4) syllables. The following prosodic parameters were automatically calculated via Praat scripts: pitch range (F0max-F0min, Hz), pitch level (F0 median, Hz), loudness (intensity, dB) and articulation rate (# syll/sec).

Measuring convergence

Given the explorative nature of this study, we started focussing on global aspects of speech coordination, i.e. those referring to similarity process undergoing at the level of the whole dialogue. We basically follow the approach proposed by Eldlund et al. (2009) in defining similarity as underlined by a) convergence, the process by which conversational partners' speech features become more similar over time until they converge; b) synchrony, when speakers' speech happen to have similar patterns over time. Due to space limitations, in this paper only results on convergence are presented. We looked for evidence of convergence by identifying cases in which speakers mean values were more similar to each other later in the dialogue. Accordingly, we splitted each game session into two halves: a window consisting of rounds 1-11 vs another window including rounds 12-22. Within each of the two windows, we compared mean values of speaker1 vs speaker2. Mean values found as significantly different in the first half but not in the second half were considered as evidence for convergence. Note that convergence can be realised on the opposite direction as a complementary manifestation of adaptation, i.e. divergence (Healy et al. 2014). Consequently, in our hypothesis mean values found as not significantly different in the first half but significantly different in the second half of the session were considered as evidence of divergence. All other cases were not taken as evidence for convergence or divergence.

Results

Prosodic convergence/divergence

Table 1 shows results of speaker1-speaker2 mean values comparison for each prosodic parameter, in the first vs. second halves of each game session. We found statistical evidence of convergence and/or divergence in four out of five dialogues: speakers in dialogue PZ become more similar in their voice loudness in the second part of the dialogue (convergence), whereas speakers in dialogue RC show complementary convergence by significantly diverging in their articulation rate in the second half of the session. In dialogues DS and CD we found both types of manifestation of overall coordination: participants in dialogue DS converge in their articulation rate, and diverge in the loudness of their voices, whereas speakers in dialogue CD converge in pitch range and diverge in pitch level. Speakers converging by some speech features yet diverging by some others in the same interaction has been reported (e.g. Bilous & Krauss, 1988, Eldlund et al. 2009).

Table 1. Comparison of speaker1 vs. speaker2 mean values in the first vs. second halves of dialogue (two-tailed t-test, t values only when significant: *=p<.05, **=p<.01, ***=p<.001). Light gray shaded boxes indicate convergence; dark gray shaded ones indicate divergence.

Dialogue	**Convergence/Divergence** speaker1-speaker2 mean values comparison, 1 half vs. 2 half of dialogue							
	Artic. rate		Pitch range		Pitch level		Loudness	
	1st half	2nd half	1st half	2nd half	1st half	2nd half	1st half	2nd half
CD	n.s.	n.s	2.18*	n.s.	n.s.	4.18***	2.29*	2.58*
DS	3.21**	n.s.	2.14*	2.16*	n.s.	n.s.	n.s.	2.16*
PP	n.s.	n.s.	n.s.	n.s.	-8.27***	-4.94***	4.66***	7.10***
PZ	n.s.	n.s.	n.s.	n.s.	-10.46***	-6.71***	-3.52**	n.s.
RC	n.s.	-2.69*	n.s.	n.s.	n.s.	n.s.	4.88***	4.89***

Prosodic convergence and speaker empathy

We explored the possible influence of speaker empathy (a subdimension of Friendliness) on the convergence process. We determined the degree of asymmetry in converging/diverging of each speaker in a pair by measuring mean values differences between the first and the second halves of the dialogue (only for the prosodic parameters involved). In Table 2, speakers whose absolute values were either greater (implying that they "converged more"), or smaller (implying that they "diverged less") with respect to their conversational partners, are the ones who consistently exhibit the higher scores for empathy in the pair (at least 10 T scores, 1 s.d.).

Table 2. Mean values differences (2^{nd}–1^{st} halves of dialogue) for each speaker in dialogues where convergence and/or divergence were observed, along with individual T scores for "Empathy" as assessed by the BFQ-2.

Dialogue	Speaker	Convergence 2^{nd}-1^{st} halves (mean values)	Divergence 2^{nd}-1^{st} halves (mean values)	Empathy (BFQ-2 T scores)
CD	sp1	10.12	9.50	58
	sp2	18.31	-2.50	70
DS	sp1	0.01	-0.90	56
	sp2	0.46	0.49	65
PZ	sp1	-0.04	-	59
	sp2	-2.53	-	76
RC	sp1	-	-0.43	61
	sp2	-	0.03	72

Discussion and conclusions

Results of this explorative study indicate that, at the whole dialogue level, Italian speakers tend to adapt their speech through a variable number of prosodic parameters, and by using different coordination strategies. These results are in line with those reported on languages investigated so far. In our data, we also observed that degree of asymmetry in convergerce/divergence appears to be related to speaker empathy. Though very preliminary, such observations are encouraging for future directions of our research.

References

Bilous F.,R., Krauss, R.M. 1998. Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. Language and Communication 8, 183-194.

Boersma, P. 2001. Praat, a system for doing phonetics by computer. Glot International 5(9/10), 131-151.

Caprara, G.V., Barbaranelli, C., Borgognani, L., Vecchione, M. 2007. Big Five Questionnaire-2, Giunti: Firenze.

Eldlund J., Heldner M., Hirschberg J. 2009. Pause and gap length in face-to-face interaction. In: Proceedings of Interspeech 2009, 2779-2782, Brighton, UK.

Giles, H., Taylor, D.M., Bourhis R.Y. 1973. Towards a theory of interpersonal accomodation through speech. Language in Society 2, 177-192.

Healey P., Purver M., and Howes C. 2014. Divergence in dialogue. *PloS one* 9(6) e98598, 1-6.

Levitan, R., Benus, S., Gravano A., Hirschberg J. 2015. Acoustic-prosodic entrainment in Slovak, Spanish, English and Chinese: A cross-linguistic comparison. In Proceedings of SIGDial 2015, 325-334, Prague, Czech Republic.

PAGE (Prosodic And Gestural Entrainment in conversational interaction in diverse languages) project: http://page.home.amu.edu.pl/

Syllable cueing and segmental overlap effects in tip-of-the-tongue resolution

Nina Jeanette Sauer
Goethe-University Frankfurt, Phorms Education Frankfurt

Abstract

The tip-of-the-tongue (TOT) phenomenon refers to a temporary word finding failure. To induce TOTs in the lab, a common method is to ask for terms after providing created definitions. When in a TOT, syllable cues were presented in order to manipulate TOT resolution. After the presentation of the correct first syllable of the target word, TOTs could be resolved faster and more accurately than after the presentation of an incorrect syllable of some other word or the control condition (Experiment 1: *syllable cueing effect*). The presentation of the extended syllable of the word (the first syllable with one more segment) facilitated TOT resolution and boosted lexical retrieval even more than the regular syllable (Experiment 2: *segmental overlap effect*).

Key words: tip-of-the-tongue (TOT), resolution, cueing, syllable, segmental overlap

Introduction

The tip-of-the-tongue phenomenon (TOT) represents a temporary impairment in speech production. When experiencing a TOT, one has access to semantic (concept) and syntactic information (lemma) but only partial access to phonological information (lexeme). While the complete word form cannot be retrieved, one has a strong feeling of knowing the word and "recall is felt to be imminent" (Brown & McNeill 1966, p. 325). Often, speakers are able to retrieve the first letter or phoneme, the number of syllables and also words with similar sound and similar meaning (Brown 2012, p. 196).

In order to induce TOTs in a laboratory setting, definitions were presented on a computer screen, for example, "a lift consisting of a series of linked compartments moving continuously" for *paternoster*. In the cueing paradigm so far, syllable cues were embedded in words or pseudowords, and presented in word lists in order to manipulate TOT resolution (for an overview, see Hofferberth-Sauer & Abrams 2014). Abrams, White, and Eitel (2003) illustrated, for example, that the *entire* first syllable is required for TOT resolution – the first phoneme or first grapheme alone had no effect. In the present studies, syllable cues were presented in isolation. The advantage of this procedure is that the syllable itself has no semantic and syntactic information. The presentation of isolated correct, incorrect, and extended syllables is new in TOT research.

Previous studies

In the pre-tests, definitions had been collected and verified (Hofferberth, 2011). In two pilot studies (Hofferberth 2012), the design of the experiment was evaluated, and more definitions were collected and validated. Thereafter, two experiments were performed. The first experiment (Hofferberth 2014; Hofferberth-Sauer & Abrams 2014) will be presented here only marginally while the focus is on the second experiment (cf. 3.). All the data was collected within my Ph.D. project (Sauer 2015).

Experiment 1

In the first experiment, definitions were presented on a computer screen. When in a TOT, one of three cues was presented. It was shown that after the presentation of the correct syllable (e.g., *pa* for *paternoster*), TOTs could be resolved about twice as fast compared to after an incorrect syllable (e.g., *co*) and to the control condition (*xxx*). The correct syllable also led to significantly more accurate answers (M = 73.5%, SD = 18.6%) compared to the control condition (M = 24.3%, SD = 16.4%, $t(47) = 16.39$, $p < .001$), and to the incorrect syllable (M = 16.0%, SD = 13.6%, $t(47) = 20.06$, $p < .001$). The control condition led to significantly more accurate TOT resolutions compared to the incorrect syllable ($t(47) = 3.71$, $p = .001$). The incorrect syllable did not block TOT resolution (not leading to more inaccurate answers), but there was an inhibition effect: There were fewer accurate answers and more unresolved TOTs. After demonstrating the cueing effect of the first syllable in Experiment 1, a further experiment was conducted in order to test if the syllable border plays a role (*syllable preference effect*).

Experiment 2

Method

Participants

69 under- and postgraduates (42 female, 27 male) between 21 and 35 years (M = 27.9 years, SD = 4.3) participated in this study.

Apparatus and material

The material was visually presented on a computer screen using the program *Presentation*. There were 240 definitions of German nouns presented in order to induce TOTs (the English examples here are only for demonstration purposes).

Procedure

The subjects were told to press a button on the keyboard as fast as possible indicating that they know the word (KNOW), that they do not know the word (DON'T KNOW), or that the word is on their tip of the tongue (TOT). They had 10 seconds to react to the definition. After pressing KNOW, they typed in the answer, and another definition was presented. After pressing DON'T KNOW, the next definition appeared on the screen. After pressing TOT, a cue was presented visually: either the regular syllable (e.g., *pa* for *paternoster*), the extended syllable (e.g., *pat*), or the control condition (marked by *xxx*). The cue was presented for 25 seconds. In this time, the subjects had to type in their answer.

Results

TOT rate

The number of TOTs varied between 21 (8.8%) and 194 TOTs (80.8%). Through 16560 stimuli overall, 5600 TOTs were induced, i.e., the TOT rate was 33.8% with 81 TOTs per person on average (SD = 14.7%). Out of the 5600 TOTs, 3385 TOTs (60.5%) were resolved in the given time of 25 seconds, with reaction times (RTs) between 571 ms and 24948 ms (M = 4049 ms, SD = 4325 ms). There were 50.3% accurate answers, and 10.2% inaccurate answers.

Cue analysis

The number of accurate TOT resolutions differed between the three types of cues ($F(2, 136) = 415.65$, $p < .001$). With the extended syllable, TOTs were accurately resolved significantly more often (M = 72.0%, SD = 18.7%) in comparison to the regular syllable (M = 60.3%, SD = 19.0%, $t(68) = 7.00$, $p < .001$), and to the control condition (M = 18.7%, SD = 13.0%, $t(68) = 26.26$, $p < .001$). The regular syllable led to significantly more accurately resolved TOTs ($t(68) = 19.80$, $p < .001$).

The RTs were significantly shorter after the presentation of the extended syllable (M = 2330 ms, SD = 887 ms) in comparison to the regular syllable (M = 2803 ms, SD = 1166 ms, $t(67) = 3.92$, $p < .001$), and to the control condition (M = 3017 ms, SD = 1592 ms, $t(62) = 2.89$, $p = .005$). There was no significant difference between the regular syllable and the control condition ($t(62) = 0.78$, $p = .436$).

Discussion

While Experiment 1 showed the syllable cueing effect, i.e., the correct first syllable helped to overcome transmission deficits from the lemma to the lexeme level, Experiment 2 showed the segmental overlap effect, i.e. a

speaker needs even more than the first syllable for successful TOT resolution. It was demonstrated that the extended syllable (e.g., *pat* for *paternoster*) significantly speeded up lexical access (shorter RTs), and significantly increased TOT resolution (more accurate answers) compared to after the regular syllable (e.g., *pa*) and to the control condition (*xxx*). The key factor was not the syllable per se but the information content: the bigger the segmental overlap between cue and target, the faster and better the TOT resolution. Therefore, it is helpful to get as much information as possible about the beginning of the target word. The unit of the syllable only plays a marginal role.

Syllable cueing and segmental overlap effects do not have to exclude each other but rather can both be explained within speech production models that allow for an interactive activation spreading and have a syllable level *below* the phoneme level. For an interpretation and discussion of these results within different models of speech production see Sauer and Schade (2016).

References

Abrams, L., White, K.K., Eitel, S.L. 2003. Isolating phonological components that increase tip-of-the-tongue resolution. Memory & Cognition, 31, 1153-1162.

Brown, A.S. 2012. The tip of the tongue state. New York, Psychology Press.

Brown, R., McNeill, D. 1966. The "tip of the tongue" phenomenon. Journal of Verbal Learning and Verbal Behaviour, 5, 325-337.

Hofferberth, N. J. 2011. The tip-of-the-tongue phenomenon: Search strategy and resolution during word finding difficulties. Proc. 4th ISCA Tutorial and Research Workshop on Experimental Linguistics, ExLing 2011, 83-86. Paris, France.

Hofferberth, N. J. 2012. On the role of the syllable in tip-of-the-tongue states. Proc. International Conference of Experimental Linguistics, ExLing 2012, 57-60. Athens, Greece.

Hofferberth, N. J. 2014. Resolution of lexical retrieval failures. Reaction time data in the tip-of-the-tongue paradigm. Proceedings of the International Seminar on Speech Production. ISSP 05-08 May 2014, 194-197. Cologne, Germany.

Hofferberth-Sauer, N.J., Abrams, L. 2014. Resolving tip-of-the-tongue states with syllable cues. In Torrens, V. and Escobar, L. (eds.), The processing of lexicon and morphosyntax, 43-68. Newcastle, Cambridge Scholars Publishing.

Sauer, N.J. 2015. Das Tip-of-the-Tongue-Phänomen. Zur Rolle der Silbe beim Auflösen von Wortfindungsstörungen. Doctoral dissertation, Frankfurt am Main, Johann Wolfgang Goethe-Universität. doi: 10.13140/RG.2.1.1229.8645

Sauer, N. J. and Schade, U. 2016. Über die Entstehung und Auflösung von Versprechern und Tip-of-the-Tongue-Zuständen. Manuscript in preparation.

An experimental study of English accent perception

Elena Shamina
Department of Phonetics, Saint Petersburg State University, Russia

Abstract

The study aims at proving the observation that in English oral speech perception, sociolinguistic evaluation prevails over personal one. The total of 10 speech samples by 2 native English speakers with no special phonetic or acting training imitating various English accents were evaluated by 26 native English speakers on a number of scales related to sociolinguistic and personal factors. When listening to the same persons speaking in different English language varieties the respondents ascribed to them very different social qualities, such as social class, education and occupation. The personality properties ascribed, such as character traits and age, are shown to depend on the social factors, associated with the accent.

Key words: sociolinguistics, perception, English accents, social and personal qualities

Introduction: sociolinguistic experimental data

Experimental studies have been used in sociolinguistics to demonstrate a consistent correspondence between pronunciation and social class in the English speaking societies (Wells 1982). They have shown that sociolinguistic evaluation is inherent in (at least English) speech perception and essentially depends on the sociolinguistic profile of the listener (Labov 1972), and that some varieties of English, including foreign accents, may be stigmatized (Coupland, Bishop 2007; Абрамова 2009). Validity of the social characteristics, such as socio-economic status, education, occupation, place of residence, ascribed to English speakers only on the basis of their pronunciation has been ascertained (Shamina 2011; Шамина 2012). Also, the data gathered from polls and questionnaires points to some informants having strong emotional reactions to certain accents.

Material and procedure

This particular study is undertaken in complete agreement with the previous research in the field of sociophonetics. It aims at proving the observation that in English oral speech perception, sociolinguistic evaluation prevails over personal one. The experimental procedure involved 2 native English speakers (both male and well-educated) who had no special phonetic or acting training but claimed that they could imitate various accents supplying

recordings of 11 - 25 seconds long on a neutral topic that had no relation to their social or personal characteristics. The varieties of English represented, except formal RP, were dialects of Manchester, Liverpool, Newcastle, Somerset, Yorkshire, West London (Hammersmith), Cockney, Southern Irish, (Southern) American English, as well as French English (defined as such by the speakers themselves). Respondents (26 native English speakers, both men and women in the age range of 21 – 64, of different social status and speakers of different national and regional varieties of English) were contacted via Internet and asked to evaluate the speech samples on a number of scales related to sociolinguistic factors, such as social class, occupation, education, place of residence and also personal factors, such as age and personal qualities. Their answers were then analysed.

Results and discussion

Perception of social properties

As in the previous research into the matter (Shamina 2011), the respondents were rather accurate in placing the speech samples on the map of world Englishes. But what is of most interest here is that when listening to the *same* persons speaking in different English language varieties the listeners ascribed to them very *different* social characteristics. For example, when Speaker 2 spoke in Somerset dialect his social status was evaluated as upper class and upper middle class by almost half of the respondents, but when he spoke in Newcastle dialect his perceived social position dropped dramatically and he was thought of as a representative of the working class by more than a third of the respondents. The level of education ascribed to the speakers, too, was a function of the variety of English spoken. For instance, Speaker 1 was considered to be university-educated when speaking formal RP by 73% of the respondents, to have an intermediate kind of education when speaking Yorkshire dialect by 63% and Cockney by 50% of the respondents correspondingly, and to be uneducated when speaking Southern American English by 85% of the participants. The figures, once again, prove that English accents have stigmatized social values.

Descriptions of occupation suggested by the respondents for the speakers, as should be expected, were closely connected to their social status and education. When the speaker was presumed to be from Southern Ireland and of working or lower middle class, he was supposed to have such jobs as “driver, driving instructor, pizza delivery person, technical support, call centre” and even “criminal”. On the other hand, when a speech sample was recognized as coming from a middle class person with a university degree living in Somerset, the suggested job descriptions included “philosopher, lecturer, artist, researcher, writer”, etc. Interestingly, when the respondents

heard Speaker 1 imitating a foreign (French) accent they were more reticent in their social judgment and tended to place him in the middle of the social ladder (lower and upper middle class in 73% of the responses). They were also rather at a loss when defining his professional qualifications and mentioned, among others, such inconspicuous occupations as "traveler, poet, teacher, tourist agent, student".

Perception of personal properties

Furthermore, the personality properties ascribed to the speakers by the respondents, such as character traits or even age, may be shown to depend on the social factors, associated with the accent. The speakers' age was determined by their education and occupation (which in their turn were interrelated with the social class): the higher the education of the speaker presumably was, the older he was thought to be. The age of Speaker 1 varied from 20-30 years old (in 100% of the answers) as a not very well educated (88%) working class member (77%) speaking Northern English (Manchester) to 30-40 (70%) or even 40-50 (12%) as a university-educated (73%) middle class (85%) RP speaker. According to the respondents' opinion, there were no uneducated people in the age group of 50-60 years old.

Personality traits that the respondents had to choose from the list offered in their answer sheets to describe the speakers (in the form of 8 pairs of adjectives with contrastive meanings, such as "industrious – lazy" or "introvert – extravert"), varied greatly for each speaker. In ascribing them, the respondents obviously relied not on the quality of the speakers' voices (individual timbre), but on the associations their accents have in the present day English speaking societies. The same speaker, in the opinion of the listeners, sounded responsible, considerate and generous when speaking with Liverpool accent, pushy, selfish but polite when speaking with standard pronunciation and irresponsible, lazy and sloppy when imitating Southern American speech. The stigmatized character of such evaluations is evident in the seeming unanimity of the respondents who generally coupled Yorkshire accent with being extravert and industrious, Cockney accent with being responsible, and considered a French person struggling to speak English polite.

Conclusion

The study data are consistent with the results of the earlier research into sociolinguistic values of English accents. What it emphasizes is an astonishing fact that in perceiving accented speech speakers of English concentrate almost exclusively on the social factors, and evaluation of the

personal properties is predetermined by those. This explains why, surprisingly, no respondent in the experiment noticed that the speech samples were recorded by the same 2 people imitating different English language varieties. The peculiarity of oral English speech perception can only be summed up in the slogan: they are what they sound like.

The study contributes to further understanding of sociolinguistic processes taking place in the English speaking societies and its results may help developing appropriate English language user strategies by non-native speakers.

Acknowledgements

The author would like to express sincere appreciation of Evgenia Sokolova's assistance in conducting the experiment.

References

Abramova, I.E. 2009. Phonetic variation outside the natural language environment. Petrozavodsk, Petrozavodsk University Press. (In Russian)

Coupland, N. and Bishop, H. 2007. Ideologised values for British accents. Journal of Linguistics, vol.11, issue 1, 74-93.

Labov, W. 1966. The social stratification of English in New York City. Washington, D.C., Center of Applied Linguistics.

Shamina, E.A. 2011. Subjective evaluation of the phonetic representation of some national and regional varieties of the English language. In S. Androsova (ed.), Proceedings of the 1st International Conference "Phonetics without Borders", 96-98. Blagoveshchensk, Russian Federation.

Shamina, E.A. 2012. On objectivity of subjective evaluation of some national and regional English accents. In L.A. Verbitskaya, N.K. Ivanova (ed.). Homo speaking: XXI century research, 150 – 155. Ivanovo, Ivanovo State University of Chemical Technology Press. (In Russian).

Wells, J.C. 1982. Accents of English, vol. 1 .Cambridge, Cambridge University Press.

Phonetic words duration simulation using Deep Neural Networks

Alexander Shipilo
Saint-Petersburg State University, Russia

Abstract

Deep Neural Networks (DNN) are widely used in speech prediction and speech modeling. The current paper describes the implementation of DNN for the task of duration prediction of speech units (allophones and syllables that form the structure of phonetic word, intonation phrase). It is well-known that numerous factors influence the duration of segments. However, the level of confidence of characteristics differs significantly. It was found that deep neural network that predicts allophones duration shows better results than the network that predicts the duration of syllables.

Key words: deep NN, duration modeling, phonetic words.

Introduction

One of the challenging tasks in text-to-speech systems is the problem of duration modeling of speech units. Despite recent research refers to the problem of lengthening and shortening the speech units, unit selec-tion systems demonstrate better naturalness (Lobanov, Tsirulnik, 2007, 2008).

The duration of speech segments varies significantly depending on the position within intonation unit, phonetic word, the number of elements in the speech unit (Svetozarova, 2014). Each allophone unit has its own intrinsic duration value. It is known that a lot of factors influence the segment duration.

Python Toolkit for Deep Learning (PDNN) was used in the current research. The general architecture of the developed system is shown on the fig. 1.

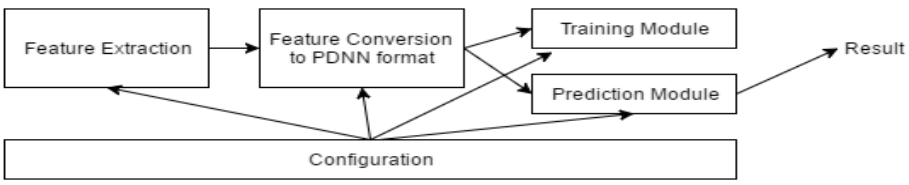


Figure 1. The architecture of the system.

Material

The Corpus of Professionally Read Speech (CORPRES) was used in the current research (Skrelin et al., 2010). During the pilot experiment the recordings of one female speaker (approx. 6 hours of speech, 155591 allophones, 61591 syllables) were chosen. Each recording has following manual checked annotation level:

F0 marks (stylized according to (Skrelin, Kocharov, 2009)
Ideal transcription
Real transcription
Word boundaries
Pitch movements
Boundaries of intonation units

CORPRES doesn't contain syllable and phonetic words annotations. To generate these levels a python script was created. Syllable boundaries were estimated according to Shcherba syllabification theory (Matusevich, 1976). According to allophone boundaries level a new level was automatically generated. It contains the allophone boundaries and boundaries of non-phonemic units (vowel insertions etc.).

The segment duration in the material is pre-processed. Each allophone segment is normalized according to the tempo coefficient estimated by

$$T = (D1 / N) / (D2 / N),$$

where T – tempo coefficient, D1 – sum of average duration values of allophones within intonation unit, D2 – sum of real durations within intonation unit.

Experiments

Four experiments were performed. The first two deal with syllable duration prediction (models 1, 2), others – with the allophones duration prediction (models 3,4). Let us consider the experiment techniques.

Unfortunately, the prediction of real duration of a segment is a rather difficult task. To simplify it, rounded values were predicted. Model 1 recognizes the percent deviation from the average of all syllables in the material, model 3 – the percent deviation of the required allophone. For example, let us consider the segment that is lengthened by 10 percent. This value was rounded to the nearest possible percent deviation value. If the required value is 110 %, the required coefficient is rounded to the nearest

possible value accurate to 25 percent (e.g. 110 to 100 %, 120 to 125 % etc.). Table 1 shows the features that were used in the model.

Models 3,4 predict the rounded number that is required to multiply by the minimum level of auditory perception that is equal to 30 ms.

Each model consists of two hidden layers, each layer contains 2048 elements.

Table 1. The features for the models.

Features	
Model 1,2	Model 3,4
	Allophone (previous, current, following)
	Number of allophones in the syllable
Syllable index from the beginning/end of phonetic word	Syllable index from the beginning/end of phonetic word
Number of phonetic words in intonation unit	Number of phonetic words in intonation unit
Number of syllables in intonation unit	Number of syllables in intonation unit
	Allophone index from the beginning/end of syllable
Phonetic word index from the beginning/end of intonation unit	Phonetic word index from the beginning/end of intonation unit
The reduction level of a syllable	The reduction level of a syllable
The pitch movements within syllable, phonetic word, intonation unit	The pitch movements within allophone, syllable, phonetic word, intonation unit

Results and discussion

Table 2 demonstrates the results of the experiments.

Table 2. Results.

Model	Prediction accuracy, %
1	20
2	51
3	45
4	79

As we can see from the table 1, model 4 shows the best result, model 1 – the worst. Models that simulate syllable durations show worse results, than models that simulate the allophone ones. Models 2 and 4 show better results (51 and 79 percents) in comparison to models 1 and 3. The reasons for it are the fact that (1) the deviation depends on the duration of the average of the target element, (2) the deviation is the relative characteristic. Let us consider the average unstressed vowel allophone (for example /u/) equals 50 ms. In this case ten percent lengthening means that the duration changes by 5 ms. On the other hand, the ten percent change of the stressed allophone of phoneme /a/ (the average duration in the material is 109 ms) means that the duration changes by approximately 11 milliseconds. If we predict the real allophone duration (models 2, 4), the problem of differences in averages disappears.

The results confirm the hypothesis that the selected features can be used as predictors of segment durations, but neural net provides no information about the rate of confidence of the features. To answer this question additional study is required.

References

Lobanov, B.M, Tsirulnik, L.I., Rules of Speech Corpus Segmentation into Phonetic Units and the Strategy of Unit Selection in Speech Synthesis, http://www.dialog-21.ru/digests/dialog2007/materials/html/60.htm

Lobanov, B.M, Tsirulnik, L.I., Computer Synthesis and Speech Cloning, Minsk, 2008 / in RussianMatusevich, M.I. 1976. Modern Russian Language. Phonetics (in Russian) Sovremennij Russkij Yazik. Phonetika

Matusevich, M.I., Modern Russian Language. Phonetics, 1976 / in Russian/ Sovremennij Russkij Yazik. Phonetika

Skrelin P., Kocharov D., Automatic processing of prosodic design of the utterance: relevant prosodic features for automatic interpretation of intonation model, 2009, AP-2009, Saint-Petersburg / in Russian.

Skrelin, P., Volskaya, N., Kocharov, D., Glotova, O., Evdokimova V. CORPRES - Corpus of Russian professionally read speech. In: Sojka, P., Horák A., Kopeček, I., Pala, K. (eds.) TSD 2010. LNCS, vol. 6231, pp., 392-399. Springer, Heidelberg (2010)

Svetozarova N.D., “Short” stressed vowels in the Russian language, Issues in Phonetics 6, 2014 / in Russian.

Transcription: what is meant by accuracy and objectivity?

Pavel Skrelin, Nina Volskaya
Department of Phonetics, Saint Petersburg State University, Russia

Abstract

The paper deals with the relationship and discrepancy between phonetic (acoustic) characteristics of the speech signal and their phonological interpretation with the aim of their reflection in segmental transcription and prosodic annotation of the speech corpora.

Key words: phonetics and phonology, transcription, speech corpora

Introduction

The presentation draws attention to the interaction between acoustic, phonetic and phonological aspects of the speech signal and their reflection in transcription. Accuracy of phonetic transcription plays an important role in the annotation of speech corpora. The requirements for precision to a great extent depend on the annotators' expertise and on what the corpus is designed for. If the corpus is to be used in TTS or ASR applications the selected phonetic signs must be as close as possible to acoustic (spectral) features of sounds analyzed in their physical boundaries. The traditional "manual" transcription is based on perception of a word or at least a syllable and represents a human model of speech perception and sound interpretation. As a result transcriptions using different methods and aimed at different applications may differ. At the same time comparison of the results of both transcription types dealt with in the presentation provides information about speech perception mechanisms on the segmental (phonetic representation of distinctive features) and suprasegmental levels (discrepancy between acoustic and perceived forms of melodic patterns).

Second paragraph text exactly the same as the first paragraph text except for the first line indentation, which is 0.6 cm for this and subsequent paragraphs (i.e. Paragraph text).

Segmental level problems

A minimal language unit for speech perception is the syllable: due to sound co-articulation distinctive features of phonemes are not limited by the boundaries of their sound realizations (allophones) proper but are represented in their phonetic environment as well. For example, a distinctive feature of softness of the Russian plosives is actually realized in the

neighboring vowels (as in the case of bilabials). Labialization of /u/ can be indicated in the preceding fricative, but may be absent from the vowel itself (as in the case of non-standard alternation of Russian phonemes /u/ - /ɨ/ we have found in CORPRESS - the Corpus of Russian Read-Aloud Speech).

This explains the use of 2 levels of representation of phonetic transcription in the corpora annotation: the first one , based on the perception of a signal fragment of a short word or syllable length (it usually corresponds to the orthoepic norm), the second one, based on the result of the perception of the sound in its physical boundaries: it reflects the sound spectral features

This method allows us to fix and describe the real situation: phoneme stream as it is perceived and interpreted by human and the same stream as it is interpreted on the basis of realized distinctive features of phonemes.

At the same time this method makes it possible to avoid solving the phonological problem, which ensues from the tensions between the abstract units (phonemes) and their material representation in the form of articulation and perception units (syllables).

Prosodic level problems

In analyzing intonation for Russian speech corpora – CORPRESS and CoRUSS (Skrelin et al. 2010; Kachkovskaia et al. 2016) – we came across situations where annotators' opinions regarding the type of a particular intonation pattern differed mostly due to the mismatch between their phonological decision and the visual acoustic representation of the intonation curve.

A few examples. In Russian, the Intonation Construction 6 (IC6) (Bryzgunova, 1970), used non-final intonation units and questions seeking repetition or clarification, is described as the (high) rising nuclear tone which levels off in the post-nuclear part. In fact, acoustically, the post-nuclear syllables form a declination line which may cover up to 4-6 semitones depending on the length of the post-nuclear part (Fig.1).

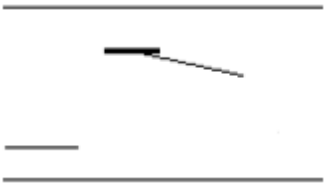


Figure 1. Schematic representation of the IC6 : nuclear syllable is marked by a bold line.

Phonologically and perceptually, though, the contour is described as "rising", and the declining part is perceptually ignored.

Another clear case for such a mismatch which complicates matters further is the use of phonetically rising-falling tone (IC3) typical for yes-no questions in Russian: though the abrupt fall on the post-nuclear part is much more prominent than in the previous case for IC6 (Fig.1) and can reach, though not necessarily, the speaker's minimum pitch level, the contour is nevertheless phonologically interpreted as rising (Fig.2).

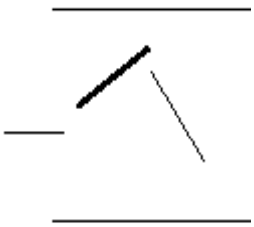


Figure 2. Schematic representation of the IC3 : nuclear syllable is marked by a bold line.

Note: For speakers of some other languages but Russian (German, English , Finnish) this contour shape is interpreted as falling. In English intonation system, for example, it belongs to the phonologically falling compex rising- falling tone , the Jackknife (O'Connor & Arnold, 1973).

This case is particularly tough both for phonological interpretation and automatic tone identification, since for any algorithm which relies on the phonetic aspect — tone-shape and F0 track only, this tone is obviously (and erroneously) falling.

Acoustically, any tone can take a number of shapes, depending on the segmental make-up of the nuclear syllable and the word itself and the location of the accented syllable proper. The case presented in Fig.3 below, shows an ambiguous situation when the tone type interpretation is unclear without postnuclear syllables, and the decision in favour of either IC6 or IC3 should be taken with other prosodic parameters in consideration, namely, the nuclear syllable duration, which is normally longer in IC 6.

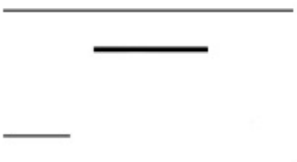


Figure 3. Schematic representation of the IC6 and IC3with nuclear syllable in the final position.

Conclusion

In real speech situation the distictive features cruicial for the phonological decision-taking may not be present in the sound itself (which may be absent altogether) but reflected in its right or /and left neighbours. This poses the problem of formal represenation of the sound stream itself in automatic interpretation (recognition) which is based on acoustic parameters of segments or F0 curves. As long as we do not exactly know how the speech signal characteristics which a person uses for phonological interpretation correlalte with its objective evidence we need to use two ways of formal representation (transcription): objective and abstract.

References

Bryzgunova E. A. 1980. Intonation [intonacija], in: Russian Grammar, N. Shvedova, Ed. Moscow: Nauka, vol. 2, pp. 96 – 122.

Kachkovskaia T., Kocharov D., Skrelin P., Volskaya N. 2016. CoRuSS - a new prosodically annotated corpus of Russian spontaneous speech. in: Proceedings of LREC 2016.

O'Connor J.D., Arnold G.F. 1973 Intonation of Colloquial English. Longman, London.

Skrelin P., Volskaya N., Kocharov D., Evgrafova K., Glotova O., Evdokimova V. 2010. CORPRES - Corpus of Russian Professionally Read Speech. in: Text, Speech and Dialogue, ser. Lecture Notes in Computer Science, P. Sojka, A. Hor´ak, I. Kopecek, and K. Pala, Eds. Springer Berlin Heidelberg, 2010, no. 6231, pp. 392–399.

Grammatical change and hindcast model statistics – A comparison between Medieval French and Brazilian Portuguese

Eduardo Correa Soares
CLILLAC-ARP, Département d'Études Anglophones, École Doctorale Sciences du Langage, Université Paris Diderot – Paris VII, France

Abstract

This paper presents a methodology to analyse the ongoing linguistic change in Brazilian Portuguese[BP] as regards the pro-drop parameter. I propose to apply a hindcast statistical regression model to a sample of data from Medieval French[MF], whose outcome is the obligatory subject use in Modern French and to compare to a sample from BP. The results suggest that the change in such languages contrasts and is related to different reasons. While the change in MF appears to have uniformly gone toward non-pro-drop parameter, the BP change seems a by-product of semantic preference of null subjects to corefer to non-animated and non-specific antecedents.

Key words: Hindcast statistical model, grammatical change, pro-drop parameter, Brazilian Portuguese, Medieval French.

Introduction

This paper proposes a new methodology to address the grammatical change regarding the pro-drop parameter in Brazilian Portuguese[BP]. I propose that statistical hindcast regression model comparing Medieval French[MF] and BP may verify whether some assumptions about BP are akin to what came ou in MF. This model is applied to two samples of data from MF and BP. The results show seemingly diverging patterns of linguistic change.

BP is taken to be a language on the way to become non-pro-drop (Tarallo 1983, Galves 1987, 1992, 1998, Duarte 1993, 1995, *inter alia*). In many standard pro-drop contexts in other Romance languages (for instance, European Portuguese, Spanish and Italian), an overt pronoun is indeed obligatory in nowadays colloquial spoken BP (see Duarte 1995, Barbosa, Duarte & Kato 2005, *inter alia*), such as in the example in (1) below.

(1) então a gente lê pra ele_1 sentado ali... *(ele_1) gosta...
So the people read.pres.3s for him seated there he like.pres.3s
"So there we read for him_1 when seated down and he_1 likes that."
(NURC-RJ, inquiry_011, data_set: "70s")

Such contexts and data lead many works to suggest that BP is changing due to the simplification of agreement marking system, the so called Taraldsen's generalization (Roberts 1993, 2014, Kato 1999, 2000, *inter alia*).

In this vein, it has been proposed that BP is following the same path by which French has passed from the MF to Modern French (notable exceptions to this claim are Kaiser 2009 and Roberts 2014). In MF, overt and null pronouns have been in apparent free variation, as in (2) below.

(2) Aucassins$_1$ s' en est tornés / (...) Vers le palais _$_1$ est alés / *il*$_1$ en monta les degrés / une canbre _$_1$ est entrés / si _$_1$ començа # a plorer
"Aucassin$_1$ departed/ to the palace he$_1$ went / he$_1$ went upstairs / (into) a bedroom he$_1$ entered / this way he$_1$ began to weep"
(SRCMF, aucassin, data_set: "XII_century")

In the next section, I put forth a hindcast model statistics, by applying inferential logistic regression to data from MF and from BP.

Methodology

I propose to use a hindcast model to compare the change regarding the pro-drop parameter in BP and French. This methodology consists of (i) analysing a set of data from a specific period of time whose outcome is already known; (ii) statistically describing what has taken place and testing for some parameters and (iii) predicting possible similarities and differences from another set of data by changing or adding one or more parameters.

I have analysed MF change (Adams 1987) whose outcome has been the non-pro-drop status of modern French. I have compared this hindcast analysis of MF data to BP data in order to evaluate the status of the current so-called "on-going" change in BP. I have taken 9 texts from the historical corpus of MF *SRCMF*[1], 6 interviews of BP *NURC-RJ* corpus (3 carried out in the 70s and 3 in the 90s)[2] and 3 movie subtitles produced after 2010 from the *OPUScorpora* project[3]. These texts were automatically annotated. The sample was gathered by a concordance toolkit. The MF subcorpus was thus constituted of 1500 sentences (a half of them without subject), distributed into 3 subsets of data according to the year of the text (*group1*, the IXth and Xth centuries; *group2*, the XIth and XIIth centuries; and *group3*, from the XIIIth century on). The BP corpus was equally formed by 1500 sentences (50% of subjectless sentences) and split into 3 subsets: *group1*, data from 70s; *group2*, from the 90s; and *group3*, data from 2010 on. The collected data was then analysed with a Generalized Linear Model using the software R, with the packages *lme4*, *languageR* and *stats*.

Results

Table 1 sums up the logistic regression analysis and the results obtained. In French, the so-called impoverishment of agreement marking has predominantly affected singular forms and 3rd person plural. The fixation of non-pro-drop in MF is taken to be a strong effect of such an impoverishment

(Adams 1987, Roberts 1993a). If Taraldsen's generalization is correct, it is expected that 1st and 2nd person plural are significantly more null subjects than the others. But this prediction does not hold. What the data have shown is a gradual increasing in the number of overt subjects regardless the verbal inflection, and no significant difference along the time and among the person markings. In BP, however, the number of null subjects is stable across the discourse persons and the periods in the last 50 years, except for 1st and 3rd person singular. In a further statistical regression, I have analysed the features animacy and specificity (previously suggested in the literature about BP by Cyrino *et al*. 2000). In MF, these features were not significant in any statistical regression. In BP, however both non-animated and non-specific are relevant in the increasing number of 3rd person singular null subjects (P-value: 0.00615 and 0.00771 respectively).

Table 1. Logistic regression analysis of Medieval French (MF) data and Brazilian Portuguese (BP) data (int = intercept term)

period/person	1_sing	2_sing	3_sing	1_pl	2_pl	3_pl
group1	int	int	int	int	int	int
group2	MF(ns) BP(.)	MF(ns) BP(ns)	MF(ns) BP(**)	int	MF(ns) BP(ns)	MF(ns) BP(ns)
group3	MF(ns) BP(*)	MF(ns) BP(ns)	MF(ns) BP(***)	int	MF(ns) BP(ns)	MF(ns) BP(ns)
Signif. codes: Pr(>|z|) 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ns' 1						

Discussion

This pilot corpus study suggests that the null subjects in BP are becoming scarcer in a way different from MF. Firstly, the null subject in BP is crucially likely to be 3rd person singular. This person is the less marked form in BP (Kato 1999). In MF, no significant difference concerning person, animacy or specificity was found. The MF change can also be related to other factors (e.g. the use of clitic subject pronouns). In BP such a difference seems to be a semantic-functionally motivated by-product of two factors – the semantic features animacy and specificity. This difference can be crucial to shed light on the *partial* pro-drop status of BP (Biberauer *et al.* 2010) and the *non*-pro-drop status of modern French (Adams 1987).

Notes

1. See http://srcmf.org and Prévost & Stein (2013) for more information.
2. This corpus is available online in http://www.letras.ufrj.br/nurc-rj/.
3. See http://opus.lingfil.uu.se/OpenSubtitles2016/ and Lison & Tiedmann (2016).

Acknowledgements

I am thankful to CAPES Foundation for providing me the financial support of this research and to my supervisors Philip Miller, Barbara Hemforth and Sergio Menuzzi, who give me *to-the-point* advice to carry out my projects.

References

Adams, M. 1987. From Old French to the Theory of Pro-drop. Natural Language and Linguistic Theory 5: 1-32.

Barbosa, P., Duarte, M. E. L. & Kato, M. A.. 2005. Null subjects in European and Brazilian Portuguese. Journal of Portuguese Linguistics 4. v. 2: 11-52.

Biberauer, T. Holmberg, A. Roberts I., Sheehan, M. 2010. Parametric Variation. CUP.

Cyrino, S. M.L.; Duarte, M.E. L., Kato, M. A. Visible subjects and invisible clitics in Brazilian Portuguese. In: Kato, M.A. & Negrão, E.V. (eds.). 2000. 55-104.

Duarte, M. E. L. 1993. Do pronome nulo ao pronome pleno. In Roberts, I., Kato, M.A. (eds.): 107-28.

Duarte, M.E.L. 1995. A Perda do Princípio "Evite pronome" no Português Brasileiro. Campinas, SP, UNICAMP: Ph.D. Dissertation.

Galves, C. 1987. A sintaxe do português brasileiro. Ensaios de lingQistica13: 31-50.

Galvez, C. 1993. O enfraquecimento da concordância no Português Brasileiro. In: Roberts & Kato (eds.): 387-408.

Galves, C. 1998. Tópicos e sujeitos, pronomes e concordância no português do Brasil. Cadernos de Estudos Lingüísticos, 34: 19-32.

Kaiser, G. A. 2009. Losing the null subject. A contrastive study of (Brazilian) Portuguese and (Medieval) French. In Proceedings of the Workshop Null-subjects, expletives, and locatives in Romance: 131–156.

Kato, M. A. 1999. Strong pronouns, weak pronominals and the null subject parameter. PROBUS, 11, 1: 1-37.

Kato, M. A. 2000 The partial pro-drop nature and the restricted VS order in Brazilian Portuguese. In: Kato, M. A. & Negrão, E. V. (eds). 2000. 223-258.

Kato, M. A. & Negrão, E. V. (eds). 2000. The Null Subject Parameter in Brazilian Portuguese. Frankfurt-Madrid: Vervuert-IberoAmericana.

Lison, P., Tiedemann, J. 2016. OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. In Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)

Prévost, S.; Stein, A. 2013. Syntactic Reference Corpus of Medieval French (SRCMF). ENS de Lyon/ILR Stuttgart.

Roberts, I. 2014. Taraldsen's Generalization and Language Change. Prepublished ms.

Roberts, I. & Kato, M. A.(eds.) 1993. Português Brasileiro: uma viagem diacrônica (Homenagem a Fernando Tarallo). Campinas, SP: Editora da UNICAMP.

Tarallo, F. 1983. Relativization Strategies in Brazilian Portuguese. University of Pennsylvania: Ph.D. Dissertation.

The phonetics of Russian North Bylinas

Svetlana Tananaiko[1], Marina Agafonova[2]
[1]Department of Phonetics, Saint-Petersburg State University, Russia
[2] Department of Phonetics, Saint-Petersburg State University, Russia

Abstract

The Internet site presenting the bylinas of Russian North from Sound Records Archives of Institute of Russian Literature was created in 2014. The aim of the site is to give free access to the unique Russian folklore sound records, made throughout XX century, for everybody interested, especially those who study anthropology, folklore, dialects and dialect phonetics of Russian, because on this site the presented sound fragments are analyzed in all these aspects. The article describes revealed phonetic characteristics of North bylinas and suggests a theoretical interpretation of the dynamics of the dialect phonetics changes.

Key words: dialect phonetics; Northern Russian dialect zone; Russian North bylinas

The Internet Presentation of Russian Bylinas

The Internet site "Corpus of Russian Folklore. Bylinas" presenting the bylinas of Russian North from Sound Records Archives of Institute of Russian Literature was created in 2014. The aim of the site is to give free access to the unique Russian folklore sound records, made throughout XX century, for everybody interested, especially those who study anthropology, folklore, dialects and dialect phonetics of Russian, because on this site the presented sound fragments are analyzed in all these aspects.

The Phonetics of Russian North

"Northern Russian Dialects, spread in the North and North-East of European Russia, and also in some regions of Siberia, have preserved a lot of archaic sound characteristics, which disappeared not only in Standard Russian, but also in Middle Russian and South Russian dialects. It can be explained both by linguistic and extra linguistic reasons, as well as by the specificity of colonization of Northern Russia, its peasant economy and peasant everyday life." (Tananaiko 2001).

The most specific characteristic of Northern Russian dialect zone is the set of phonemes different from other dialect zones. In the archaic Northern Russian dialect phonemic system there are more vowels and less consonants than in Standard Russian, what can be explained, on the one hand, by the preservation of special phonemes replacing etymological *yat* and ancient */o/* under ascending tone, and

on the other hand, by the absence of /š′:/ and the presence of only one affricate instead of two (Avanesov 1949).

Phonetic realization of vowels and consonants, even those common with Russian Standard, is in these dialects different from the Standard. The vowels are diphthongs or diphthongoids, the palatalized sibilants are lisping and so on (Meshchersky 1972).

The rules of phoneme distribution and the rules of alternations are also different from the Standard. For example, the unstressed vocalism retains unstressed /o/ and /e/, in the consonant system there are consonance simplifications (/mm/ instead of /bm/, /s′/ instead of /s′t′/) (Kolesov 2006).

Figure. 1 The example of the presentation of a speech portrait of a bylina performer on the “Corpus of Russian Folklore. Bylinas. Sound Analogue” Internet site page.

The Material: Russian North Bylinas

All these and many others Northern dialects characteristics can be traced in various texts, but in bylinas, traditional Russian heroic epics, known for preservation of the most established, typically traditional texts, least subject to changes because of their essential genre property, these characteristics are supposed to manifest themselves in their brightest. Apparently the very bylina genre, which belongs to folklore tradition, and is not supposed to contain any modern vocabulary, promotes the strongest possible retention of archaic phonetic features.

Bylina, known for preservation of the most established, archaic folklore language formulas and corresponding archaic language features, was developing in Russian folklore as a live genre until the middle of the XX century. So in spite of traditionalism, the bylinas phonetics couldn't help being influenced by modern linguistic processes which have been changing the phonetic aspect of folklore heritage.

The bulinas under study were recorded in Pechora and Mesen' region, both belonging to Pomor dialect group of Northern Russian dialect zone. The majority of these dialects retain the archaic pronunciation features, which are especially significant for the phonetic characteristic of these bylinas. The influence of modern language manifests itself in inconsistent realization of dialect features, random phonetically unjustified substitutions of some forms for others, that is, in the effects which usually indicate the destruction of the integral phonetic system of a dialect.

The Results: Phonetical Features of the Bylinas

The study of the material revealed a set of phonetic features, typical for these regions. It's important to mention that only segmental features were studied, because the traditional melodiousness, which is quite essential for the performance of bylinas, prevented any prosodic analysis. Generally the preservation and stability degree for the phonetic features characterizing the archaic northern dialects, is quite high.

The most persistent dialect features registered in the studied records were the retention of unstressed /o/ and /e/ – these two characteristics, which are essentially basic for Northern Russian vocalism, can be explained by the type of word stress, word rhythmic and the absence of vowel reduction in Northern dialects. The most persistent consonant feature is the specific use of affricates, when there is only one affricate in a dialect, either palatalized or velarized. These are the characteristics which are inevitably present in the pronunciation of all recorded performers. Also there are several morphological features, but their appearance is not as regular as of aforesaid

characteristics, and each performer has their own set of morphological features.

From the point of view of theoretic linguistics, the results can be explained in the following way.

Language, perceived as a system of signs, is a complex self-organizing system, getting in the process of its development a definite functional structure and functioning as an entity. During the transition to new formations, for example during currently perceived swift, caused by various factors, destruction of Russian dialects, in the developing systems, that is, in the systems of language units of different levels, fluctuation amplitude growth will be traced, and it makes dialect systems, especially on phonetic level, rather chaotic, causes disappearance or neutralization of inherent phonetic oppositions and prosodic characteristics.

The results of the study show that the integral language system of these dialects is currently under destruction, and on phonetic level it is right to describe not a dialect phonetic system as a single whole, but to enumerate separate phonetic features, characterizing not only Pomor group dialects, but all the dialects of Northern dialect zone. The preservation of the very features, which are most inherent, most basic for the whole Northern dialect zone, reflects the progressive destruction of earlier phonemic oppositions.

In the end the various degrees of preservation and stability of different elements of the destroying phonetic systems demonstrate that even in faraway small Northern villages, where the records were made and where the performers were senior people, who recited the archaic text full of traditional formulas, - even there we witness the fast inevitable modifications of the sound form of speech, that reflect the language functioning dynamics in dialects.

References

Avanesov R. 1949. Essays of Russian Dialectology. Moscow.

Corpus of Russian Folklore. Bylinas. Sound Analogue: Internet site. URL: http://www.zvukbyliny.pushkinskijdom.ru/.

Kolesov V. et al. 2006. Russian Dialectology. Moscow.

Meshchersky M. (ed.) 1972. Russian Dialectology. Moscow.

Tananaiko S. 2001. Russian Dialects in Non-Slavonic Surrounding. In Verbitskaya L., Vasilkova V., Kozlovsky V., Skvortsov N. (eds.), Comparative Collection: Miscellany of Sociological and Humanitarian Studies, 173-185, Saint-Petersburg.

Association experiment in practice of linguistic and cultural dominants research

Svetlana Takhtarova[1], Diana Sabirova[2]
[1]Dept of Theory and Practice of Translation, Kazan Federal University, Russia
[2]Dept of European Languages and Cultures, Kazan Federal University, Russia

Abstract

The paper is devoted to experimental definition of the changes happening in structure of cultural dominants of the German ethnosociety on the example of a linguistic and cultural concept of *Ordnung.* To provide well-grounded conclusions on the status of the problem and determine the axiological characteristics of the concept the authors carried out an associative experiment. The respondents were asked to write several words to the given words incentives. The experiment confirms that cultural constants are dynamic formations which bound to change. The changes characteristic of *Ordnung* as a cultural dominant inevitably involve modification of the German communicative style that is shown, in particular, in greater tolerance to deviations from norms and standards, smaller degree of criticality and straightforwardness.

Key words: associative experiment, concept, cultural dominants.

Introduction

Cultural concepts, representing the most important category of cultural linguistics, are actively studied as exemplified in the material of different languages and cultures. The main characteristic of linguocultural concept is, as it is well known, its value component (Karasik 2004). The culture dominants, most important concepts for a given culture, constitute the core value of worldview peculiar to a specific culture.

The *Ordnung* concept, which is the subject of this article, traditionally considered as one of the key cultural landmarks of the German ethnosociety (Bartminsky 2005, Medvedev 2007, Ter-Minasova 2007, Markowsky 1995, Matussek 2006). Vezhbitska notes that Germans should have *Ordnung* (order) and live in a world where Ordnung "reigns". In fact, only *Ordnung*can guarantee their inner peace (Wierzbicka 1999). According to Bauzinger untranslatability of German words *Ordnungsamt, Ordnungswidrigkeit, Ordnungsstrafe, ordnungspolitische Massnahme* proves the order concept to be of idioethnic character in German society. In this context the order is not only a social principle, limiting every single person to a particular behavioral pattern or framework, but also a norm, which every person adheres to without any coercion (Bausinger 2002). At

the same time, the cultural dominants, despite its rigidity, can change over time, similarly to the way the culture and the society evolve.

Materials and Methods

We conducted an open associative experiment to identify how *the Ordnung* concept is understood in modern German society and to determine its axiological characteristics. The experiment features 120 informants that we provisionally divided into three age groups: young people and students (20-27 years old), employed respondents (28-60 years) and senior citizens. During the experiment, respondents were asked to write a few words they associate with the word-stimulus *Ordnung*. Besides association questionnaire respondents were offered evaluation questionnaire, in which they had to indicate their attitude to the word-stimulus as "+" - positive, "-" - negative or "0" - indifferent. Thus, the purpose of the experiment, which involves determining the value component of the concept under consideration, both explicitly, through informant directly evaluating the given concept, but also implicitly, through the analysis of obtained during the experiment associations to a given word-stimulus.

Results

The conducted experiment has allowed to define the following features of Ordnung concept.

Firstly, most responses given by elderly people, i.e. third age group, constituting associations they have given to the word-stimulus represent axiomatic phrases and clichés: *Ordnung muss sein* (31%) and *Ordnungist das halbe Leben* (26%). It is indicative, in our view, that such phrases appear only sporadically in the responses of informants representing the first and the second group.

Secondly, such verbal responses as*wichtig, notwendig, sehrwichtig, sehrpositiv* were given by the representative of the third group, thus, confirming normative-evaluative nature of the analyzed concept. The responses of the informants comprising the first and the second group are way less "axiological" - 4% and 12%, respectively. Moreover, verbal responses submitted by youth group respondents reflect not only the positive but also the negative perception of the stimulus-word: *einschränkend, überschätzt, bremst Kreation, Druck, nichtimmer*. In general, negative associations are insignificant (16%), but their presence in the responses of young respondents is, in our opinion, of symptomatic character.

Thirdly, many informants of the youth group associate *Ordnung* with purity and establishing order, which is evidenced by the following, rather frequent responses: *Sauberkeit, Sauber, Aufräumen, Zimme*r. Similar words

are given by the representatives of the second age group, although much less frequently. For the older generation the order is associated primarily with the "mental" order and structured and well-organized life: *Gedanken, Sicherheit im Leben.* Confirmation that is The fact that in many questionnaires informants of this group provided not only single words as responses to the word-stimulus, but detailed answers confirm the idea that *Ordnung*is perceived by the oldest age group as an immutable value.
Ich liebe sie, weil sie das eigene Leben und das der anderen erleichtert; sie sollte anzustreben sein, um besser zuleben; notwendig, um in eigener Umweltbestehen zu können.

Fourthly, the associations of the youth group have been more varied and diverse in terms of semantics. Thus, in particular, the responses of this group contain following words, which are absent in the response given by the other two groups of informants: *Hierarchie, Gleichmäßigkeit, Planung, Organisation, Recht, Organisiertheit, Struktur, Kalender, Eltern*. The last word-response is probably due to the fact that the order is instilled by parents and children education begins, first of all, with meeting their own room cleanness requirements. Thereby, the associations are closely connected with the above-named frequent responses given by the representatives of the group, denoting the cleaning and order. Connection with the cleaning procedure is peculiar to responses of the informants from the second group, evidenced by the following associations: *Putzfrau, Schreibtisch, Zimmer, Schrank.*

It is noteworthy that unlike antonymy synonymy is not relevant element in the responses of all three groups of respondents. Antonymous verbal responses like *Chaos, Unordnung* were registered only sporadically.

Thus, the concept of Ordnung, remaining the culture dominant is undergoing some changes in its content and value components. In particular, it can be argued that for the younger generation, this concept has a more utilitarian, practical significance. Associations given by the representatives of student-youth groups have far fewer positive words, which indicates a change in the axiological component of the analyzed concept. Proof of this are the results of the axiological survey, which are, in our opinion, very significant in this respect. In particular, it was found that for the vast majority of informants of the oldest group *Ordnung* concept has positive connotation - 98% of respondents demonstrated their positive attitude towards this concept.

Answers of the second group are not so unambiguous - 56% defined their attitude to the order as positive and 44% as neutral.

Attitude of informants from the youth group to the *Ordnung* concept proved to be most ambivalent: positive attitude to the order shown by 48%

of the respondents, neutral - 44%. 8% of informants in this group defined their attitude towards this concept as a negative.

Conclusions

The conducted experiment allows for a conclusion that cultural constants represent dynamic formations, content of which may change reflecting alterations in the systems of values specific to a particular ethnosociety. In this context, the study dedicated to the study of the value component of lingocultural concepts is of particular importance, as the results of such studies are relevant for establishing and sustaining effective cross-cultural communication.

References

Karasik, V.I. 2004. Language Circle: Personality, Concepts, Discourse. Moscow, Gnozis.

Bartmin'skiy, Ye. 2005. Language imageofthe World: Essays on Ethnolinguistics. Moscow, Indrik.

Medvedeva, T.S. 2007. Representation of *Ordnung*concept in German linguistic picture of the world. Herald of Udmurskiy University. Philology, 5(2).

Ter-Minasova, S.G. 2007. War and Peace of Languages and Cultures: Theory and Pracitice. Moscow, Astrel': Khranitel'.

Markowsky, R. 1995. Studienhalber in Deutschland: interkulturelles Orientierungstraining für amerikanische Studenten, Schüler und Praktikanten. Heidelberg, Asanger.

Matussek, M. 2006. Wir Deutschen. Warum uns die anderen gern haben können. – Frankfurt /Main, S. Fischer Verlag

Wierzbicka, A. 1999. Semantic Universals and Language Description. Moscow, Yazyki Russkoy Kul'tury.

Bausinger, H. 2002. Typisch deutsch. Wie deutsch sind die Deutschen? München, Beck HG - Verlag.

Filled pauses and lengthenings detection using machine learning techniques

Vasilisa Verkhodanova, Vladimir Shapranov, Alexey Karpov
SPIIRAS, Saint Petersburg, Russia

Abstract

This paper addresses the issue of filled pauses and lengthenings detection and classification in Russian using machine learning techniques, such as ELM. We use such parameters as formants and energy variation and MFCC coefficients. The experiments on FPs detection and classification, that are carried out on the joint material of SPIIRAS task-based dialogs corpus, Russian casual conversations from Binghamton Open Source MultiLanguage Audio Database, reports from the appendix No5 to the phonetic journal "Bulletin of the Phonetic Fund" belonging to the Department of Phonetics of Saint Petersburg University and small part of SWITCHBOARD corpus. For evaluation of the experiments results we calculate the F1 score. The best achieved F1 score was 0.42.

Key words: speech disfluencies, filled pauses, spontaneous speech processing, Russian, ELM

Introduction

The need of detecting speech disfluencies automatically emerged mainly from the problems of automatic speech recognition (ASR): disfluencies are known to have an impact on ASR results, they can occur at any point of spontaneous speech, thus they can lead to misrecognition or incorrect classification of adjacent words. Since the INTERSPEECH 2013 Computational Paralinguistics Challenge (ComParE) (ComParE, 2013) appeared a lot of works on detection of fillers using the different machine learning approaches, since ComParE raised interest in automatic detection of fillers providing a standardised corpus and a reference system.

In (Medeiros et al., 2013) authors focused on detection of filled pauses basing on acoustic and prosodic features as well as on some lexical features. Experiments were carried on a speech corpus of university lectures in European Portuguese Lectra. Several machine learning methods have been applied, and the best results were achieved using Classification and Regression Trees: for detecting words inside of disfluent sequences performance was about 91% precision and 37% recall, when filled pauses and fragments were used as a feature, without it, the performance decayed to 66% precision and 20% recall. In (Prylipko et al., 2014) authors presented a method for filled pauses detection using an SVM classifier, applying a Gaussian filter to infer temporal context information and performing a

morphological opening to filter false alarms. For the feature set authors used the same as was proposed for ComParE (ComParE, 2013), extracted with the openSMILE toolkit (Eyben et al, 2010). Experiments were carried out on the LAST MINUTE corpus of naturalistic multimodal recordings of 133 German speaking subjects in a so called Wizard-of-Oz (WoZ) experiment. The obtained results were recall of 70%, precision of 55%, and AUC of 0.94.

Though evidence on filled pauses and lengthenings (further jointly referred as FPs) differs across languages, genres, and speakers, on average there are several disfluencies per 100 syllables, filled pauses being the most frequent disfluency type (O'Connell et al., 2004). In Russian speech filled pauses and lengthenings (jointly referred as FPs later on) occur at a rate of about 4 times per 100 words, they also occur at approximately the same rate inside clauses and at the discourse boundaries (Kibrik et al., 2014). In this paper we present the results of machine learning experiments on detection of FPs on the mixed and quality diverse corpus of Russian spontaneous speech with a addition of 20 minutes from SWITCHBOARD (Godfrey et al, 1992).

Corpus

The corpus we use for the experiments comprises various material. There are dialogs collected in St. Petersburg in the end of 2012 - beginning of 2013 (Verkhodanova et al., 2014). This part consists of 18 dialogs from 1.5 to 5 minutes, where people in pairs fulfilled map and appointment tasks. Participants were students: 6 women and 6 men from 17 to 23 years old with technical and humanitarian specialization. Recordings were annotated manually into different types of disfluencies, the FPs being the majority - 492 phenomena (222 filled pauses and 270 lengthenings). There are also recordings from Multi-Language Audio Database (Zahorian et al., 2011), that consists of approximately 30 hours of sometimes low quality, varied and noisy speech in each of three languages, English, Mandarin Chinese, and Russian taken from open source public web sites, such as http://youtube.com. From the Russian part we have taken the random 6 recordings of casual conversations (3 female speakers and 3 male speakers) that were manually annotated into FPs (284 FPs:188 filled pauses and 96 sound lengthenings). There are also12 recorded scientific reports (linguistics, logic, psychology, etc) from the appendix No5 to the phonetic journal "Bulletin of the Phonetic Fund" belonging to the Department of Phonetics of Saint Petersburg University (Dep. of Phonetics). They were all recorded in 70s-80s in Moscow except one that was recorded in Prague. All speakers (6 men and 6 women) were native Russian speakers. The number of manually annotated FPs is 285 (225 filled pauses and 60 lengthenings). Another part we added for making our corpus more quality diverse is the records from the SWITCHBOARD corpus (Godfrey et al., 1992): 3 telephone dialogues,

approximately 6 minutes each. The number of manually annotated FPs is 113 (67 filled pauses and 46 lengthenings). In total, the data set we used is about 2.5 hours and comprises 1174 FPs, the duration of a single FP lies between 9ms and 2.3s, the average duration is 360ms.

Experiments on FPs detection using ELM

In this study we describe experiments on FPs detection using the Extreme Learning Machines (ELM), a particular kind of Artificial Neural Networks that solve classification and regression problems. We used the Python ELM implementation described in (Akusok et al., 2015), number of sigmoid neurons was 600.

The feature set used in the experiments consisted of 21 standard deviations (for F0 and first three formants, energy, voicing probability and its derivative, 14 MFCC coefficients), and of 3 mean values (for energy, voicing probability and its derivative). The formants value was taken from Praat (Boersma et al., 2016) and all other parameters – from openSMILE (Eyben et al., 2010). Parameters were calculated in a window of 100ms with a 10ms step, and within each window we calculated standard deviation for every parameter from the feature set and mean value for energy.

To create train and test sets out of the data we selected random 10% of the data for test set, and the rest was used as the train set. This operation was performed 10 times producing 10 different pairs of train and test sets. The data has been separated into two classes: “FPs” and “Other”. Since the classes were not balanced (there were about 12 times more “Other” instances than FPs ones) we downsampled the train set to avoid the bias towards the class “Other” (Prylipko et al., 2014). Thus we created subset containing randomly chosen 8% of the instances of the class “Other” and all the FPs data. To train the classifier we use this downsampled training set.

ELM method yields a real number for every sample that was classified as a FP event if this number exceeded a certain threshold. This threshold was determined by a grid search in a way maximizing the F1 score on training set. As the result we achieved F1 score of 0.42.

Conclusion

In this paper we presented experiments on detection of filled pauses and lengthenings using acoustic-only features for machine learning classification (Extreme Learning Machines). For the experiments we used diverse material differing in quality, recording sites and situations. The feature set consisted of 21 standard deviations (for F0 and first three formants, energy, voicing probability and its derivative, 14 MFCC coefficients), and of 3 mean values

(for energy, voicing probability and its derivative). As the result we achieved F1 score of 0.42.

Acknowledgements

This research is supported by the grant of Russian Foundation for Basic Research (project No 15-06-04465).

References

Akusok, A., Bjork, K. M., Miche, Y., Lendasse, A. 2015. High-performance extreme learning machines: a complete toolbox for big data applications. Access, IEEE, 3, 1011-1025.

ComParE INTERSPEECH: Computational Paralinguistic Challenge, 2013. http://emotion-research.net/sigs/speech-sig/is13-compare

Department of Phonetics of Saint Petersburg University. http://phonetics.spbu.ru/

Prylipko, D., Egorow, O., Siegert, I., Wendemuth, A. 2014. Application of Image Processing Methods to Filled Pauses Detection from Spontaneous Speech. In Proc. of INTERSPEECH 2014, 1816-1820, Singapore.

Eyben, F., Wollmer, M., Schuller, B. 2010. OpenSMILE: the Munich Versatile and Fast Open-Source Audio Feature Extractor. In Proc. 18th ACM International conference on Multimedia, 1459-1462.

O'Connell, D., Kowal, S. 2004. The History of Research on the Filled Pause as Evidenceof the Written Language Bias in Linguistics. Journal of Psycholinguistic Research, vol. 33(6), 459-474.

Kibrik, A., Podlesskaya, V. (eds.). 2014. Rasskazy o Snovideniyah: Korpusnoye Issledovaniye Ustnogo Russkogo Diskursa [Night dream stories: Corpus study of Russian discourse], Litres.

Godfrey, J.J., Holliman, E.C., McDaniel, J. 1992. SWITCHBOARD: Telephone Speech Corpus for Research and Development. In Proc. of International Conference on Acoustics, Speech, and Signal Processing (ICASSP-92). vol. 1, 517-520.

Verkhodanova, V., Shapranov, V. 2014. Automatic Detection of Filled Pauses and Lengthenings in the Spontaneous Russian Speech. In: Proc. 7th International Conference Speech Prosody, 1110-1114, Dublin, Ireland.

Zahorian, S.A., Wu, J., Karnjanadecha, M., Vootkur, C.S., Wong, B., Hwang, A., Tokhtamyshev, E. 2011. Open-Source Multi-Language Audio Database for Spoken Language Processing Applications. In Proc. INTERSPEECH 2011, pp. 1493-1496, Florence, Italy.

Boersma P., Weenink D. 2016. Praat: doing phonetics by computer [Computer program]. Version 6.0.11, retrieved 20 January 2016 from http://www.praat.org/

Psycholinguistic evidence for the composite group

Irene Vogel, Angeliki Athanasopoulou
Department of Linguistics and Cognitive Science, University of Delaware, USA

Abstract

It is widely accepted that speech is phonologically structured in terms of phonological constituents composing a Prosodic Hierarchy (PH). There is less consensus, however, regarding the constituents themselves. We focus here on the controversy surrounding a prosodic constituent between the Phonological Word and the Phonological Phrase, the Clitic Group in (Nespor and Vogel 1986/2007). While in some analyses it has been excluded, elsewhere it has been replaced by a revised Composite Group(κ) (Vogel 2009). Here we present psycholiguistic data from language acquisition and adult speech production that support the existence of κ across languages.

Key Words: language acquisition, speech encoding, phonological word, composite group

Introduction

The Composite Group (κ), which has replaced the Clitic Group, is the most controversial constituent in the PH, and in fact, it is often excluded. The κ consists of a Phonological Word (ω) and certain affixes and/orfunction words, and possibly additional ωs in the case of compounds. It thus provides a constituent between the Phonological Word (ω) and Phonological Phrase (φ) which may serve as the domain of phonological phenomena across languages. The data presented below provide independent support for the κ based on two types of psycholinguistic studies, language acquisition and language processing. We first discuss the acquisition of prosody in English and Greek, and then speech processing studies in Dutch, Italian, Romanian, and Nepali.

Acquisition of Prosody and the Composite Group

It has been argued that the acquisition of prosodic phenomena proceeds according to the PH, from lower to higher constituents (Athanasopoulou 2016, Demuth & Fee 1995, Vogel & Raimy 2002). Thus, the acquisition of phenomena that involve constituents in the range of the κ can provide evidence with regard to the presence of this constituent in the PH between the ωand φ. That is, if the κ exists, we should observe a developmental order of ω→κ→φ phenomena.

It has been demonstrated that English word stress is acquired quite early, around age 2 (Kehoe et al. 1995), while compound and phrasal stress (e.g.,

green*house* vs. *green* ***house***) are not fully acquired until the age of 11years or later (Athanasopoulou 2016, Shilling 2010, Vogel & Raimy 2002).Interestingly, the production of phrasal stress is mastered after compound stress.Thus, we can place the acquisition of compound stress between that of the ω and φ, providing support for an intermediateκ constituent. The acquisition order is thus as predicted:ω→κ→φ.

The acquisition of Greek compound(ω), clitic (κ), and phrasal stress (φ) further supports the presence of κ in the PH.Stress in compounds (e.g., *kokinomális* "redhead") is acquired first, at the age of 6 (Athanasopoulou 2016) and possibly earlier (Tzakosta & Manola 2012) whilephrasal stress (e.g., *kókina****maliá*** "red hair") is acquired last (Athanasopoulou 2016). Clitic stress (e.g., *kípeló tis* "her cup"; compare with *kípelo* "cup") appears as early as 2 years (Tzakosta 2004), but it is not fully acquired until later, crucially, after compound stress and prior to phrasal stress (Athanasopoulou 2016). This three-step acquisition sequence provides further support for the κ constituent and matches the one we saw in English: ω→κ→φ.

Table 1 summarizes the findings regarding the order of acquisition of the different prosodic patterns in English and Greek. The results support the claim that prosodic development follows the PH and crucially, they show that the presence of theκ between the ω and φ is necessary to account for the order of acquisition of these prosodic phenomena.

Table 1. Prosodic patterns tested and predictions

PH	English	Greek	Order of acquisition
φ	phrasal stress	phrasal stress	third
κ	compound stress	clitic stress	second
ω	word stress	compound stress	first

The Composite Group as Speech Encoding Unit

It has been proposed that planning for speech production is based on the Phonological Word, as opposed to a lexical word or other morpho-syntactic constituent (Levelt 1989). This predicts that the encoding time for speech strings with more ωs will be longer than those with fewer ωs. To test this prediction, the encoding time for a range of constructions, measured as the reaction time (RT) of the participants to a stimulus, is compared to the encoding time for a baseline stimulus.

Wheeldon and Lahiri (1997, 2002) examined the RT of Dutch speakers producing utterances consisting of a generic structure (e.g., *Ik zoek* 'I seek') followed by nothing(baseline) and structures with either one or two ωs: (i) one ω structures included full pronouns (e.g., *het* 'it'), lexical words(e.g., *water* 'water'), and clitic + nounstructures (e.g., *het water* 'the water') and

(ii) two ωs structures included compounds (e.g., *oog lid* 'eyelid')and phrases(e.g., *ver water* 'fresh water'). Based on the proposal above, the prediction was that the structures with oneωwould show shorter RTs than the baseline, while the structures with twoωs would show longer RTs. Contrary to the prediction, however, only phrases had longer RTs, while the compounds had similar RTs to the structures with oneω (i). Analogous RT patterns have also been found for Italian, Romanian, and Nepali where compounds and clitic structures had similar encoding times to single words rather than to phrases (Vogel and Wheeldon 2010, Vogel and Spinu 2009, Koirala 2012). Table 2 summarizes the results for all the languages (NT = not tested).

Table 2. RTs to different structures in comparison to the baseline.

			RT (vs. baseline)			
Structures	# ωs	# κs	Dutch	Italian	Romanian	Nepali
Words (full pronouns)	1	1	shorter	shorter	shorter	shorter
Clitic structures	1	1	shorter	NT	shorter	NT
Compounds	2	1	shorter	shorter	shorter	shorter
Phrases	2	2	longer	longer	longer	longer

Overall, we see the same pattern: compounds behave like single words while clitics do not significantly increase the encoding time. One account for this pattern is to reassess the structure of compounds as a single (recursive) ω' despite their internal composition with two ωs (Wheeldon and Lahiri 1997, 2002).This would not only alter the definition of prosodic constituents, but it would also obscure structural and other phonological distinctions, resulting in serious drawbacks (Vogel 2009). On the other hand, if the κ is included in the PH, the results can be simply accounted for avoiding these drawbacks: it is the number of κs, not ωs, hat determines the encoding time. As we can see in Table 2, this account yields the correct predictions for all the structures, since the κ could have one ω (e.g., clitic structures) or two ωs (e.g., compounds). Overall, we see that having a constituent between ω and φ explains better the encoding time patterns across languages.

Conclusions

In the present paper, we synthesized the findings from several studies in language acquisition and speech processing to assess the psychological reality of the controversial κ constituent in PH. The results from both groups of studies demonstrate that the observed behaviors are best accounted for if an intermediate constituentκis included in the PH between ω andφ. Thus, while there is theoretical controversy regarding the κ, psycholinguistic

findings from language acquisition and speech encoding in several languages provide independent support for this constituent in the PH.

References

Athanasopoulou, A. 2016. *Prosodic development in Greek and English.* University of Delaware: Doctoral dissertation.

Demuth, K. and Fee, J. 1995. *Minimal Words in early phonological development.* Brown University & Dalhousie University.

Kehoe, M., Stoel-Gammon, C., and Buder, E. 1995. Acoustic correlates of stress in young children's speech. *Journal of Speech and Hearing Research 38,*2, 338-350.

Koirala, C. 2012. The composite group as the units of speech production in Nepali. Talk presented at the 33rd Annual Conf. of the Ling. Society of Nepal.

Levelt, W. 1989. *Speaking: from intention to articulation.*Cambridge, MA:MIT Press.

Nespor, M. and Vogel, I. (1986/2007). *Prosodic phonology.* Dordrecht: Foris.

Shilling, H. 2010. Compound and phrasal stress acquisition: When a greenhouse becomes different to a green house. University of Birmingham: MA dissertation.

Tzakosta, M. 2004. The acquisition of the clitic group in Greek. *Proc. of the 24th Annual Meeting of Greek Linguistics*, 693-704. Thessaloniki, Greece: Faculty of Philosophy, Aristotle University of Thessaloniki.

Tzakosta, M. and Manola, D. 2012. Perception and production of compounds by preschool children: pedagogical consequences. In Malafantis et al. (eds.),*Proc.of the 7th Intern. Conf. of the Greek Pedagogical Soc. – Greek Pedagogy and Educ. Research*, vol. 2, 1119-30. Athens: Diadrasi.

Vigário, M. 2011. Prosodic structure between the prosodic word and the phonological phrase: recursive nodes or an independent domain? *The Linguistic Review 27,* 4, 485-530.

Vogel, I. 2009. The Status of the Clitic Group. In Grijzenhout, J. and Kabak, B. (eds.),*Phonological Domains: Universals and Deviations*, 15-46. Berlin: Mouton de Gruyter.

Vogel, I. and Raimy, E. 2002. The Acquisition of Compound vs. Phrasal Stress in English. *Journal of Child Language 29*, 2, 225-50.

Vogel, I. and Spinu, L. 2009. The domain of palatalization in Romanian. In Masullo, P., O'Rourke, E., and Huang, C. (eds.), *Selected Papers from LSRL 37,* 307-20. Philadelphia: John Benjamins.

Vogel, I. and Wheeldon, L. 2010. Units of speech production in Italian. In Colina, S., Olarrea, A., and Carvalho, A. (eds.), *Romance Linguistics 2009*,95-110. Philadelphia: John Benjamins.

Wheeldon, L.and Lahiri, A. 1997. Prosodic Units in Speech Production. *Journal of Memory and Language 37*, 356-81.

Wheeldon, L. and Lahiri, A. 2002. The minimal unit of phonological encoding: prosodic or lexical word. *Cognition 85*, B31-B4

Index of names

Agafonova, M. 175
Alexeeva, S. 25
Athanasopoulou, A. 29, 187
Barabanov, A. 33
Benali, I 37
Botinis, A. 41
Caffò, A. 151
Campana, M. 45
Chaida, A. 41, 51
Chen, T.-Y. 119
Chernova, D. 55
Chukaeva, T. 59
Evdokimova, V. 59
Evgrafova, K. 59
Fedchenko, V. 63
Ferragne, E. 139
Frolova, O. 103
Karpava, S. 71
Karpov, A. 183
Kasevich, V. 75, 79
Kharlamova, A.V. 67
Khokhlova, M. 79
Kisilier, M. 83
Kocharov, D. 33
Kochetkova, U. 87
Kontostavlaki, A. 51
Krzonowski, J. 139
Lai, W.W.S. 91
Lapertosa, L. 151
Lastochkina, A. 79
Levi, J. 95
Litvinova, O. 107
Litvinova, T. 107
Lorenz, D. 99
Lyakso, E. 103
Manerov, R. 111
Manerova, K. 111
Martin, Ph. 1
Menshikova, I. 75, 79
Mirzagitova, A. 115
Mitrofanova, O. 115
Moiseev, M. 33
Myers, J. 119
Nagy, K. 123
Ng, M.L. 91
Niebuhr, O 11
Nikolaenkova, O. 41
Nirgianaki, E. 41
Opitz, A. 127
Panicheva, P. 131
Panova, E. 135
Pechmann, Th. 127
Pélissier, M. 139
Pérez, C.P. 143
Refice, M. 151
Ryzhkova, E. 107
Sabirova, D. 179
Salishev, S. 33
Sandryhaila-Groth, D. 147
Sauer, N.J. 155
Savino, M. 151
Shamina, E. 159
Shapranov, V. 183
Shipilo, A. 163
Shuvalova, E. 79
Skrelin, P. 33, 59, 167
Soares, E.C. 171
Sotiriou, A. 51
Takhtarova, S. 179
Tananaiko, S. 175
Tizón-Couto, D. 99
Uchitel, I. 63
Verkhodanova, V. 183
Vogel, I. 29, 187
Volskaya, N. 167

ExLing 2016
Ebook ISSN: 2529-1092
Ebook ISBN: 978-960-466-161-9