ExLing 2018

Proceedings of 9th Tutorial and Research Workshop on Experimental Linguistics

28-30 August 2018
Paris, France

Edited by Antonis Botinis

université PARIS DIDEROT PARIS 7

ExLing 2018

Proceedings of 9th Tutorial and Research Workshop on Experimental Linguistics

28-30 August 2018, Paris, France

Edited by Antonis Botinis

université PARIS DIDEROT PARIS 7

ExLing 2018
Proceedings of 9th Tutorial and Research Workshop on Experimental Linguistics

Published by ExLing Society
Electronic edition
ExLing 2018
Athens, Greece
ISSN: 2529-1092
ISBN: 978-960-466-198-5
DOI: 10.36505/ExLing-2018

Foreword

This volume includes the proceedings of ExLing 2018, the 9th Tutorial and Research Workshop on Experimental Linguistics, in Paris, France, 28-30 August 2018. The first ExLing Workshop was organised in Athens, Greece, in 2006, and has been regularly held after that in different places. This is the second time we met in Paris, the first one was in 2011.

Following the spirit of the ExLing conferences, new and established researchers came to Paris to discuss developments in linguistic research and experimental methodologies. This time the focus was on prosodic models, in order to gain theoretical perspectives and interdisciplinary knowledge.

We are happy to see that our initial attempt is more and more becoming an established forum for new generations of linguists. As in the previous conferences, our colleagues have come from different parts of the world, and we hope they have had a rewarding exchange of scientific achievements and expertise. This is indeed the core of the ExLing Workshops, which promotes new ideas and methodologies in an international context.

We would like to thank all ExLing 2018 participants as well as the Laboratoire de Linguistique Formelle at the University Paris Diderot. We also thank our keynote speakers Emanuela Cresti, Mark Liberman, Philippe Martin, Jörg Peters, and colleagues from the International Advisory Committee as well as Workshop assistants for their contribution to the successful outcome of the Workshop.

Antonis Botinis

Contents

Tutorial papers

Research papers

The annotation of Information Structure in spoken Japanese

Emanuela Cresti, Massimo Moneglia
LABLITA, University of Florence, Italy
https://doi.org/10.36505/ExLing-2018/09/0001/000334

Abstract

This paper presents the main results of a pilot aimed at verifying the consistency of the Language into Act Theory model for the annotation of Information Structure in spoken Japanese. The segmentation of the Japanese speech flow into utterances through the detection of terminal prosodic breaks and the segmentation into information unit through non-terminal breaks works fine in Japanese. The main Information unit types characterizing the L-AcT approach (Comment, Topic, Parenthesis, Appendix) and their main properties, fit well with the Japanese data-set.
Key words: Spoken Japanese, Information Structure, Prosodic Structure.

Introduction

This paper presents a pilot aimed at verifying the consistency of the Language into Act Theory model (L-AcT, Cresti 2000; Moneglia & Raso 2014; Cresti & Moneglia 2018a) for the annotation of information structure in Japanese. The pilot is intended at grounding the development of an annotated mini-corpus to be stored in the IPIC Database (Panunzi & Gregori 2012) which is devoted to the Cross-linguistic Comparison of Information Structure. At present, IPIC stores resources of Italian, Brazilian and Spanish (Panunzi & Malvessi-Mittmman 2014; Nicolas-Martinez & Lombán fothcaming)

The Japanese data set relies on the Nagoya University Conversation Corpus - NUCC (Fujimura et al. 2012) and corresponds to approx. 80 hours of conversation for 1.5 million transcribed morphemes. Transcripts are in Japanese characters, recently automatically transliterated into Latin characters. NUCC contains 129 natural dialogues and conversations between friends, family members and colleagues, representing a large variety of contexts. For this reason, it can be the source of a selection of samples fitting with the IPIC corpus design model (Cresti & Fujimura 2018). The pilot considers around 100 excerpts derived from four recordings.[1]

L-AcT model foresee the alignment of each utterance in the corpus to its acoustic counterpart through the speech software WinPitch and the annotation of information structure according to a specific methodology and tagset (Moneglia & Raso 2014). In 2. we will briefly sketch the main assumptions for what regards the prosodic cues that are necessary to this end and we will verify the consistency of this model to the Japanese data set. More specifically, in 3 we will consider the criteria for the segmentation of the speech flow into

ExLing 2018: Proceedings of 9th Tutorial and Research Workshop on Experimental Linguistics, 28-30 August, Paris, Frannce

utterances and in 4. the segmentation of the utterance into information unit types.

The general feature of the L-AcT model

L-AcT assumes that the speech flow can be segmented into reference units by means of both pragmatic and prosodic cues. In this framework, a reference unit may belong to two types, respectively *utterance* and *stanza*. The utterance is defined as the counterpart of a Speech act (Austin 1962) and is the primary reference unit for speech (Biber et al. 1999). A stanza expresses the flow of thought (Chafe 1994) and corresponds to a sequence of week speech activities packaged together.[2] The boundaries of both reference units are marked by *prosodic breaks* ('t Hart et al. 1990; Swerts 1999) that are perceived with the quality of being *terminal* (Moneglia & Cresti 2005)

Every reference unit is composed of an Information Pattern which can be simple or complex. Each unit of the pattern necessarily corresponds to a prosodic unit. The prosodic units of a complex pattern are separated the one from the other by *non-terminal breaks*.

Given that information units match in one-to-one way to prosodic units, the prosodic annotation grounds the identification of information units in the flow of speech. Therefore, in order the model to be applied in a language, two preliminary operations are compulsory: 1) identification of terminal breaks; 2) identification of non-terminal breaks.

According to L-AcT, the core of the Information pattern is one specific information unit (Comment) devoted to the expression of illocutionary force. For this reason, a Comment unit is necessary and sufficient for a complete Information pattern. The latter may be simple, which is to say composed of only a Comment or complex. In Complex utterances other optional information unit types may support the Comment, each one corresponding to a dedicated prosodic unit and to a specific information function. Information functions are classified into two basic types, depending on whether they work in fulfilling the semantic content of the utterance or in its communicative support (Discourse markers).

Information unit types with their tags and formal definitions are detailed in Moneglia & Raso 2014. The aim of the pilot is to verify the Adequacy of L-AcT model for the segmentation of spoken Japanese according to key operational principles. We will verify breaks detection, the consistency of the Comment principle, and the consistency of the main textual Information functions; i.e Topic (TOP), Parenthesis (PAR); Appendix (APC).

Terminal breaks, non-terminal breaks and the pragmatic independence of the reference unit

Although major prosodic breaks are prominent also to non-natives, they cannot really judge their terminal or non-terminal nature. The following two examples

correspond to opposite judgements given by non-natives, both not fitting with speech act performance. The major break in figure 1, which is connected to a rising contour, is perceived as a continuation, while the major boundary in figure 2, showing a falling contour, is perceived terminal.

(1) J1:
十三？　うち 十三…
jusan ?　uchi　jusan::
thirteenth ? we　thirteenth::
'tirteenth?' 'we (are) thirteenth…'

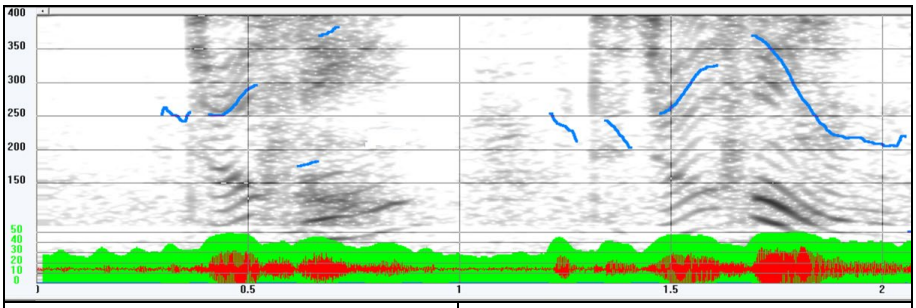


Figure 1. Terminal break with rising contour.

(2) *M3A18:
もう あんた 今ごろ 全部, 葉っぱ-が 出そろって な-あかん　よ。
mou anta　imagoro　　zenbu　/ happa-ga　desorotte　na-akan　yo //
already　you now　　every　/ leave-SUB come-out must　PR FIN //
'As a whole for now / leaves had to be already born'

As the transcript shows, competent speakers easily recognise that the first break in (2) is terminal, since it corresponds to a concluded speech activity (*request of confirmation*) that is followed by a second speech activity (*supposition*). If a stretch of speech can be interpreted in isolation as a speech act the prosodic break is judged terminal.

On the other way around, in (3) competent speaker do not assign the value of independent speech act to the first prosodic unit. The break is perceived non-terminal since it cannot be interpreted in isolation, and the prosodic unit is considered part of a sequence interpreted as *self-conclusion.*

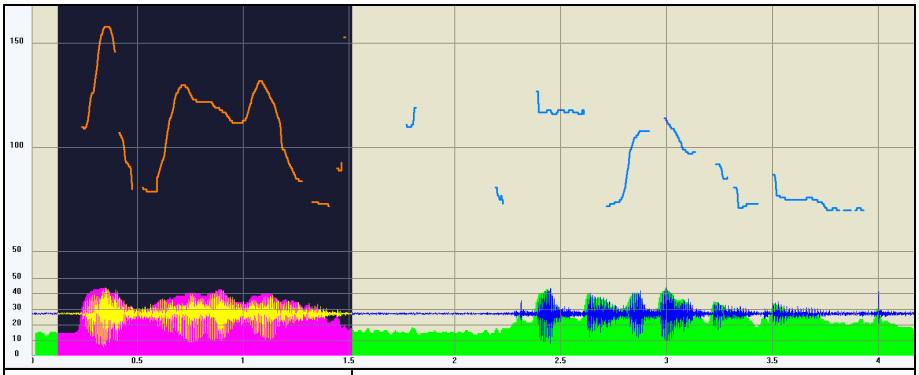


Figure 2. Terminal break with falling contour.

Therefore, the identification of the terminal quality in a major boundary does not follow from intrinsic prosodic properties (rising vs falling boundary tones), but strictly requires the access to the language competence which grounds the pragmatic interpretation. Based on this competence the linguist determines whether the prosodic unit can be interpreted or not in isolation. When it doesn't, the unit is part of a larger utterance and the perceived prosodic break is considered non-terminal. Therefore, the assignment of a value to prosodic breaks and pragmatic judgements go hand in hand.

The Comment principle and the structure of information within the reference unit

L-AcT foresees that when the utterance is segmented into information units these are marked by prosodic boundaries. (2) as well as the following examples, allow to verify: a) the segmentation of the utterance into information units according to non-terminal breaks detection; b) the correspondence of the information units to the typology of information function foreseen in L-AcT.

As we will see, the first prosodic unit in (2) corresponds to a Topic unit of a complex utterance, however, what is more interesting in (2) to our ends is the nature of the second unit. L-AcT assumes that within an utterance, characterized by an illocutionary value, one and only one unit identifies the information unit bearing the illocutionary information. We call Comment this unit.

This core assumption of the theory is confirmed in (2). Indeed, listening in isolation to its second unit, competent speakers find that it can receive a pragmatic interpretation. The Comment principle hold in all terminated sequences of the pilot, grounding the application of the L-AcT model. For instance, let's consider the following dialogue between wife and husband, where she complains about a delay in planting tulips and the husband note that indeed nothing flourished.

(3) *F1A8:
あ、と、 チューリップ とか て 今、もう 植え -たら 安い ねん けど ね 、球根 。
a to/PHA chu^rippu toka-te/TOP ima, mou/PARue –tara yasui-nen kedo ne /COM kyuukon //APC
ah well / tulip such-as / right now / plant-if cheap but PR / bulb //
ah well / the tulips / if you, (had) already, planted (them) it would be less costly / the bulbs '
%ill: expression of disagreement

(4) *M3A:
チューリップ なんか, 1つ-も出て へん やん うち .
chu^rippu nanka /TOP hitotsu-mo de-te hen yan /COM uchi //APC
tulip such-as / anything go-out not isn't /our place //
(for what regards) tulips / nothing flourished / in our place //
%ill: ascertainment

As the F0 tracks in Figure 3 and 4 shows, both utterances are segmented into prosodic units by non-terminal breaks and present complex information patterns. Breaks are clear to perception and are marked by F0 resets.

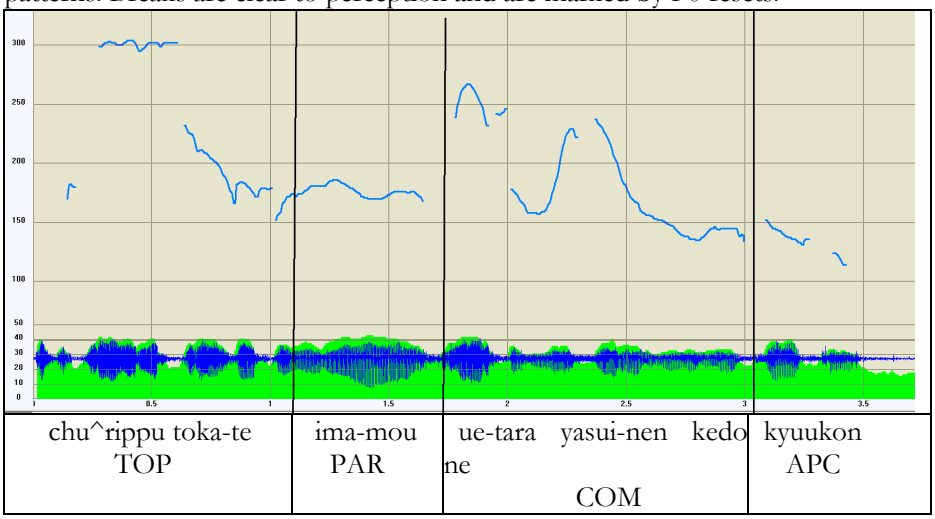


Figure 3. Prosodic units and F0 resets at non-terminal boundaries.

Working with competent speakers we first verified that one only unit play the role of Comment and can be interpreted in isolation and, in parallel, that all the other units can be erased from the signal without prejudice for the interpretability of the utterance. The Comment principle works fine in Japanese.

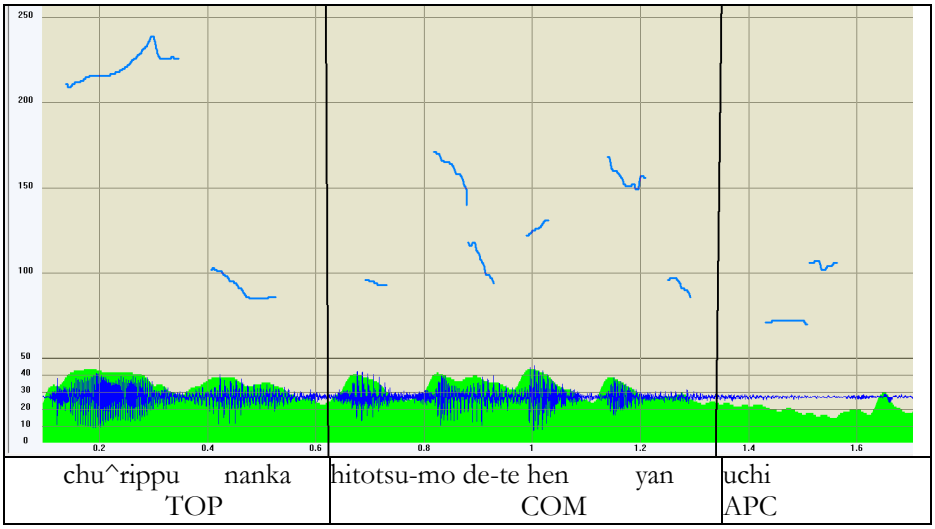


Figure 4. Prosodic units and F0 resets at non-terminal boundaries.

The information function of the Topic is defined in L-AcT at the pragmatic level. The Topic *specifies to the addressee what the illocutionary activity performed by the Comment is about*. From a formal point of view, it must precede the Comment and should bear a strong prosodic prominence (*prefix* prosodic form according to Signorini 2004; Mittman 2012; Cavalcante 2016; Cresti & Moneglia 2018b).

Appendix and Parenthesis constitutes supplementary strategies foreseen in L-AcT for packaging information. The Appendix is a *textual integration of the Comment's content*, it has low semantic relevance and behaves as an adjunct at the end of the utterance. The Appendix occurs necessarily after the Comment unit and is performed via a prosodic unit of the *suffix* type, (with a low-falling profile and weak intensity). In this frame, it is sharply distinct from the Topic, since it does not specify the domain of relevance of the Comment. Topic/Comment and Comment/Appendix are well-formed prosodic patterns.

Parenthesis *insert a metalinguistic evaluation in the utterance* and it is characterised by a jump in F0 average with respect to the other units. Contrary to Topic and Appendix, Parenthesis does not have a canonical order with respect to the Comment; it can appear in whatever position, but not in first position, therefore the sequence Parenthesis/Comment is not a well-formed prosodic pattern.

The above previsions of the L-AcT model match directly with Japanese data for what regards the informational role of Topic and Appendix, which respectively precedes and follow the Comment. The Topic unit is coherent with the informational definition of Topic: the self-conclusion in (2) is relative to the period of the year; the disagreement in (3) and the ascertainment in (4) regards "tulips". The Comment does not regard neither the "place" nor the "bulbs", which coherently with the function of Appendix integrates the information already given.

Japanese is a Topic-language (Li & Thompson 1976; Shibatani 1982) and our spontaneous speech pilot confirm this. Also, from the prosodic point of view, Japanese fits with the general feature of the model. The prefix-root prosodic pattern supporting the Topic-Comment is a clear prosodic structure also in Japanese. The Topic bears a strong prosodic prominence, while the Appendix is weak and bears no movement.

Parenthesis strictly follow the properties foreseen in L-AcT. Competent speaker verified that Parenthesis in (3) can be erased without prejudice for the well formedness of the Topic Comment Appendix prosodic pattern. On the contrary, they also verified that, if the Topic unit is deleted, the resulting pattern (Parenthesis / Comment / Appendix) does not makes sense, this, despite the fact that "ima mou" [right now] might in principle be a reference for the act of disagreement.

Conclusions

Information units necessarily correspond to prosodic units as foreseen in the tradition of studies on Information Structure. Considering specifically terminal and non-terminal prosodic breaks highlight information units in spoken Japanese. These cues, that are very prominent to perception and can be verified with the support of F0 analysis are therefore viable for a large corpus annotation campaign. Beyond of that, L-AcT works fine in all its basic principles and specifically for the illocutionary definition of Comment; i.e the information unit specifying the illocutionary activity that permits the pragmatic interpretation of the utterance. More specifically the basic information unit types which pattern the utterance (Topic, Parenthesis and Appendix) structure the information in the utterances also in Japanese.

Notes

1. The acoustic source of NUCC transcripts is not delivered. The copy-right owner granted the wav files specifically for this pilot.
2. We will limit our argument to the utterance only in this paper.

References

Austin, J. 1962. How to do things with words. London: Arnold.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finnegan, E. 1999. The Longman Grammar of Spoken and Written English. London: Longman.

Cavalcante, F.A. 2015. The topic unit in spontaneous American English: a corpus-based study. Belo Horizonte: UFMG.

Cavalcante, F.A., Ramos, A.C. 2016. The American English spontaneous speech minicorpus. Architecture and comparability. CHIMERA: Romance Corpora and Linguistic Studies 3(2), 99-124.

Chafe, W. 1994. Discourse, consciousness, and time. University of Chicago Press.

Cresti, E. 2000. Corpus di italiano parlato. Firenze: Accademia della Crusca

Cresti, E. 2010. La Stanza: un'unita di costruzione testuale del parlato. In Ferrari, A. (Ed.), Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione e giustapposizione. Atti del X Congresso SILFI, 713-732. Firenze: Franco Cesati.

Cresti, E., Fujimura, I. 2018. The information structure of spontaneous spoken Japanese and Italian in comparison: a pilot study. In Manco, A. (Ed.) Le lingue extra-europee e l'italiano: aspetti didattico-acquisizionali e sociolinguistici. Milano: Officinaventuno.

Cresti, E., Moneglia, M. 2018a. The illocutionary basis of information structure. In Adamou, E., Haude, K., Vanhove, M. (Eds) Information Structure in Lesser-described Languages. 359-402. Amsterdam: Benjamins.

Cresti, E., Moneglia, M. 2018b. The definition of the Topic within Language into Act Theory and its identification in spontaneous speech corpora. Revue Romane 5, 30-62.

Fujimura, I., Chiba, S., Ohso, M. 2012. Lexical and grammatical features of spoken and written Japanese in contrast: Exploring a lexical profiling approach to comparing spoken and written corpora. In Raso, T., Mello, H., Pettorino, M. (Eds), Proc. of the International GSCP 2012 Conference: Speech and Corpora. 393-398. Firenze: FUP.

't Hart, J., Collier, R., Cohen, A. 1990. A Perceptual Study on Intonation. An Experimental Approach to Speech Melody. Cambridge University Press.

Panunzi, A. Gregori, L. 2012. DB-IPIC. An XML Database for the Representation of Information Structure in Spoken Language. In Mello, H., Panunzi, A., Raso, T. (Eds), Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation. 133-150. Firenze University Press.

Panunzi, A., Malvessi-Mittmann, M. 2014. The IPIC resource and a cross-linguistic analysis of information structure in Italian and Brazilian Portuguese. In Raso, T., Mello, H. (Eds) 129–151. Spoken Corpora and Linguistic Studies, Amsterdam: Bejamins.

Nicolas-Martinez, C., Lombán, M. (forthcoming). Mini-Corpus del Español para DB-IPIC, CHIMERA.

Signorini, S. 2005. Topic e soggetto in corpora di italiano parlato. Florence: University of Florence. (PhD Thesis).

WinPitch: https://www.winpitch.com/

Brain waves and prosodic structure

Philippe Martin
LLF, UFRL, Université Paris Diderot, France
https://doi.org/10.36505/ExLing-2018/09/0002/000335

Abstract

There is a general agreement among prosodists about the definition of the sentence prosodic structure (PS) as a hierarchical organization of minimal prosodic units called accent phrases (AP). However, disagreements appear among models about the characteristics of the accent phrases, the role of their pitch accents, and the function of the prosodic structure itself in the linguistic system. This paper describes some properties of the Incremental Dependency (ID) model and their relations with brain waves, compared to the dominant Autosegmental-Metrical (AM) model.

Instead of being described as containing only one content word (a noun, an adjective, a verb or an adverb), accent phrases in the Incremental Dependency model are defined by the time taken to pronounce them orally or process them in silent reading. In non-lexically stressed languages such as French or Korean, AP duration varies from 250 ms to about 1250 ms, values corresponding to the range of delta brain oscillations. This suggests that delta waves are involved in the coding and decoding of accent phrases. A eurhythmic process aiming to balance the duration of successive AP determines the actual its duration in the limits of delta wave variations.

Whereas the AM approach does not consider interactions between pitch accents, the ID model describes melodic movements of pitch accents as markers of dependency relations existing between accent phrases, relations which determine the sentence prosodic structure.

Furthermore, while the AM model views the prosodic structure as an emanation of the sentence syntactic structure, the ID model considers the prosodic structure as independent from grammar and actually generated in the speech process chunk by chunk before syntax and lexical selection. The prosodic structure appears then as a syntactic preselection device to process quickly in real time the flow of information, given the limits of short-term memory for speech (about 3 seconds).

Key words: prosodic structure, delta brain waves, autosegmental-metrical, Incremental-Dependency, theta brain waves.

Introduction

The dominant Autosegmental-Metrical model in prosodic phonology envisions the prosodic structure as a hierarchy of accent phrases (AP). Accent phrases are prosodic minimal units containing one single content word and their associate grammatical words. As content words, verbs, adjectives, adverbs and nouns, are normally stressed, accent phrases contain one single stressed syllable (excluding eventual emphatic stress). This hierarchy has been for a long time limited to one level (cf. the Strict Layer Hypothesis, Selkirk, 1978), but more recent studies admit the existence of two levels in the prosodic structure (Michelas,

2011). These levels are respectively called intermediate (intonation) phrase (ip) (lower level) and Intonation Phrase (IP) (highest level). Therefore, in a two-levels prosodic structure, ip groups accent phrases, IP groups ip's and the PS groups IP's.

In this view, pitch accents, located on stressed syllables and more precisely on stressed vowels of accent phrases, do not play any role in the indication of the prosodic structure. This role is devoted to boundary tones, which are located on the last syllable of intermediate phrases and Intonation Phrases.

The relation between the prosodic structure, which defines how accent phrases are grouped together in relation (or not) with syntactic units, can be viewed in the AM approach as an emanation of syntax, i.e. a hierarchy inferred from syntax with some adaptation rules. The PS is therefore more or less congruent with syntax as perfect congruence cannot be usually realized given among other reasons the limitation of the prosodic structure to two levels (except for short sentences), whereas the syntactic structure may need more levels.

In practice, descriptions proceed form a transcription of acoustic (or perceived) data using ToBI notations, in order to establish a set of well-formed sequences of tones levels, infer a prosodic grammar from the set of these sequences and finally determine adaptation rules that would predict well-formed tone sequences from a given sentence and its syntactic and possibly semantic structures.

In the AM view, the prosodic structure just as the syntactic structure is static, meaning that no temporal effect is involved in the definition of its description. Indeed, the general use of ToBI notation eludes the transcription of any duration factor of accent phrases, ip or IP, except for the perceived pauses noted (rarely) on a perception scale from 1 to 5.

The Incremental Dependency model (IP), developed since 1975 (Martin, 1975), takes almost all possible opposite views on the prosodic characteristics found in the AM approach. The only point of agreement pertains to the definition of the prosodic structure itself as a hierarchy of accent phrases (called earlier prosodic words), whereas the main differences are as follows:

Accent phrases are not defined by the category of words they contain but by the time it takes to pronounce them, either orally or silently.

The prosodic structure is not generated from syntax through some deep structure, but the other way around where the PS precedes the encoding and decoding of the syntactic structure.

Pitch accents located on accent phrases stressed syllables are not without interaction with each other, on the contrary they are markers of dependency relations existing between accent phrases.

The dependency relations indicated by pitch accents are generated and decoded dynamically along the time scale and not generated from some deep structure.

The prosodic structure is recursive and not constrained by the Strict Layer Hypothesis.

All of these characteristics find an explanation in the brain wave features involved in speech production and perception. In particular, it has been shown the theta and delta brain waves timing properties do correspond closely respectively to the syllables (Ghitza, 2011) and to the accent phrases ranges of duration (Martin, 2018). The generation of the sentence prosodic structure itself, which cannot be avoided in either spontaneous or read speech whether produced orally or silently, finds its origin if the limited capability of our short-term memory dealing with speech.

Accent phrases

In lexically-stressed languages such as English or Italian, accent phrases usually include only one content word and therefore only one stressed syllable, as nouns, verbs, adverbs and adjectives are normally stressed at some morphological boundary, their eventual associated grammatical words being unstressed (excluding emphatic stress). In non-lexically stressed languages such as French or Korean, the position of stressed syllables is governed by rhythmic constrains whose effects are hidden in lexically-stressed languages by morphological stress rules. These rhythmic constrains are easy to observe in the pronunciation of long words (ex. *paraske**vi**dékatriapho**bie*** in French) which requires at least one extra stressed syllable on top of the final default position of the rhythmic stress. On the other hand, sequences involving one syllable word (ex. *par le **fait** **que***) require a gap of at least 250 ms to keep the final syllable of the preceding AP perceived as stressed.

These observations lead to confer to accent phrases a minimal duration of 250 ms (including the gap preceding a single syllable accent phrase), and a maximal duration of about 1250 ms. If thanks to speech synthesis manipulation, two consecutive stressed syllables are separated by more than 1250 ms, listener will tend to perceive an extra intermediate stressed syllable, even if not showing any particular acoustic characteristic of stress, such as a longer duration of a perceived pitch movement.

This means that the actual content of a given accent phrase will depend on the speech rate, the stressed syllable being always placed in final position in non-lexically and non-tonal stressed languages. While an average speaking rate corresponds to about 4 to 5 syllables per second, resulting to accent phrases containing a maximum of 6 to 7 syllables, very slow speech rate will result in only one syllable in each AP (a detached pronunciation syllable by syllable), and a very fast speech tempo will pack up to 10 or 11 syllables in a single AP (cf. *parole de jeunes* in French).

Between the 250 ms – 1250 ms extreme values, a eurhythmy principle applies in order to reduce the variation of duration between consecutive accent phrases (Wioland, 1985). As the number of syllables of every successive word

cannot be changed, speakers have to choose to either adapt the speech rate for every accent phrase to achieve a relative eurhythmicity, speaking faster for accent phrase with many syllables and slower for those with few syllables, or selecting consecutive words to obtain a similar number of syllables in consecutive accent phrases, possibly at the expense of congruence with syntax. The first strategy is preferred for spontaneous speech, the second for read speech.

Prosody before syntax

The precedence of prosodic structure generation over syntax can be shown in many ways (Keating and Shattuck-Hufnage, 2002 , Martin 2018), but one of the most convincing argument comes from the reading process, where this property may seem at first the least probable.

Reading operates by saccades, where the eye focalizes on successive words by jumps which do not exceed some 10 to 20 characters (Reyner et al., 2010). This means that a reader who never read a given text before can only adapt its (partial) prosodic structure to a few consecutive words, eventually helped by some punctuation marks. In configurations such as (A) (B C) where the number of syllables of the words B and C would exceed some 10 to 20 characters, the reader could not normally anticipate the syntactic relation existing between B and C, and would instead group A and B. Discovering C at the next reading saccade, the only solution would then to put prosodically C at the same level than B, ending the sequence AB. The prosodic hierarchy will then be [A B] [C], not congruent with the syntactic (A) (B C). In the French example *deux alpinistes allemands ont trouvé le cadavre d'un homme dans un glacier* "two German mountaineers found the corpse of a man in a glacier", readers will group prosodically *ont trouvé* with *le cadavre* and then add *d'un homme* at the next step of the prosodic structure elaboration.

Accent Phrases dependency relations

When we start a sentence, either orally or silently "in our head", we have already decided about its modality, declarative or interrogative, and their variants, imperative or implicative for the declarative, and surprise or doubt for the interrogative. The terminal conclusive melodic contour on the last pitch accent correlated with this modality is therefore planned in the future of the process, and will mark also the end of the sentence (except for deferred complements or theme-rheme constructions not discussed here). Therefore, the melodic realization as non-final of the other pitch accents that precede depend on the final prosodic event planned by the speaker in the future instantiating a dependency relation with another accent phrase in the future. This can be generalized to all non-terminal pitch accents, which may define a dependency relation with some other pitch accent located later in the prosodic structure.

In French, following the terminology of Delattre (1966) but with a somewhat different interpretation, at least five phonological categories of melodic contours located on pitch accents can be considered: the *major continuation*, which indicates a dependency relation towards the *terminal conclusive contour declarative* or *interrogative*; the *minor continuation*, which indicates a dependency relation towards the major continuation; the *neutralized* contour, which indicates a dependency towards either the minor continuation, the major continuation or the terminal conclusive contour.

The symbols attached to these categories are respectively C0 ↓terminal conclusive declarative, Ci ↑terminal conclusive interrogative, C1 ↗ major continuation, C2 ↘ minor continuation, Cn → neutralized contour. The falling and rising arrows for C2 and C1 implement the contrast of melodic slope used in French to indicate the dependency "to the right" between accent phrases, where a falling contour C2 is depending of a rising contour C1 located somewhere later in the sentence, and the rising contour C1 a dependency towards the final conclusive declarative C0. C1 and C2 melodic variation are above the glissando threshold (Rossi, 1971), which means their melodic change is effectively perceived as such by listeners, contrary to the neutralized contour Cn, and possibly C0 which usually reaches the lowest frequency level in the sentence.

Accent phrases local congruence with syntax

In non-lexically non-tonal stressed languages, accent phrases contain one or more content words with their syntactically associated grammatical words. These micro syntactic constructions constitute the elementary building blocks stored as such in speakers and listeners long-term memory. In the speech decoding process, listeners retrieve these stored elements already partially syntactically organized, and assemble them according to the hierarchy defined by the incremental prosodic structure. Therefore, referring to the example given above, accent phrases containing sequences such as **allemands* <u>*ont*</u> "Germans have" will be excluded as unlikely be part of the listener lexicon.

Given this local congruence property, the function of the prosodic structure in the linguistic system appears more clearly: to allow the speaker to quickly assemble hierarchically partial syntactic structures in the very short allowed time window before the speech sound vanishes in the short-term memory.

Prosodic structure phonetics and phonology

The prosodic annotation in the Autosegmental-Metrical framework is usually done with the ToBI notation system, basically transcribing high and low melodic targets (and not contours) with H and L symbols. Numerous variants allow to represent complex melodic shapes, as well as their position relative to

pitch accents (i.e. H* or L*) or syntactic boundaries (i.e. %H or %L "left", H% or L% "right".

Besides the often-questionable validity of actual transcription made from sometimes elusive fundamental frequency curves, this annotation process will basically mix phonological and phonetic data, to be sorted at a later stage. Unfortunately, this is not always easy to do in practice, as the specific function of the prosodic structure being inferred from syntax in the AM approach is not clearly defined.

By contrast, the incremental dependency model poses the prosodic structure as a frame for micro syntactic patterns, which is actually generated before the overall syntactic structure in the speech planning process operated by the speaker, in spontaneous as well as in read speech, oral or silent. From this view of the prosodic structure function and the assumed dependency system of relations between accent phrases, it is relatively easy to discover features relative to the necessary contrast between categories of contours and to sort them from the phonetic details such as those relative to regional or idiosyncratic variations.

Brain waves

Brain waves are generated by constant electrical exchanges of the order of microvolts that take place between neuronal regions through long chains of synapses. Brain waves are usually classified by their frequency range. In the speech domain, brain oscillations of interest are delta, oscillating in the 0.8 Hz – 4 Hz range, theta 4 Hz – 10 Hz and gamma 30 Hz - 80 Hz (these approximate values may vary depending on the subject). Beta brain waves, which are endogenous (i.e. not trigger directly by an external stimulus), oscillates between 14 Hz and 21 Hz are also involved in speech processing.

It has been showed recently that theta waves are directly related to the perception of syllables, delta to the processing of accent phrases (Martin, 2018), and gamma to the perception of higher frequency acoustic consonants characteristics, such as fricatives. Syllables need a minimum of 100 ms duration to be processed, although the perception of their acoustic features themselves can take much less time (Ghitza, 2011). On the other hand, syllables separated by gaps of more than 250 ms will be perceived as stressed.

As mentioned above, accent phrases have a minimal duration of 250 ms (one stressed syllable preceded by the necessary separation from an eventual previous stressed syllable), and a maximal duration of 1250 ms (if longer, an extra stressed syllable will be perceived in the interval).

All these observations can be explained by looking at the temporal dynamics of theta and delta brain oscillations. Their frequency range converted into periods, 4 Hz – 10 Hz as 100 ms – 250 ms for theta, and 0.8 Hz – 4 Hz as 250 ms – 1250 ms for delta, correspond respectively to the range of syllabic duration and to the inter stressed syllable duration, closely linked to the accent phrases duration. As the delta and theta oscillations are phase locked, i.e. they

synchronize each other, the mechanism of syllable and stressed syllable can be interpreted as follows.

In the absence of sound and in particular of speech sound, the delta and theta waves are idle, the oscillate freely inside their duration range while being constantly phased locked. When the first stressed or unstressed syllable of a speech sequence appears, both the theta and delta spikes align on this new temporal event (Gilbert and Boucher, 2007). The upcoming syllables then synchronize the next theta oscillations until a stressed syllable appears, characterized by some perceivable acoustic difference with the preceding syllables, such as a longer duration, a difference in intensity, a marked pitch change. This event triggers a delta spike, which is in phase with the theta spike triggered by the syllable. The eurhythmy effect would prevent variations of successive delta periods to be too large.

Although in lexically-stressed languages the stressed syllable is not necessarily in final position, it has been suggested (Martin, 2018) that the delta spikes initialize a retrieving process of stored accent phrases, containing a micro syntactic structure in the listener long-term memory. This seems also to be the case and trigger a delta spike even in silent speech where stressed syllables are present with no external acoustic stimulus (Magrassi et al., 2015).

This bottom-up mechanism is concurrent with a top-down process where from a first syllabic segmentation the listener can retrieve from long-term memory the content of the incoming segmented accent phrase as well as its stressed syllable position. The switch between bottom-up and top-down processes appears to be instantiated by beta oscillations (Pefkou and al., 2017).

References

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

Ghitza, O. 2011. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm, Frontiers in Psychol. 2, 130.

Gilbert, A., Boucher, V. 2007. What do listeners attend to in hearing prosodic structures? Investigating the human speech-parser using short-term recall, Proc. Interspeech 2007, 430-433.

Keating, P., Shattuck-Hufnage, S. 2002. A Prosodic View of Word Form Encoding for Speech Production, UCLA Working Papers in Phonetics 101, 112-156.

Magrassi, L, Aromataris, G., Cabrini, A., Annovazzi-Lodi, V., Moro, A. 2015.

Sound representation in higher language areas during language generation, PNAS 112 (6) 1868-1873.

Martin, Ph. 1975. Analyse phonologique de la phrase française. Linguistics 146, 35-68.

Martin, Ph. 2018. Intonation, structure prosodique et ondes cérébrales. London: ISTE.

Michelas, A. 2011. Caractérisation phonétique et phonologique du syntagme intermédiaire en français de la production à la perception. Thèse de doctorat, université de Provence.

Pefkou, M., Arnal, L.H., Fontolan. L., Giraud, A.-L. 2017. Theta- and beta-band neural activity reflect independent syllable tracking and comprehension of time-compressed speech. Journal of Neuroscience 16, 37 (33), 7930-7938.

Reyner, K., Slattery, T.J., Béanger, N.N. 2010. Eye movements, the perceptual span, and reading speed. Psychonomic Bulletin 17(6), 834-839.

Rossi, M. 1971. Le seuil de glissando ou seuil de perception des variations tonales pour la parole. Phonetica 23, 1-33.

Selkirk, E.O. 1978. On prosodic structure and its relation with syntactic structure. In Fretheim, T. (ed.) 1978, Nordic Prosody, 111-140. Trondheim: TAPIR.

Wioland, F. 1985. Les structures rythmiques du français. Paris: Slaktine-Champion.

Gender differences in adolescents' written texts

Georgia Andreou, Maria Liakou, Fotini Anastassiou, Vassiliki Tsela
Department of Special Education, University of Thessaly, Greece/ Hellenic Open University, Greece
https://doi.org/10.36505/ExLing-2018/09/0003/000336

Abstract

The aim of this study was to compare adolescents' lexical abilities in relation to their gender, examining the differences between their age too. The study compared lexical abilities in written narrative and non-narrative texts of three hundred typically developed adolescents. The results showed performance differences between female and male participants, with a tendency of an advantage of females over males due to higher persistence and systematic effort levels on the part of the females.
Key words: Gender differences, lexical abilities, written texts, adolescence

Introduction

Gender differences are more evident in adolescence than in childhood. Such disparities have been attributed to the dissimilarities in the rate in which their brains mature. As it has been demonstrated, a girl's brain evolves psycho-motorically and it ultimately develops faster than a boy's. The faster maturation of girls means "lower final slanting, i.e. increased cerebral symmetry and increased verbal ability" (Andreou, 2012). Gender differences in their grading performance assume a pattern from around the age of 11 and remain consistent throughout schooling in secondary education (Berninger, Nielsen, Abbott, Wijsman & Raskind, 2008; Voyer & Voyer, 2014).

In the present study, criteria of lexical complexity were used in written texts of adolescents. Specifically, as far as lexical analysis is concerned we focused on the following two criteria: 1) Length of words (words with 3 or more syllables) (Berman, 2008; Berman & Nir-Sagiv, 2007), i.e. words from three syllables and above, which is an indicator of advanced student development and an important criterion of vocabulary complexity; 2) The use of abstract nouns (Berman, 2008; Berman and Nir-Sagiv, 2007; Nippold, 2007), which is an advanced vocabulary index.

Finally, in the present study, we adopted the distinction between narrative and non-narrative texts in order to avoid the fragmentation of speech into unrelated categories without clear boundaries. Narrative texts are those that reflect specific experiences, mostly the ones that belong to the past, and those that put an emphasis on human actions and experiences. Non-narrative texts are more demanding because they focus on a broad field of description or on the development of concepts, arguments and information without relying on narration (Georgakopoulou & Goutsos, 1999). Adolescents must demonstrate their ability to write in a variety of texts. The text holds a key position in the

whole curriculum of teaching language lessons since students are constantly being asked to recognize the type of texts and meet their demands.

Purpose of the study

Based on the above, the aim of this study was to compare adolescents' lexical abilities in relation to their age while examining the differences between the two genders. Given the difficulty of producing different types of written texts, it was also predicted that girls' writings will show a higher performance than boys' writings in both age groups but also in both types of texts.

Method

Participants

Our research took place in public, urban high schools in Greece. The sample consisted of two groups of 300 typically developing adolescents (N=300) whose average age was 14.5 (N= 150 early adolescents) and 16.6 (N=150 late adolescents) years old respectively.

Procedure

A text on racism was distributed to the students and they were asked to produce one non-narrative and one narrative text, following the instructions given. In total, this study analyzed 600 written texts, half of them being narrative and half non-narrative ones.

Analysis

The Mann-Whitney U Test was used to detect possible differences in written texts between early and late adolescents. It is the non-parametric alternative to the independent t-test. SPSS 15 was used for our statistical analysis and Monte Carlo simulation methods were used to obtain the p-value. When the p-value is less than the significance level α (a =0.05), the result is said to be statistically significant.

Results

The statistical analysis of Mann-Whitney (see tables 1, 2) revealed that there is a statistically significant difference in narrative texts between the two groups of adolescents (boys vs girls) in the: a) length of words (words with 3 or more syllables) and b) abstract nouns.

Table 1. Mean percentage of length of words (words with 3 or more syllables) and abstract nouns by gender (boys vs. girls- early adolescents) in narrative texts (S: Significant - NS: Non Significant).

Type of Adolescents	Length of words	Abstract nouns
Boys (early adolescents)	37.1	1.6
Girls (early adolescents)	41.7	2.4
p-value	0.216 (NS)	0.030 (S)

Table 2. Mean percentage of length of words (words with 3 or more syllables) and abstract nouns by gender (boys vs. girls- late adolescents) in narrative texts (S: Significant - NS: Non Significant).

Type of Adolescents	Length of words	Abstract nouns
Boys (late adolescents)	43.4	2.1
Girls (late adolescents)	47.6	2.0
p-value	0.063 (NS)	p= 0.572 (NS)

Also, the statistical analysis of Mann-Whitney (see tables 3, 4) showed a statistically significant difference in non-narrative texts between the two genders in the: a) length of words (words with 3 or more syllables) and b) abstract nouns.

Table 3. Mean percentage of length of words (words with 3 or more syllables) and abstract nouns in terms of gender (boys vs. girls- early adolescents) in non-narrative texts (S: Significant - NS: Non Significant).

Type of Adolescents	Length of words	Abstract nouns
Boys (early adolescents)	41.5	3.3
Girls (early adolescents)	49.7	6.4
p-value	0.010 (S)	<0.001 (S)

Table 4. Mean percentage of length of words (words with 3 or more syllables) and abstract nouns in terms of gender (boys vs. girls- late adolescents) in non-narrative texts (S: Significant - NS: Non Significant).

Type of Adolescents	Length of words	Abstract nouns
Boys (late adolescents)	44.8	4.5
Girls (late adolescents)	57.0	5.4
p-value	<0.001 (S)	0.110 (NS)

Discussion – Conclusions

Gender analysis revealed that girls showed better performance than boys, and so the initial hypothesis was confirmed. Indeed, in several cases these differences were statistically significant, while in others the differences were within the limits of statistical significance. The findings above, in the age

comparison, showed that in almost all the lexical criteria girls are superior to the boys in both types of texts.

In conclusion, gender differences are more evident in adolescence than in childhood. Gender differences in the production of written texts have been observed in this study. In most research, this particular finding is common, that is, girls have a higher degree of linguistic proficiency than boys. The written texts of girls systematically outweighed the boys and the difference was greater in the theoretical-language lessons (as here is the production of written text) in relation to applied sciences and mathematics (Berninger et al., 2008; Voyer & Voyer, 2014). In addition, this is an indication that girls may have encountered less difficulty in producing texts or that they may have tried more than the boys or that they may have been more focused on the language part of the research. Therefore, the relationship between gender and lexical skills has shown differences between the two genders, given that that girls tended to outperform boys because of their more persistent and systematic effort.

References

Andreou, G. 2012. Language: Theoretical and Methodological Approach (3rd edition), Athens: Pedio Publications.

Berman, R.A. 2008. The psycholinguistics of developing text construction. Journal of Child Language 35, 735-771.

Berman, R.A., Nir-Sagiv, B. 2007. Comparing narrative and expository text construction across adolescence: A developmental paradox. Discourse Processes 43 (2), 79-120.

Berninger, V.W., Nielsen, K.H., Abbott, R.D., Wijsman, E., Raskind, W. 2008. Gender differences in severity of writing and reading disabilities. Journal of School Psychology 46, 151–172.

Georgakopoulou, A., Goutsos, D. 1999. Text and Communication. Athens: Ellinika Grammata.

Nippold, M. 2007. Later Language Development: School-Age Children, Adolescents, and Young Adults. Austin, TX, Pro-Ed.

Nippold, M.A., Hegel, S.L., Sohlberg, M.M., Schwarz, I.E. 1999. Defining Abstract Entities: Development in Pre-Adolescents, Adolescents, and Young Adults. Journal of Speech, Language, and Hearing Research 42, 473-481.

Voyer, D., Voyer, S.D. 2014. Gender Differences in Scholastic Achievement: A Meta-Analysis. Psychological Bulletin 140 (4), 1174-1204.

Analysis of prosodic correlates of emotional speech data

Katarina Bartkova[1], Denis Jouvet[2]
[1]Université de Lorraine, CNRS, ATILF, F-54000 Nancy, France
[2]Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France
https://doi.org/10.36505/ExLing-2018/09/0004/000337

Abstract

The study of expressive speech styles remains an important topic as to their parameters detection or prediction in speech processing. In this paper, we analyze prosodic correlates for six emotion styles (anger, disgust, joy, fear, surprise and sadness), using data uttered by two speakers. The analysis is focused on the way pronunciations and prosodic parameters are modified in emotional speech, compared to neutral style. The analysis concerns speech pronunciation modifications, presence of pauses in sentences, and local prosodic behavior, with an emphasis set on the analysis of the prosody over prosodic groups and breathing groups.

Key words: expressive speech, emotions, prosodic groups, prosodic correlates.

Introduction

Prosody not only conveys information on the linguistic content of the vocal messages, but also on speaker's attitude and emotional state. In speech sciences, expressive speech is now attracting a lot of interest, as for example for speech synthesis (Schröder 2009) and for automatic recognition of emotions (Lanjewar 2013). For such studies, the collection of emotional speech data plays an important role. Several approaches have been envisaged for collecting emotional data; this includes the recording of natural emotions, the recording through induced situations, and the recording of acted speech by professional actors (Scherer 2003). This last mode is currently the most frequently used. Expressive speech synthesis has been developed for corpus-based (Iida 2003) and parametric-based approaches (Yamagishi 2004). As the recording of emotional data is difficult, several techniques have been investigated for adapting speech synthesis systems using small amount of emotional data (Inanoglu 2009) or for converting neutral speech into emotional speech (Tao 2006). It should be noted that parametric speech synthesis, needs to know the sequence of units, that is the sequence of sounds and pauses, for being able to produce the synthesized speech signal.

Following a previous study (Bartkova 2016) that has analyzed prosodic features on a global level, this paper focus on analyzing more local distributions associated to prosodic groups and breathing groups.

Corpus and features

The speech corpus used contains French sentences in six emotional styles (anger, disgust, joy, fear, sadness, and surprise), uttered by two professional speakers, one male and one female. Every emotional style contains about 50 sentences of various length. Each speaker has also uttered the same sentences in a neutral, reading style. This makes possible to study how pronunciations and prosodic parameters vary from neutral to emotional style.

For each sentence, the speech material has been aligned with its corresponding text using an automatic speech-text forced alignment procedure, and then manually checked and corrected if necessary. An automatic process has also been applied to localize prosodic boundaries, based on vowel duration, F0 slope, F0 delta, and pause occurrences,

Statistics on prosodic groups

Emotional speech is generally faster than neutral speech. Compared to neutral speech, the speaking is significantly higher (up to 28% faster) for fear and anger, slightly higher for disgust, joy and surprise, but significantly lower (about 20% slower) for sadness.

As the same sentences have been pronounced in neutral and emotional styles, it was possible to compare the presence and position of pauses. About 95% of the pauses observed in emotional data appears at the same place in the neutral pronunciation. However, except for sadness, 13% to 30% of the pauses observed in neutral pronunciation disappear in emotional data, which explains the increase in speaking rate for these emotional styles.

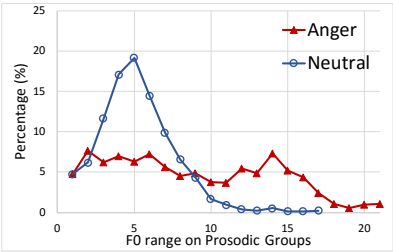


Figure 1: Histograms of F0 ranges over prosodic groups (average over the two speakers) in emotional and neutral speech styles.

Compared to neutral speech, the most noticeable differences in F0 range distributions are observed for anger and sadness (displayed in Figure 1). Larger F0 ranges are much more frequent for anger, and slightly more frequent for fear, surprise and joy (not represented in Figure 1 for lack of space); and smaller F0 ranges are more frequently observed for sadness.

Analysis of delta F0 at end of breathing groups

The histograms in Figure 2 display the distributions of the delta F0 measures at the end of the breathing groups (for lack of space, only two emotions are displayed). To better see the occurrences of falling and rising delta F0 values at the end of the breathing groups, the horizontal axis is symmetric (from -15 up to +15 semi-tones). Anger and fear styles comprise larger amounts of negative delta F0 values than the neutral style. The other noticeable difference is observed for sadness, displaying a sharp histogram of delta F0 values centered around zero; exhibiting rather flat F0 patterns. A tendency of using relatively flat or slightly falling F0 pattern is also observed in disgust.

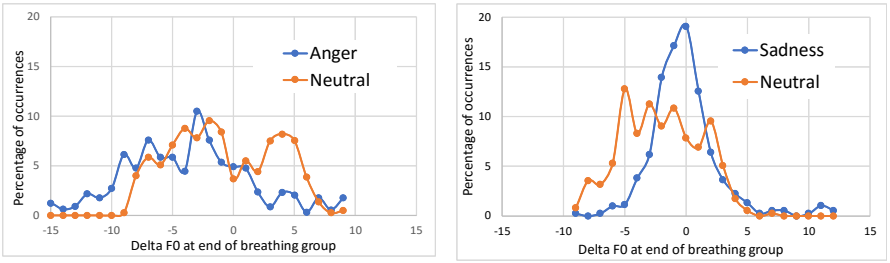


Figure 2: Histograms of delta F0 values at end of breathing groups (average over the two speakers) in emotional and neutral speech styles

Segmental level analysis

Sequences of phone segments associated to each word have been compared between emotion and neutral data to obtain statistics on the modification of the pronunciation of the words from neutral to emotional speech. Unlike other studies (Tahon 2016), few phoneme changes are observed in the data. The most common phoneme changes concern phonetic feature assimilations (mostly nasalization or voice feature assimilation) that are slightly more frequent in emotional speech than in neutral style. Some cases of liquid omissions are also observed, mainly in consonantal clusters. However, the main difference is mostly the omission of the schwa like vowel. In the emotional style a high number of schwa are omitted, and this vowel omission is place sensitive. In fact, the first and the last breathing groups of each record (a record corresponds to one or a few sentences) contain the highest number of schwa omissions (compared to neutral speech), 24% of the omissions are observed in the first breathing group and 27% in the last breathing group. Also, the number of schwa omissions is slightly emotion dependent, the highest percentage of schwa omissions is observed for disgust, fear and joy.

Conclusion

This paper has presented an analysis of prosodic correlates of emotional speech using a corpus of acted emotional speech. In comparison to neutral data, for the anger emotion, the speaking rate is higher, the number of pauses is significantly lower, larger F0 ranges over the prosodic groups are more frequent, as well as large negative delta F0 at end of breathing groups. Joy, disgust and fear emotions exhibit also higher speaking rate than neutral style. However, sadness exhibits a quite different behavior: compared to neutral speech, the speaking rate is lower, with a high number of inserted pauses. Unlike the other five emotional styles, in sadness style, there is a shift of the F0 range histogram towards lower values. As for the delta F0 at end of breathing groups, small values (rather flat F0 pattern) are more frequent.

Acknowledgments

This work has been partially supported by the CPER LCHN (Contrat Plan Etat Région "Langues, Connaissances et Humanités Numériques").

References

Schröder M. 2009. Expressive speech synthesis: Past, present, and possible futures. Affective information processing, 111-126. London: Springer.

Lanjewar R. B., Chaudhari D.S. 2013. Speech Emotion Recognition: A Review. Int. Journal of Innovative Technology and Exploring Engineering (IJITEE), 2.

Scherer K.R. 2003. Vocal communication of emotion: A review of research paradigms. Speech communication 40(1), 227-256.

Iida A., Campbell N., Higuchi F., Yasumura M. 2003. A corpus-based speech synthesis system with emotion. Speech Communication 40(1), 161-187.

Yamagishi J., Masuko T., Kobayashi T. 2004. HMM-based expressive speech synthesis-Towards TTS with arbitrary speaking styles and emotions. Proc. Special Workshop in Maui, Maui, Hawai.

Inanoglu Z., Young S. 2009. Data-driven emotion conversion in spoken English. Speech Communication 51(3), 268-283.

Tao J., Kang Y., Li A. 2006. Prosody conversion from neutral speech to emotional speech. IEEE Trans. on Audio, Speech, and Language Processing 14(4), 1145-1154.

Bartkova K., Jouvet D., Delais-Roussarie E. 2016. Prosodic Parameters and Prosodic Structures of French Emotional Data. Speech Prosody 2016, Boston, United States.

Tahon M., Qader R., Lecorvé G., Lolive D. 2016. Optimal Feature Set and Minimal Training Size for Pronunciation Adaptation in TTS. SLSP'2016, 108-119.

Prosody and temporal productions in Greek

Antonis Botinis[1], Athina Kontostavlaki[1], Evgenia Magoula[2], Olga Nikolaenkova[3], Charalambos Themistocleous[4]

[1]Lab of Phonetics & Computational Linguistics, University of Athens, Greece
[2]Department of Primary Education, University of Athens, Greece
[3]Department of General Linguistics, Saint Petersburg State University, Russia
[4]Department of Swedish, University of Gothenburg, Sweden

https://doi.org/10.36505/ExLing-2018/09/0005/000338

Abstract

In this study, we have investigated the temporal patterns of syllable onsets, nuclei, and codas in Greek. The main findings showed significant effects of stress and focus on syllable constituents, which suggest complex effects of lexical and sentence prosody on intrasyllabic constituents. Nevertheless, there was no significant effect of syllable constituents and focus on the overall syllable duration despite the fact that consonants at coda position differed in their intrinsic duration. This finding suggests that the syllable exercises control over the duration of intersyllable constituents.

Key words: syllable structure, onset, nucleus, coda, focus, stress, Greek

Introduction

Syllable duration depends on a variety of factors such as its internal structure and position, whether it is stressed or focused (Botinis 1989) and whether it carries edge-tones (Themistocleous 2014); Therefore, the duration of segments that make up the syllable is conditioned by its structure and role as a prosodic constituent. This is also evidenced by controlled experiments and studies on free speech (e.g., Greenberg, Carvey, Hitchcock, Chang, 2003). In this study, we look at the temporal patterns of syllable onsets, nuclei, and codas in Greek, in order to uncover prosodic effects on segmental durations.

Experimental methodology

Four male and five female Athenian speakers produced a set of twelve test words (Table 1) in the carrier phrase [ˈeleɣe ___ θetiˈka] (he was saying ___ positively) either in a fairly neutral way or in focus (see Botinis 1989 relevant discussion). We conducted two regression analyses: in the first, the response variable was the duration of syllable constituents and in the second, the response variable was the total syllable duration. The syllable consituent (onset, nucleus, and coda Cŋ, Cn, Cŋ), stress (stressed vs. unstressed), focus (information focus vs. neutral), and their interactions were the predictors in both tests. Statistics were conducted in R (R Core Team, 2016).

Table 1. Test words with antepenultimate and penultimate lexical stress.

Antepenultimate stress	Penultimate stress	Gloss
[ˈsiŋ.fonos], singular	[siŋ.ˈfonus], plural	Concession
[ˈsiŋ.vulos], singular	[siŋ.ˈvulus], plural	Advisor
[ˈsin.θetos], singular	[sin.ˈθetus], plural	Complex
[ˈsin.ðezmos], singular	[sin.ˈðezmus], plural	Link
[ˈsiŋ.xronos], singular	[siŋ.ˈxronus], plural	Synchronic
[ˈsiŋ.ɣrapse], past	[siŋ.ˈɣrapsi], future	wrote/will write

Results

The results are presented in Figures 1-2 and Tables 2-3. Lexical stress and sentence focus have a lengthening effect on overall syllable duration: unstressed non-focused syllables are 126 ms, stressed and unfocused syllables are 160 ms and stressed and focused syllables are 163 ms. Also, there are statistically significant effects of stress and focus on duration. Stress has significant effects on syllable nuclei and coda that differ from the intercept (i.e., syllable onsets) in their durations. Also, codas and nuclei differ from onsets and there are significant interactions of foci and stresses on nuclei and codas.

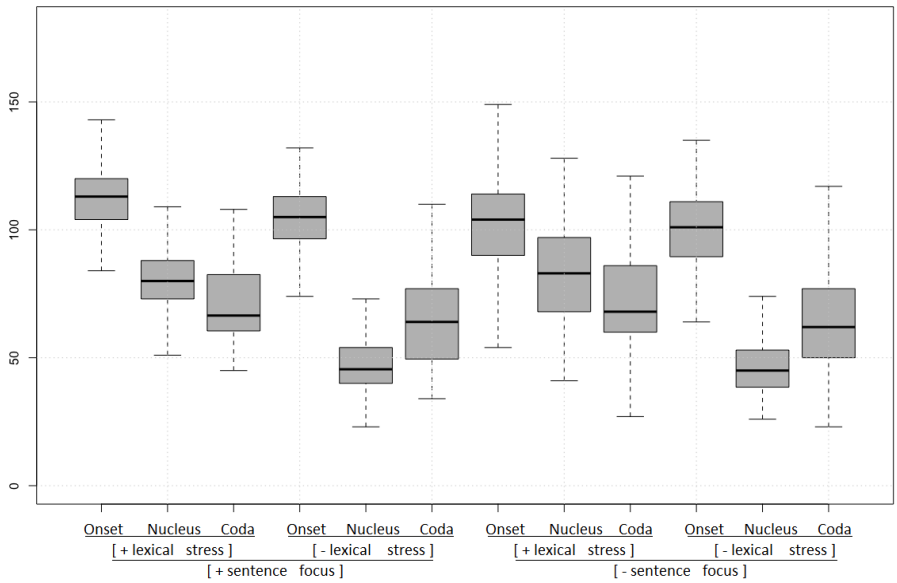


Figure 1. Onset, nucleus and coda syllable constituents as a function of lexical stress and sentence focus.

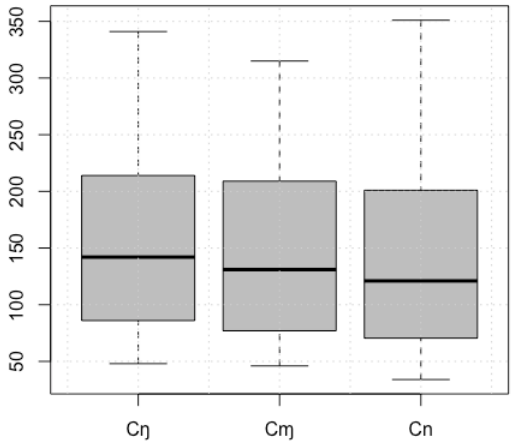


Figure 2 Total syllable duration in ms. for syllables ending with /ɱ, n, ŋ/.

Table 2. Linear regression model for syllable constituent, stress, and focus on duration. The model had a residual standard error: 16.51 on 2277 degrees of freedom; multiple R-squared: 0.5996; Adjusted R-squared: 0.59; F(11)= 310, df=2277, $p < 0.001$.

(Intercept)	108.995	1.153	94.53	0.001
Nucleus	-29.576	1.631	-18.13	0.001
Coda	-36.79	1.631	-22.56	0.001
Unstressed	-5.637	1.696	-3.32	0.001
FocusN	-7.589	1.627	-4.66	0.001
Nucleus:Unstressed	-26.004	2.399	-10.83	0.001
Coda:Unstressed	-1.488	2.399	-0.62	0.535
Nucleus:FocusN	11.015	2.3	4.78	0.001
Coda:FocusN	7.95	2.3	3.45	0.001
Unstressed:FocusN	3.923	2.398	1.63	0.102
Nucleus:Unstressed:FocusN	-9.259	3.392	-2.73	0.006
Coda:Unstressed:FocusN	-6.008	3.392	-1.77	0.076

Nevertheless, a linear regression model with total duration as the dependent variable shows no significant effects of focus and syllable constituent. We observe that stress and syllable constituent (as it is evident by the significnat effect of the coda consonant [n] over the intercept) affect syllable duration ($p < .05$) significantly.

Table 3. Linear regression model for syllable constituent, stress, and focus on total syllable duration. The model had a residual standard error: 14.61 on 751 degrees of freedom; multiple R-squared: 0.1813, Adjusted R-squared: 0.17; F(11)= 15.12, *df*=751, $p < 0.001$.

(Intercept)	77.25	1.92	40.27	0.001
Cn	11.47	2.71	4.23	0.001
Unstressed	5.59	2.69	2.08	0.05

Discussion

The syllable is one of the fundamental units of speech production across different dimensions including representation, processing and implementation. In this study, we have provided acoustic evidence from Greek syllables that realize the first morpheme of multimorphemic words. The results show that onsets are clearly longer than nuclei and codas. One of the main reasons for that is that the intrinsic duration of sibilant consonants in Greek is long (Themistocleous 2017), which to a certain degree explains this effect. Overall, the syllables do not significantly differ in their total duration despite the fact that their coda consonants differ in their intrinsic duration. One account for this phenomenon is that syllables 'regulate' the duration of their constituents so that there is a tendency to lengthen or shorten their segments to accommodate for their intrinsic duration (about the role of syllables vs. segments, see Hickott 2014). So, this study provides evidence that favors a predominance of syllable over subsyllabic constituents and explains why despite different consonantal realizations of syllables retain their overall duration.

References

Botinis, A. 1989. Stress and prosodic structure in Greek. Lund University Press.

Greenberg, S., Carvey, H., Hitchcock, L., Chang, S. 2003. Temporal properties of spontaneous speech—a syllable-centric perspective. Journal of Phonetics 31(3-4), 465-485.

Gregory, H. 2014. The architecture of speech production and the role of the phoneme in speech processing. Language, Cognition and Neuroscience 29:1, 2-20.

Hickok, G. 2014. The architecture of speech production and the role of the phoneme in speech processing. Language, Cognition and Neuroscience, 29(1), 2-20.

Themistocleous, Ch. 2014. Edge-tone effects and prosodic domain effects on final lengthening. Linguistic Variation 14(1), 129-160.

Themistocleous, Ch. 2017. Effects of Two Linguistically Proximal Varieties on the Spectral and Coarticulatory Properties of Fricatives: Evidence from Athenian Greek and Cypriot Greek. Frontiers in Psychology 8, 1945.

Influence of semantics on the perception of corrective focus in spoken Italian

Sonia Cenceschi, Licia Sbattella, Roberto Tedesco
Dept of Electronics, Information and Bioengineering, Politecnico di Milano, Italy
https://doi.org/10.36505/ExLing-2018/09/0006/000339

Abstract

This study is a web-based, psychoacoustic test for adult, Italian native-speakers, investigating detection of different prosodic phenomena in Standard Italian utterances. The purpose was to investigate the influence of semantics on human ability to recognise different prosodic aspects, in order to understand the basic pieces of information involved into the psychoacoustic process of verbal comprehension. In particular, one section of the test regarded the ability to recognize the presence of a Corrective Focus, which is a spoken constituent that is a direct rejection of an alternative. Results show Corrective Focus seems difficult to detect into isolated audio utterances. Semantics seems to improve detection accuracy; phonotactics, instead, seems not to add useful information; finally, our test confirms correlation with prominent syllables.

Key words: contrastive focus, prosody, perception, semantics, psychoacoustics

Introduction

Usually, speaker express emotions or introduce a new topic/concept into the dialog by adding acoustic stress or pronouncing more clearly one or more syllables of their speech. In the context of a dialog, the Corrective Focus (CF) is a particular kind of stress, where the current speaker's intention is to correct a concept introduced by the other speaker in the previous dialog turn (Gussenhoven 2008). The acoustic realization of CF depends on culture and language of the speaker (Bosch and van der Sandt 2009).

The Standard Italian language is strongly syllable-timed: syllables take approximately an equal amount of time to be pronounced, and they are temporally stretched by speakers when they intend to underline a word. Prominence is also characterized by changes in fundamental frequency excursion and intensity-related parameters, with respect to their average values. Finally, listener's expectations (i.e., how the speaker believes her/his interlocutor will react) affects how prominence is perceived (Tamburini, Bertini and Bertinetto 2014).

The aim of this study is to investigate how much semantics – conveyed by syntax and lexicon – and phonotactics affect the human ability to detect CF in real sentences. This experiment focuses on isolated sentences, so the listener's expectations are not considered because they can only be detected in dialogs.

The experiment

The subjects were presented with simple questions; in particular, for the test on CF, the question was: *In which one of these audios do you perceive the wish to correct the interlocutor? – Example: "I want BREAD, not meat"* (the uppercased word was the one with CF). Then, the subjects started hearing three audio fragments – selected at random in two sets: one containing utterances (more precisely, Units of Intonation - UIs) with CF, and one with all other UIs – and checked audios that she/he recognized as carrying a CF. The interface allowed the subject to hear each audio fragment several times.

The UIs were taken by the SI-Calliope corpus (Cenceschi, Sbattella and Tedesco 2018), recorded by professional speakers (i.e., the corpus contains recited speech), 7 women and 7 men. Each test was conducted for three UI typologies: *real words* (where the audio fragments contained regular Italian words, and thus preserved all the syntax, lexicon, phonotactics, and acoustic information), *pseudo-words* (where the audio fragments contained invented words, with a "sound" similar to the one of real Italian words, and thus only preserved phonotactics and acoustic features, while removing syntax, and lexicon), and *pitch envelopes only* (where the audio was restricted to a pitch contour, reducing the acoustic features to a minimum, and removing all other pieces of information). We also collected subjects' age, region, and gender.

To generate pseudo-word UIs, we started from the CoLFIS corpus of Italian words (Bertinetto 2005), where we removed every word containing characters in the {w, y, j, k, x} set, and every word containing characters with diacritical signs different from acute and grave accents. Then, the remaining words where split into syllables by means of Hyphenator 0.5.1 (Berendsen 2013), a Python module that leverages the OpenOffice hyphenation dictionary. Finally, we trained a trigram of syllables that thus encoded an approximation of the Italian phonotactics. Given a real-word UI, the algorithm leverages the trigram to generate, for each word, a random pseudo-word composed of the same number of syllables. As an example, from the real-word UI "*Domani è bel tempo!*" we obtained the pseudo-word UI "*Selèzio è bel àmmi!*".

Pitch envelopes were computed with the Praat (Boersma 2002) to Pitch command, smoothed with a value of 5Hz, and then used to generate a sound by means of the `hum` command.

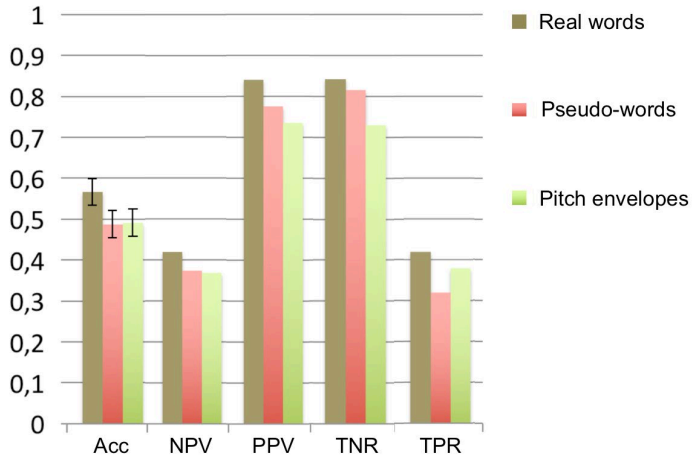


Figure 1. Accuracy (error bars: binomial at 95%), Negative Predictive Value, Positive Predictive Value, Specificity (TNR), and Sensitivity (TPR).

Results

We collected 306 tests. We found a possible correlation between accuracy in recognizing CF and the subject's origin; data, however, were not conclusive. We did not find relevant correlations with subjects' age and gender.

Figure 1 shows statistical results about perception of CF. The overall results showed that recognising CF required a combination of semantics and vocal clues. In particular, the *Accuracy* was 0.566 for real words, 0.488 for pseudo-words and 0.491 for pitch envelopes. The t-test confirmed with $p<0.001$ that UIs with real words were simpler to understand; accuracies of pseudo-words and pitch envelopes, instead, were not significantly different. This result highlights how important the "prediction" process – allowed by semantics – is in perceiving CF; our test also confirms that the fundamental frequency envelope affects the perception of CF (Terken 1991), while other acoustic features and the phonotactics do not add further information.

The *Negative Predictive Value* (the fraction of UIs recognized as not carrying CF, which are actually not carrying CF) is much lower than the *Positive Predictive Value* (the fraction of UIs recognized as carrying CF, which are actually carrying CF). Moreover, the *Specificity* (the fraction of UIs not carrying CF, which are correctly identified as such) is much higher than the *Sensitivity* (the fraction of UIs carrying CF, which are correctly identified as such). This behaviour is found in all UI typologies.

These results suggest subjects were very selective in perceiving CF: they tend not to asses its presence unless it was clearly perceivable. In this way they often missed a CF but were rarely wrong in recognizing it.

Conclusion and future works

CF seems very difficult to detect into isolated UIs; this could be justified by the fact that CF exists because there is a *dialogue*, and so it is probably better perceived if contextualized (Kakouros and Räsänen 2016). Thus, in a future experiment we could provide the subjects with a dialogue where the last UI could carry the CF. Moreover, CF recognition showed similar results for pseudo-words and pitch envelopes; this result confirms that CF is related to syllable's prominence and thus to the F0 contour and duration, while other acoustic clues and phonotactics do not add useful information. Anyway, semantics seems to play a crucial role, as real-word UIs reached a (slightly but measurable) better accuracy. Finally, we did not find relevant correlations with subjects' age and gender, while there were hints of a possible correlation with subjects' geographical origin.

Reference

Berendsen, W. 2013. Available on: https://pypi.python. org/pypi/hyphenator/0.5.1/.

Bertinetto, P.M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., Thorton, A.M. 2005. CoLFIS. Corpus and frequency lexicon of written Italian http://linguistica.sns.it/CoLFIS/Home.htm.

Boersma, P. 2002. Praat, A system for doing phonetics by computer. Glot international 5, 41-345.

Bosch, P., van der Sandt, R. (Eds.). 1999. Focus: Linguistic, cognitive, and computational perspectives. Cambridge University Press, 30-31.

Cenceschi, S., Sbattella, L., Tedesco, R. 2018. Towards Automatic Recognition of Prosody. Proc. 9th International Conference on Speech Prosody, 319-323, Poznań, Poland.

Gussenhoven, C. 2008. Types of focus in English. In Topic and focus, 83-100. Springer.

Kakouros, S., Räsänen, O. 2016. Perception of sentence stress in speech correlates with the temporal unpredictability of prosodic features. Cognitive science 40, 1739-1774. Wiley Online Library.

Liberman, M., Pierrehumbert, J. 1984. Intonational invariance under changes in pitch range and length. In Aronoff, M, Oehrle, R. (eds.) 1984, Language and Sound Structure, 157-233. Cambridge: MIT Press.

Tamburini, F., Bertini, C., Bertinetto, P.M. 2014. Prosodic prominence detection in Italian continuous speech using probabilistic graphical models. Proc. Speech Prosody, 285-289, Dublin, Ireland.

Terken, J. 1991. Fundamental frequency and perceived prominence of accented syllables. The Journal of the Acoustical Society of America 89(4), 1768-1776.

A semi-automatic assessment of lexical stress patterns in non-native English speech

Évelyne Cauvin[1], Laure Pairet[2]
[1]CLILLAC-ARP (EA 3967) - Sorbonne Paris Cité - Université Paris Diderot, France
[2]CURAPP (UMR 7319), Université de Picardie Jules Verne, France
https://doi.org/10.36505/ExLing-2018/09/0007/000340

Abstract

This study investigates the acquisition and assessment of English lexical stress placement by French native speakers. Previous research on the evaluation of stress patterns has been centred on selections of syllables in written words, but little is known about the learners' realisations in a spoken context. We aim to shape methodological criteria to semi-automatically assess auditory stress pattern realisations based on perception tests in read speech. By refining the methodology and running statistical tests, we have been able to find out more consistent data to create a more adequate variable to assess L2 learners' lexical stress proficiency.

Keywords: semi-automatic assessment, French learners of English, lexical stress models, evaluation grid, multidimensional analysis

Introduction

This study on lexical stress patterns in oral speech addresses intonation in the wider meaning of the suprasegmental field since *It also involves the study of the rhythm of speech, and (in English, at any rate) the study of how the interplay of accented, stressed and unstressed syllables functions as a framework onto which the intonation patterns are attached* (Wells [2006] 2009).

It is basically intended to be a response to the increasing pressure of finding ways to grade the proficiency of an international workforce in spoken English, which is in line with the *CEFRL* (2001). It paves the way to finding *criterial features* (Hawkins, Buttery 2010) in order to assess non-native spoken English at the suprasegmental level.

Extensive research has been done in lexical stress assignment and patterns, focusing on words out of context and based on data collected from dictionaries (Wells 2008, Roach *et al.* 2011). Descriptive findings have led to the formulation of rules for both native and non-native speakers of English (Guierre [1984] 1992, Duchet & Fryd 1998, Fournier 2010). Assessing lexical primary stress assignment is general practice to many teachers of English on an auditory basis and in written tests, yet disconnected from context and connected speech. Horgues (2010) used arrows to distinguish between right or left shift in learner speech. Cauvin (2017) devised a methodology to assess and semi-automatically grade lexical stress assignment proficiency in non-native speech production in context. Four specialists listened to the recordings of 15 French learners of

English, representative of 155 learners reading a text (*Longdale-Charliphonia* corpus[1]), and assessed 6 polysyllable realisations according to native productions found in reference pronunciation dictionaries (see above) as well as by listening to two native English speakers reading the same text as models. *Excel* spreadsheet formulas are the basis of the creation of a variable to assess lexical stress pattern realisations. Coining each word with right or wrong stress pattern assignment seems unsatisfactory due to a lack of adequacy with other prosodic variables. Nevertheless, it was a first step towards creating a comprehensive evaluative suprasegmental methodology.

Methodology and analysis

This research is based on Cauvin (2017)'s methodology and it analyses her data per learner, group and word more thoroughly by using statistical tests.

Factors at work in L2 lexical stress acquisition

As written above, 15 learners representative of a 155 learner corpus were selected according to their reading speed performance — three of them were selected in the first and ninth deciles, the first and third quartiles, and the median. Two more variables define the learners such as their age and time spent in an English-speaking country.

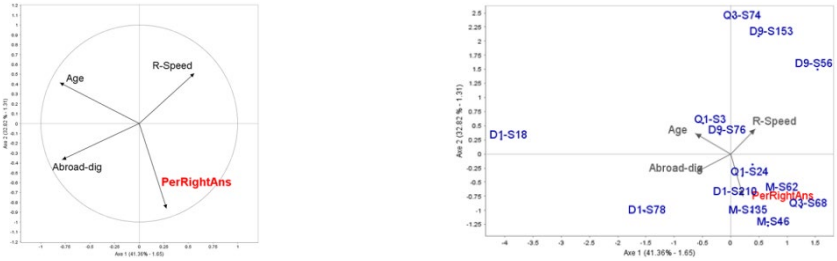


Figure 1. Representation of quantitative variables on the 1st plane of the PCA correlation circle and representation of the individuals on the first plane of the component space.

In Figure 1, the correlation circle shows that the percentage of correct answers (PerRightAns arrow) is orthogonal to the learners' reading speed (R-Speed arrow) and the time spent in an English speaking (Abroad-dig arrow). Neither the reading speed nor the time spent speaking English influences lexical stress mastery, which is acquired via specific teaching. The most proficient students are in the bottom part of the PCA factor map (Escofier, Pagès 2008).

Polysyllabic word analysis

In the reading of the *Longdale-Charliphonia* text, Cauvin (2017) selected six polysyllable words in context in Figure 2 introduced with their stress pattern[2]

(Stress Pattern), their syllable number (nSyll), their stress number (nStress), the potential presence of a prefix (Prefix), and a strong or weak suffix (Suffix).

Table 1. Polysyllabic word selection and their characteristics (their stress pattern, their syllable number, their stress number, the potential presence of a prefix, and either a strong or weak prefix).

Words	Stress Pattern	nSyll	nStress	Prefix	Suffix
Conversation	/2010/	4	2	pre	strong-suf
Limited	/100/	3	1	no-pre	weak-suf
Determined	/010/	3	1	pre	weak-suf
Trespassers	/100/	3	1	no-pre	weak-suf
Prosecuted	/1000/	4	1	no-pre	weak-suf
Musicians	/010/	3	1	no-pre	strong-suf

We devised a more complex grading scale with 5 for a correct answer, 4 for ignoring a secondary stress, 3 for creating or inverting primary and secondary stresses, 2 for stressing a syllable before the assigned placement, 1 for doing so after it and 0 for stressing the word at the end or when the stress realisation is unclear. Applying this refined assessment grid to Cauvin (2017)'s polysyllabic data resulted in a new hierarchy of word stress difficulty (Figure 3).

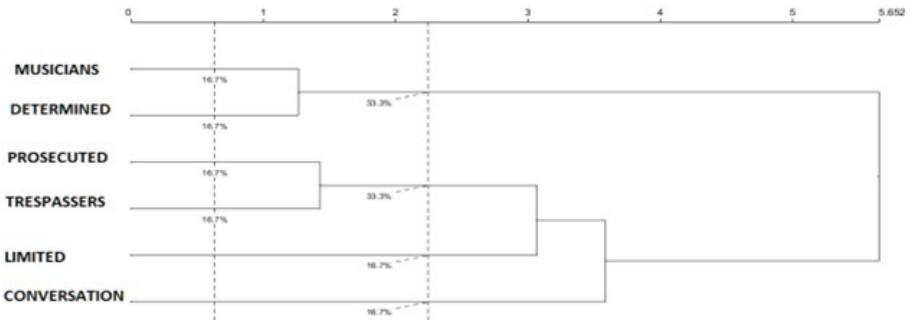
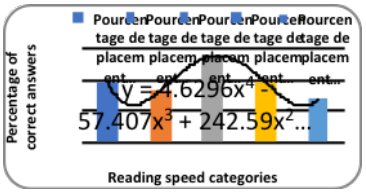


Figure 2 . Ascendant hierarchical clustering of the polysyllabic words with a dendrogram using 3 variables.

According to three assessment variables (Cauvin (2017)'s true/false grading scale, the current grading variable with six marking degrees from 0 to 5, and the sum of the learners' answers marked differently from 0), this hierarchical clustering analysis of the words shows that *musicians* and *determined* present a similar degree of difficulty, with the lowest success scores. At the bottom end of the dendrogram, *conversation* remains apart since it gathers the best results. *Limited* is next, and *trespassers* and *prosecuted* are grouped together at a higher difficulty level.

Discussion and conclusion

Contrarily to the 2017 true/false grading scale (Figure 4a), the results of the new one (Figure 4b) follow a normal distribution, similar to that found in most histograms depicting other prosodic variables, which shows a greater methodological consistency in assessing lexical stress. Finding more specific data in lexical stress assignment refines the semi-automatic assessment model.

(a) 

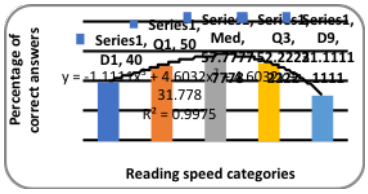
(b) 

Figure 3. Assessing quality realisations per reading speed category (in percentage) with a) a 0-1 grading scale and b) a 0-5 grading scale

Notes

1. *Longdale-Charliphonia*: [http://www.clillac-arp.univ-paris-diderot.fr/projets/]
2. Guierre ([1984] 1992)'s methodology: "1" for primary stress, "2" for secondary stress and "0" for an absence of stress.

References

Cauvin, É. 2017. Élaboration de critères prosodiques pour une évaluation semi-automatique des apprenants francophones de l'anglais. Paris 7.

Duchet, J.-L., Fryd, M. 1998. Manuel d'anglais oral pour les Concours. CNED: Didier Érudition.

Fournier, J.-M. 2010. Manuel d'anglais oral. Paris: Ophrys.

Guierre, L. [1984] 1992. Drills in English Stress-Patterns. Paris: Armand Colin-Longman.

Hawkins, J.A., Buttery, P. 2010. Criterial Features in Learner Corpora: Theory and Illustrations. English Profile Journal, 1, e5. (40-62).

Horgues, C. 2010. Prosodie de l'accent français en anglais et perception par les auditeurs anglophones. Université Paris Diderot - Paris 7.

Wells, J. C. [2006] 2009. English Intonation: An Introduction. CUP.

Aspirated voiceless stops in elderly speakers from Calabria: a pilot study

Manuela Frontera
Phonetics Laboratory, Department of Languages and Education Sciences, University of Calabria, Italy
https://doi.org/10.36505/ExLing-2018/09/0008/000341

Abstract

The aim of this preliminary study is to depict a first sketch of VOT realizations of elderly speakers from the province of Catanzaro (central Calabria, southern region of Italy), to be compared in the near future with data belonging to the Italian-Argentinean community. Thus, the present work represents the starting point of a broader research, aimed at exploring possible cases of phonetic attrition in VOT values of Calabrian migrants in Argentina. Focusing on elderly speakers, /p t k/ stop consonant productions will be analysed within different linguistic contexts. Long lag VOT values are expected to be found, especially if compared to those belonging to adolescent speakers of the same province.
Key words: voiceless stops, VOT, Calabrian varieties

Introduction

Post-aspiration of voiceless stop consonants in Calabrian regional variety is a reported feature since the late 60's. The first studies on the case were based on a perceptual approach, by which the phenomenon seemed to intrude upon the main regional provinces in different ways and diverse phonetic contexts (see Rohlfs, 1966; Falcone, 1976; Canepari, 1986). Later on, acoustic studies on the production of geminated /p t k/ consonants across Italy have demonstrated that adult speakers from the city of Catanzaro show the longest VOT values at all (Stevens, Hajek, 2010). Nevertheless, the unique study focused on VOT analysis in other productive phonetic contexts (post-nasal and post-liquid /p t k/), in the same area, has been led on adolescent speakers only, and from a socio-phonetic perspective, showing an attitude to "control" aspiration (considered as spy of a stigmatized variety), in favor of a more standardized pronunciation (Nodari, 2015).

Hence, it has been considered appropriate to try to offer a wider and complete view of this feature in the aforementioned area. Thus, a focus on a more conservative age range seems likely to provide a more objective view of the phenomenon. The present study represents, then, a first step into this new investigation.

Methods

Participants

This work reports the case of the first two speakers involved in the ongoing research, a male and a female, aged 70 and 82 and living in the province of Catanzaro. They have spent their whole life in the same country (Settingiano), speaking dialect as a first language and Italian as a L2, which are the only languages they know and speak.

Corpus

Speakers were recorded using a Scarlett CM25 microphone and a Focusrite Scarlett 2i2 audio interface via Sony Sound Forge Pro 11, with a sampling frequency of 44.100 Hz and a 24 bits resolution.

Speech material is elicited through a reading task consisting of 54 isolated words and 12 sentences in Italian language (232 stimuli). The target words all contain voiceless stops /p t k/ and subsume geminate, post-liquid (-lp, -lt, -lk, -rp, -rt, -rk) and post-nasal (-mp, -nt, -nk) contexts, equally alternating point of articulation, stressed/unstressed syllables (pre-tonic and post-tonic position in paroxytones and proparoxytones) and adjacent vowels (/i/, /a/, /u/), since each of these contexts intrinsically determines higher or lower VOT values.

Analysis

All productions are labelled and phonetically annotated, following Abramson and Whalen (2017)'s procedure. Measures are automatically extracted in Praat (Boersma & Weenink, 2018), summing up REL (release) and ASP (aspiration) durations in each target sound (Kang, Whalen, 2017). VOT is measured starting from the first release burst of the stop consonant up to the acoustic periodic onset of the following vowel (Harrington, 2013). Global mean values and standard deviations associated to different consonants and contexts are then calculated.

Results

The extracted values are compared according to C place, adjacent vowels, phonetic contexts and stress position in the target word. In the present study, geminate stops will be analysed only within the phrasal context. Further analyses and possible interactions between variables will be taken into account in future investigations.

Consonant place

In line with what found by Cho & Ladefoged (1999), the more backward the consonant place of articulation is, the longer lasts its VOT duration: /k/ mean duration in isolated words is of 78 ms (st. dev.=23) against the pick of 87 ms (st. dev.=26) in /k/ consonants pronounced in phrasal context. The sound /t/ gets a 61 ms mean VOT in the word list (st. dev.=20) and 54 ms (st. dev.=19)

in the sentences, while /p/ mean duration is equal to 59 ms (st. dev.=19) in isolated words, 52 ms (st. dev.=22) in phrasal context.

Almost all consonants show long realizations, notably higher than those found both in the spontaneous speech of adolescent speakers from the same Calabrian province (Nodari, 2015) and in the word list from adult speakers of Catanzaro (Stevens, Hajek, 2010).

Adjacent vowel

Mean VOT values show longer durations when the consecutive vowels are /i/ and /u/, in both diaphasic contexts: 76 ms vs. 66 ms for /i/ (st. dev.=29 and 28), 74 ms and 70 ms for /u/ (st. dev. =21 and 27). VOT before /a/ is averagely equal to 59 ms (st. dev.=21) in the isolated words, 51 ms (st. dev.=21) in phrasal context. Thus, this data mirrors the trend of stop consonants to reach longer VOT durations if followed by closed or high vowels (Morris, McCrea, Herring, 2008). Even in this case, post-aspiration is greater than the one produced by younger speakers (max. value=43 ms).

Phonetic context

Among the main contexts of aspiration, the post-lateral position (-lC) seems to intrude upon VOT values slightly more than the post-nasal and post-rhotic ones, above all in isolated words (Table 1). The obtained values almost double the means found in Nodari (2015).

Table 1. Voiceless stop consonants' VOT in different phonetic contexts.

	VOT MEANS (MS)	
CONTEXT	WORD LIST	SENTENCES
-NC/-MC	67 (18)	60 (27)
-LC	72 (23)	66 (24)
-RC	69 (20)	65 (28)
CC	-	58 (27)

Stress position

Dealing with the differences between stressed and unstressed syllables, the data from elderly speakers confirms the presence of longer VOT values in post-tonic position, both in paroxytones (73 ms, SD=28; 63ms, SD=26) and proparoxytones (72 ms, SD=26; 67ms, SD=47), with respect to the pre-tonic positions (min value=59 ms).

Conclusions

The reported results give rise to some noteworthy remarks: firstly, the obtained data matches the trends tested in the previous works on VOT variation, and confirms the effects of post-liquid/post-nasal positions on /p t k/'s degrees of aspiration. Second, most importantly, elderly speakers seem to show post-aspiration levels definitely higher with respect to their younger compatriots, in

every investigated context. In order to confirm these important preliminary results, data from a greater amount of speakers and further specific analyses will be provided in future investigations.

References

Abramson, A.S., Whalen D. 2017. Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. Journal of Phonetics 63, 75-86.

Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved 11 May 2018 from http://www.praat.org/

Canepari, L. 1986. Italiano standard e pronunce regionali, Padova: CLEUP.

Cho T., Ladefoged P. 1999. Variation and universals in VOT: evidence from 18 languages. Journal of Phonetics 27, 207-229.

Falcone, G. 1976. Calabria, in M. Cortelazzo (Ed.) Profilo dei dialetti italiani, Pisa: Pacini.

Harrington, J. 2013. Acoustic Phonetics. In Hardcastle, W.J., Laver, J., Gibbon, F.E. (Eds.) The Handbook of Phonetic Sciences (2nd Ed.), 81-129. Chichester: Blackwell Publishing.

Kang, J., Whalen, D.H. 2017. get_vot. In https://github.com/HaskinsLabs/get_vot.

Morris, R.J., McCrea, C.R., Herring, K.D. 2008. Voice onset time differences between adult males and females. Journal of Phonetics 36, 308-317.

Nodari, R. 2015. Descrizione acustica delle occlusive sorde aspirate: analisi sociofonetica dell'italiano regionale di adolescenti calabresi. In Vayra, M. Avesani, C., Tamburini, F. (Eds.), Il farsi e disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio, Studi AISV 1. 139-153. Milano: Officinaventuno.

Rohlfs, G. 1966. Grammatica storica della lingua italiana e i suoi dialetti (I). Torino: Einaudi.

Stevens, M., Hajek, J. 2010. Post-aspiration in standard Italian: some first cross-regional acoustic evidence. Proc. Interspeech 1557-15. Makuhari, Japan.

Prosodic accuracy and foreign accent in cultural migrants

Manuela Frontera[1], Emanuela Paone[2]
[1]Department of Languages and Education Sciences, University of Calabria, Italy
[2]Department of Humanities, University of Calabria, Italy
https://doi.org/10.36505/ExLing-2018/09/0009/000342

Abstract

The present study examines the relationship between language learning motivation and prosodic accuracy in L2 Italian speakers. 4 Romanians and 4 Arabs are selected among cultural migrants living in Italy. Each L1 group includes learners with moderate and high motivational indexes, 2 inexperienced and 2 experienced speakers (indexed by length of residence). Non-native subjects' prosodic accuracy is assessed on 4 declarative sentences in Italian, through acoustic measurements and a perception test. The same sentences are produced by 4 native Italians, acting as a control group. Preliminary results suggest a possible correlation among prosodic accuracy, motivation indexes and LOR levels.

Key words: prosody, foreign accent, motivation, cultural migrants, L2 Italian.

Introduction

Several studies have highlighted that motivation plays an important role in second language acquisition process, strikingly in cultural migrants. Previous studies have claimed that suprasegmental features, as well as segmental features, may lead to the perception of a foreign accent in speech (Flege, 1988; Major, 2001). SLA research has underlined that the development of these aspects represents a huge obstacle for L2 learners, especially in adulthood , even though different investigations suggest that high linguistic motivation levels can play a significant role to achieving native-like pronunciation (Moyer, 1999; Bongaerts et al. 1997).

This study will therefore address the following research questions:

1. Do high motivational levels trigger better prosodic accuracy?
2. How do NN and N speakers' prosodic trends diverge from one another?
3. Is foreign accent still detectable in highly motivated learners?
4. Do other variables, such as the learners' L1 typology and the length of residence have an effect on their prosodic accuracy and consequently on the perceived foreign accent?

Experimental methodology

Speakers

Eight participants, 7 woman and 1 man aged 22-36 years, 4 Romanians and 4 Arabs, are selected among cultural migrants living in Italy. Related sociolinguistic information and indexes of linguistic motivation are provided by a survey from a previous study. Each L1 group includes learners with moderate and high motivational indexes (2+2), 2 inexperienced and 2 experienced speakers (indexed by length of residence, less or more than 3 years). Four native Italian speakers are selected as a control group.

Stimuli

Speech material is elicited through a semi-rigid reading task consisting of 4 declarative sentences in Italian (5 repetitions per speaker). The target utterances are: i) *Paolo mette poco zucchero nel caffè*; ii) *Marco ha molte zie a Piacenza*; iii) *Mi piace visitare i porti e le stazioni*; iv) *Stiamo perdendo troppe razze di animali*.

Task and participants

77 native Italians are asked to judge the degree of foreign accent perceived in 40 sentences (32 by the cultural migrants group, 8 by the control group), using a 3 point Likert-scale (from *no foreign accent=0* to *strong foreign accent=2*).

Analysis

Prosodic accuracy

First, values of total duration, silent pauses' duration, f0 max and min are extracted from native and non-native declaratives. Articulation rate (syll/sec excluding pauses) and tonal pitch range are then calculated and statistically compared. Pitch range (that is the difference between the highest and the lowest f0 values in an utterance) is calculated in Hertz and then converted to semitones (ST), since this scale best reflects the intonational equivalence and allows for a normalization across gender (see Nolan 2003).

Data is further analysed through ANOVAs, in order to assess possible correlation among prosodic adequacy (articulation rate and tonal range), motivational indexes, LOR levels and L1s.

Perceptual test

Mean scores and standard deviations of perceived foreign accent related to every L2 Italian speaker are extracted. Variations in global mean scores are tested depending on motivational indexes, levels of LOR, L1s and their interaction.

Finally, the relation between acoustic values and scores of perceived accent is observed.

Results

The Anova exploring the effect of the speakers (i.e. Romanians, Arabs and Italians) on the articulation rate and tonal pitch range showed it was significant on both the first parametre [F (2, 45)= 17,69, p= .000] and on the tonal pitch range [F (2, 40) = 3,66, p= .03]. An unpaired two sample t-test confirmed that there were no significant differences between the Italian and Romanian groups [t(29)= 2,04, p= .29]; indeed, articulation rate was quite similar (Rom.: M= 5,4 syll./sec, SD= 1,1; Ita.: M= 5 syll/sec, SD= 1,2). On the contrary, Arabs' articulation rate was much slower (M=3,4; SD=0,7), causing a significant difference with respect to Italians' articulation rate, t (23)= 2,06, p =.000. As regards tonal pitch range, native Italians employed a wider range if compared to non native-speakers' (Ita.: M=11,6 ST; SD= 1,9; Rom.: M= 9,3 ST, SD=3,8; Ar.: M=9,6; SD= 2,5). Significant differences were found between Arabs and Italians [t(20)= 2,08, p= .000] but not between Italians and Romanians [t(13)= 2,1, p= .2].

Among Arab learners significant differences were found between speakers with high and moderate motivational level, with regard to tonal pitch range only, (highly motivated Arabs: M=10,7 st; SD=1; t(8)=2,1; p=.06). Neither motivational index nor LOR had an effect on articulation rates. Moreover, variability on Romanians' articulation rate and tonal pitch range seems not to be affected by motivational levels or LOR.

As regards the perceptual test, global means of perceived foreign accent set into a 0,36-1,89 range, where the first value is associated to Romanian L1, moderated linguistic motivation and long LOR, while the latter belongs to a Romanian migrant with moderated linguistic motivation and low LOR. Indeed, statistical analyses reveal that there is no a significant interaction among the given variables (F[1,246]=00,13, p>0.5). Nevertheless, L1 and motivational index being equal, a greater LOR is generally associated to a weaker foreign accent, while lower motivational indexes are crucial in determining a stronger perceived foreign accent when L1 and LOR are similar.

Globally speaking, in two out of three cases, a strongly perceived foreign accent goes along with lower articulation rates (NN art.rate=2,60;3,62 vs. N art.rate=5,07) but slight tonal range differences (less than 1 semitone). Similarly, a native-like accent responds to almost equal articulation rates respect to the Italian means and does not seem to be intruded upon by greater divergences in the tonal pitch range (more than 1 semitone).

Conclusions

Results suggest that, as far as native speakers' judgments are concerned, it is not possible to say there is a stable correlation between a higher motivation and

a more native-like accent; however, motivation is crucial in determining a weaker perceived foreign accent when L1 and LOR are similar.

Romanian prosodic trends are globally close to native Italian values, regardless their motivational index or LOR, i.e. their prosodic accuracy (articulation rate and tonal pitch range) did not vary according to these variables, but is already quite similar to natives'. On the contrary, high motivational levels trigger better prosodic accuracy in Arabs. Even in this case, different levels of LOR do not determine significant divergences in non-native productions.

References

Bongaerts, T., Van Summeren, C., Planken, B., Schils, E. 1997. Age and Ultimate Attainment in the Pronunciation of a Foreign Language. Studies in Second Language Acquisition 19, 447-465.

Flege, J. 1988. Factors affecting degree of perceived foreign accent in English sentences. Journal of the Acoustical Society of America 84, 70-79.

Major, R.C. 2001. Foreign accent: The ontogeny and phylogeny of second language phonology. Mahwah, NJ, Lawrence Erlbaum Associates.

Moyer, A. 1999. Ultimate Attainment in L2 Phonology: The critical factors of age, motivation and instruction, Studies in Second Language Acquisition 21(01), 81-108.

Nolan, F. 2003 Intonational equivalence: An experimental evaluation of pitch scales, in Solé, M.J., Recasens, D., Romero, J. (Eds.), Proc. 15th International Congress of Phonetic Sciences, 771-774.

The perception of some personality traits in female voice

Glenda Gurrado
Department of Letters, Languages and Arts, University of Bari, Italy
https://doi.org/10.36505/ExLing-2018/09/0010/000343

Abstract

The ethological model called the *Frequency Code* developed by Ohala (1983, 1984) postulates a correlation between pitch and the expression of power: low-pitched voices communicate a meaning of dominance while high-pitched voices transmit a meaning of submission. Based on this awareness, the present research aims to investigate the relationship between the pitch variation of female voice and some specific personality markers, in particular we tested the perception of dominance.

Key words: Frequency Code, female voice, pitch, dominance/submissiveness dimension, Visual Analogue Scale

Introduction

According to the Frequency Code, fundamental frequency is an instrument of transmission of emotions. In particular John Ohala (1983, 1984) argues that some 'social messages' can be communicated by means of f0: both in animals and in humans low values of f0 convey a sense of dominance, confidence and aggressiveness whereas high values of f0 transmit a meaning of submission and vulnerability. In humans the scenario is more complex because pitch influences the prosodic structure of the entire sentence. Furthermore the sexual dimorphism of the vocal anatomy of men and women plays a key role in favour of the innate character of the Frequency Code. Indeed, as is widely known, there is a correlation between the size of the vocal folds and the pitch of voice: small vocal folds, typically those of female, produce high frequencies, on the contrary, large vocal folds, typically those of adult male, produce low frequencies. This is an important aspect that contributes to differentiating the role of women and men in society and in everyday life.

Methodology

To date the female voice has often been analysed with reference to the male voice. For this reason this study aims to verify the role played by pitch variation and other prosodic aspects of female voice in the listeners' perception of dominance. The hypothesis is that lower f0 values make a voice sound more dominant, on the contrary, higher f0 values make a voice sound more subordinate. The participants are three female speakers of Bari Italian ranging from age 22 to 25. We selected a low-pitched voice (RV), a medium-pitched voice (GM) and a high-pitched voice (CI)[1]. The speakers were asked to read an

article (read speech) and to describe a cooking recipe (spontaneous speech). They were recorded by means of a digital recorder (WAV. format, 44100 Hz, 32 bit).

A perceptive analysis was conducted in order to test the way listeners judge the speakers' personality. A group of 50 Italian listeners (36 F and 14 M, average age 23.8 years) was invited to evaluate the three voices by means of five pairs of adjectives: *small/large*, *submitted/dominant*, *lacking confidence/confident*, *vulnerable/aggressive*, *desirous of goodwill/secure*. The listeners had to mark their judgments on a Visual Analogue Scale (VAS), (Chen et al. 2004)[2]. We divided the VAS in 10 parts and placed two opposite adjectives on the two ends of the line. It was predicted that listeners would evaluate the low-pitched voice (RV) as dominant by marking the slash on the right side of the VAS; on the contrary, we supposed that the high-pitched voice (CI) would be judged as submissive by being associated with the left side of the scale.

Results and discussion

The data collected so far does not show a precise predictive tendency (see Table 1.).

Table 1. Medium values (over) and standard deviation (under) of the auditory judgments made by listeners by means of five pairs of adjectives with reference to read (R) and spontaneous (Sp) speech.

	Small/Large		Lacking confidence/ Confident		Submissive/ Dominant		Vulnerable/ Aggressive		Desirous of Goodwill/ Secure		Medium values	
	R	Sp	R	Sp	R	Sp	R	Sp	R	Sp	R	Sp
CI	5.0 1.7	4.6 1.9	6.6 1.8	5.5 2.7	5.4 1.5	6.0 1.9	4.5 1.3	5.1 2.1	6.5 2.0	5.6 1.7	5.6 1.7	5.4 2.1
G M	5.3 2.5	5.9 2.6	3.2 2.0	3.4 2.5	3.3 1.6	4.0 2.3	2.8 1.7	3.7 2.5	3.1 1.7	5.4 2.7	3.5 1.9	4.1 2.4
RV	6.1 2.0	5.8 1.9	6.7 1.9	6.0 2.3	6.8 1.7	5.5 2.0	5.8 1.7	5.0 1.9	6.7 2.3	5.8 2.4	6.4 2.0	5.6 2.1

With reference to the read speech the situation is more definite, the auditory values are statistically significant [F(2.736)= 102.629, p= .000]. In this case, the judgments tend to correlate slightly the lowest-pitched voice (RV) with those aspects of personality conveying confidence and dominance (average value: 6.4, SD: 2); on the contrary the highest-pitched voice (CI) is associated with opposite personality traits, as submissiveness (average value: 5.6, SD: 1.7). Nevertheless, by comparing CI and RV, the scores relating to the adjectives *lacking confidence/confident* and *desirous of goodwill/secure* are not statistically significant. On the other hand, the judgments made by the listeners about the spontaneous speech are mostly around the medium zone of the VAS, that is 5; furthermore, in this case the differences between CI and RV are not significant

(p= .591). Unexpectedly GM, the medium-pitched voice, gets the lowest average values, both in read 3,5 (SD: 1.9) and in spontaneous speech 4.1 (SD: 2.4), (p= 0.01). In order to assess the correlation between judgements distribution and the VAS scores, frequency indexes were calculated and reported in a graph (see Fig. 1.).

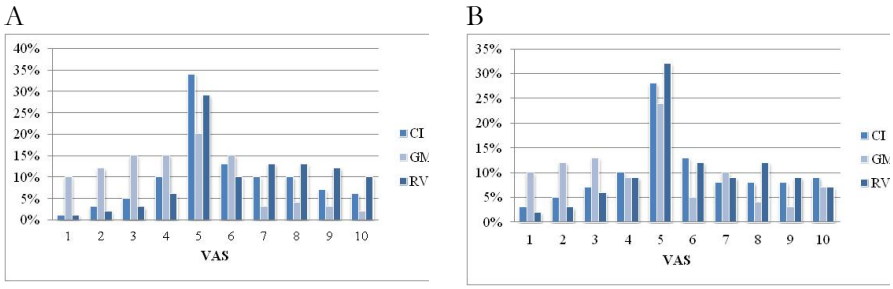


Figure 1. Percentage distribution of the perceptive judgments for read speech (A) and spontaneous speech (B)

Figure 1. (A) reveals that in read speech the distribution of values between CI and RV differs only in the left section of VAS: the 34% (CI) and 29% (RV) of all the judgements passed on these two speakers are oriented towards the centre of the VAS. Nevertheless, the 13% of the auditory values of RV reaches even the right side of the VAS (7-9 points), while for CI this percentage drops to 10%. A similar tendency is detectable also for spontaneous speech: Figure 1. (B) shows that with reference to CI and RV an important percentage of the values is oriented towards the central area of the VAS, while only the 12% for RV and the 8% for CI are placed in the right side of the VAS. A particular position is occupied by speaker GM: most of the auditory judgements is concentrated around the lowest points of the VAS, reaching a higher percentage than those achieved by the other two speakers. For read speech respectively the 12% and the 15% of the values reach the points 2 and 3 on the VAS, while for CI and RV the percentage does not exceed the 5%. Furthermore, it is remarkable how the percentage of the judgments made on GM is oriented for the 10% on the score 1, on the contrary CI and RV show a very low percentage of judgments on this side of the VAS.

The perceptive analysis revealed that the correlation between a low-pitched voice and the perception of dominance can be considered just a tendency. In fact, this particular circumstance showed a widespread uncertainty among the listeners, who were not able to form a definite opinion about the personality traits communicated by the three female voices. It can be argued that listeners' judgements were affected by several aspects: GM was perceived as the most submissive voice among the others, even though her f0 values were medium, probably the presence of a slow speaking rate and a low intensity communicated a sense of insecurity and weakness.

This study shows that the paralinguistic meanings conveyed by human voice are not influenced only by pitch variation but also by other prosodic aspects like speech rate, pauses, intensity. In order to clarify this issue, in the future we intend to focus specifically on the interaction between these parameters and verify if there are some other variables involved in the perception of personality. In particular a further research will necessarily include the synthesis of voice for the purpose to identify precisely the aspects involved in the transmission of dominance.

Notes

1. The acoustic analysis revealed that the three voices are actually different in terms of f0. Both in read and spontaneous speech, RV presents the lowest f0 values whereas CI shows the highest ones; GM can be considered a medium-pitched voice. The other parameters that contribute to differentiate the three voices are speech rate (CI is the fastest one, followed by RV and GM, who is the slowest one) and intensity (RV shows the highest values, followed by CI and GM). All data are available in Gurrado, Sorianello (2017).
2. VAS is a graduated scale horizontally or vertically oriented, 100 mm longer, generally adopted in clinical research for analysing something that cannot be precisely measured, as pain or mood.

References

Chen, A., Gussenhoven, C., Rietveldt, T. 2004. Language specificity in perception of paralinguistic intonational meaning. Language and Speech 47(4), 311-349.

Gurrado G., Sorianello P. 2017. La percezione della dimensione di dominanza/sottomissione in un campione di voci femminili. In Bertini, C., Celata, C., Lenoci, G., Meluzzi, C., Ricci, I. (eds) 2017, Origini e funzioni della variazione fonetica. Fattori biologici e sociali a confronto, Studi AISV 3, 189-210. Milano: Officinaventuno.

Ohala, J.J. 1983. Cross-language use of pitch: an ethological view. Phonetica 40, 1-18.

Ohala, J.J. 1984. An ethological perspective on common cross-language utilization of F0 of voice. Phonetica 41, 1-16.

Arabic character diacritization using DNN

Ikbel Hadj Ali, Zied Mnasri, Zied Lachiri
Signal, Image and Technology of Information Laboratory, University Tunis El Manar, Tunisia
https://doi.org/10.36505/ExLing-2018/09/0011/000344

Abstract

In this paper, automatic Arabic character diacritization is more accurately achieved using deep neural networks. Actually, though diacritic signs represent short vowels and/or indicate gemination on consonants, they are omitted in modern standard Arabic (MSA). However, most speech processing applications like speech synthesis and machine translation need such marks to convey the right meaning. Therefore in this work, automatic diacritization accuracy is enhanced using feedforward DNN. The results show that using more significant and Arabic-specific input features increases the prediction accuracy of diacritic signs.
Key words: Arabic characters, diacritic signs, feedforward DNN, input features.

Introduction

Modern standard Arabic (MSA) is natively spoken by more than 300 million people in the MENA region (Middle East and North Africa). Therefore, it's urgent to catch up the development of NLP (Natural Language Processing) to keep Arabic present in novel speech technology products. In Arabic, short vowels and geminated consonants are indicated by diacritic signs. However, in contrary to classical literary Arabic, diacritic signs are mostly omitted in modern standard Arabic MSA. Arabic speech processing applications, especially speech synthesis and machine translation, need accurate diacritization to convey the right meaning. This problem was addressed a particular attention, in many related works using linguistic and grammatical rules. Therefore, in this paper, machine learning, and especially deep neural networks are used to increase the accuracy of Arabic character diacritization.

The rest of this paper is organized as follows: section 2 introduces the Arabic phonology and linguistics; section 3 presents a brief description of the deep neural networks used in this work; section 4 shows the speech material, the conducted experiments and the yielding results. Finally the findings are discussed and commented.

Arabic diacritics

Arabic is a Semitic language which has the advantage of having a single literary version. Though there are many dialects (colloquial Arabic) which differ, not only from one country to another, but also from one region to another in the same country, there's only one standard version, which is used in literature,

journalism and science. This standard version, called MSA (Modern Standard Arabic) had inherited from Classic Arabic, which used to be the literary version since the middle ages.

Arabic diacritic signs

Arabic has 28 consonants and three vowels. One of the specific characteristics of Arabic is the ability to double all consonants (gemination) and to lengthen all vowels. However, this leads to changing the word meaning, e.g. the word *"darasa"* دَرَسَ means (to study), whereas with a geminated *"r"* it becomes *"darrasa"* دَرَّسَ (to teach) and with a long final "a" it changes to *"darasa:"* دَرَسَا (to study with somebody else).

Arabic alphabet does not include special letters for short vowels. However, these vowels are marked on consonants using three diacritic signs, i.e. *"fatha"* for /a/, *"dhamma"* for /u/ and *"kasra"* for /i/. Doubling theses diacritics at the end of a noun indicates the indeterminate form (cf. Table 1). Besides, *"sukun"* and *"shaddah"* are used to indicate stop and gemination, respectively.

In addition, different ways of diacritization of the same word change the meaning, e.g. the word consonant-based root ("d,r,s" د,ر,س) may be pronounced *"darasa"* دَرَسَ (to study) or *"durisa"* دُرِسَ (to be studied) depending on the diacritics representing the vowels introduced after each consonant

Arabic diacritization systems

Many techniques were used to achieve this task. However, all of them can be divided into three main categories, i.e. rule-based, model-based and data-driven techniques. Rule-based techniques rely on the implementation of linguistic rules to determine the diacritic sign of each character (Halabi 2016), whereas model-based techniques use n-gram language models to predict the diacritic sign of each character (Habash 2009). More recently, with the development of data-driven prediction tools, probabilistic learning techniques like HMM (hidden Markov models) were applied to predict the diacritic sign based on a set of contextual features (Rashwan 2011).

DNN in speech processing

DNN are nowadays used in most speech processing applications, such as speech synthesis, speech recognition and machine translation. In speech processing, DNN can be used either for regression, to predict continuous values, such as segment duration or fundamental frequency (pitch) values, or for classification tasks, such as segment voicing decision or diacritization sign prediction. Therefore, it has been recently performed using DNN (Rebai and Ben Ayed 2015). However results need to be enhanced using more significant characteristic features.

Experiments

Implementation

A feedforward DNN using 2 hidden layers and sigmoid activation function was used for Arabic character diacritization. In the preprocessing phase, the selected input features (cf. Table 1), were transformed into a specific code for each, whereas output targets (diacritic signs) were encoded using one-against-all code, since the task is multi-class classification. During the learning process, early stopping option was used to prevent over-fitting.

Table 1. Features classification and coding.

Feature	Value	Coding	Nodes
Identity (of previous/ current/next letter)	/a/ا , /b/ب, /t/ت …etc.	One against all	36
Type (of previous/ current/next letter)	Plosive, fricative, nasal, trill, lateral, semi-vowel…	One against all	8
Gemination (of previous/ current/next letter)	Yes/No	Binary	1
Relative position of current letter in the word	beginning/middle/ end	Coarse coding	3
Relative position of current word in the sentence	beginning/middle/ end	Coarse coding	3
Content word	Yes/no	Binary	1

A database of 28737 sentences fully-diacritized containing ca. 1.6 million characters was used. 80% of the database was allocated for training whereas the remaining 20% were used for validation and test.

Results and discussion

To assess the accuracy of DNN-based diacritization, two measures were used, i.e. total accuracy rate (TAR), calculated all over the characters in the validation set, and class-wise accuracy rate (CAR), which is calculated for each single diacritic sign using the confusion matrix. The best DNN model, tested on the validation set, has given a TAR of 84.4%. Furthermore, the matrix confusion was calculated to extract the CAR for each diacritic sign. For some signs, like *sukun (a stop on a consonant)* the CAR has reached more than 90% (cf. Table 2). Also, for letters which need no diacritic signs (like long vowels, e.g. /a:/ ا, /u:/ و and /i:/ ي) the CAR related to the class *None* was very high, 92.7% (cf. Table 2).

Table 2. Accuracy results of DNN-based diacritic signs prediction model.

Diacritic sign	Tested samples	Recognized samples	Accuracy
Fatha /a/ (ـَ)	62596	43134	68.9 %
Dhamma /i/ (ـِ)	22068	15509	70.2 %
Kasra /u/ (ـُ)	112449	98873	87.9 %
Fathaten /an/ (ـً)	269	109	40.5 %
Dhammaten /un/ (ـٌ)	14	9	64.3 %
Kasraten /in/ (ـٍ)	1170	888	75.9 %
Sukun (stop) (ـْ)	41188	37339	90.6 %
None	80246	74438	92.7 %
Total	320000	270299	84.4 %

However, the prediction accuracy needs to be enhanced for some classes like *kasra* /i/, *fatha* /a/ and *dhamma* /u/ that are essential to understand the meaning of the word. Also, less abundant diacritics like *fathaten* /an/ (ـً), *dhammaten* /un/ (ـٌ) and *kasraten* /in/ (ـٍ) need to be better modelled to enhance their prediction accuracy. Therefore, mono- and bi-directional long short term memory (LSTM and B-LSTM) deep neural networks might be used to take advantage of the recurrent aspect of speech.

References

Habash, N., Rambow, O., Roth, R. 2009. MADA+ TOKAN: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, POS tagging, stemming and lemmatization. In Proceedings of the 2nd international conference on Arabic language resources and tools (MEDAR), Cairo, Egypt (Vol. 41, p. 62).

Halabi, N., Wald, M. 2016. Phonetic inventory for an Arabic speech corpus. In Proceedings of the Tenth International Conference on Language Resources+ and Evaluation (LREC 2016), Slovenia, 734-738.

Rashwan, M.A., Al-Badrashiny, M.A., Attia, M., Abdou, S.M., Rafea, A. 2011. A stochastic Arabic diacritizer based on a hybrid of factorized and unfactorized textual features. IEEE Transactions on Audio, Speech, and Language Processing, 19(1), 166-175.

Rebai, I., BenAyed, Y. 2015. Text-to-speech synthesis system with Arabic diacritic recognition system. Computer Speech & Language, 34(1), 43-60.

INTSINT: a new algorithm using the OMe scale

Daniel Hirst
Laboratoire Parole et Langage (LPL), CNRS and Aix-Marseille University, France
https://doi.org/10.36505/ExLing-2018/09/0012/000345

Abstract

This presentation reports work in progress on an improved and simplified algorithm for coding the output of the Momel algorithm using the INTSINT alphabet, building on recent work which proposed the Octave-Median scale (ome = log2(Hz/Median)) as a natural scale for the representation of pitch. Preliminary results comparing the output of the new algorithm with that of the standard version shows that more values are less than 1 semitone from the Momel output and the RMSD value is also lower. Further work is needed to improve this new algorithm.

Key words: Intonation, symbolic coding, INTSINT, algorithm, evaluation

Introduction

Official presentations of ToBI (e.g. Beckman et al 2005) have adopted the position that *symbolic tone labels in the ToBI framework are intended to 'tag' the intonation contour and not to 'encode' it.*The authors contrast this with the approach of INTSINT (an International Transcription System for INTonation) which had been proposed as a first approximation of a prosodic equivalent of the IPA.

The original version of the INTSINT system (Hirst 1987) was based on an inventory of minimal pitch contrasts found in published descriptions of intonation patterns. The aim was to provide a tool for the systematic description of these intonation patterns, something along the lines of a narrow transcription using the International Phonetic Alphabet (IPA) Like the IPA, it was intended that INTSINT could be used for preliminary descriptions of intonation patterns, even for languages which had not previously been described.

Notice that this aim is very different from that of the ToBI system (Silverman et al. 1992), which presupposes that the inventory of intonation patterns for the language being described has already been established.

The official website for ToBI (ToBI website) makes this particularly explicit:

Note: ToBI is not an International Phonetic Alphabet for prosody. Because intonation and prosodic organization differ from language to language, and often from dialect to dialect within a language, there are many different ToBI systems, each one specific to a language variety and the community of researchers working on that language variety.

In recent years, though, there has been a revival of interest among researchers working within the ToBI framework, in the development of such a

ExLing 2018: Proceedings of 9th Tutorial and Research Workshop on Experimental Linguistics, 28-30 August, Paris, France

coding system (Hualde & Prieto 2016). There have also been two recent workshops on the subject: one at the 2015 ICPhS, in Glasgow and another at the 2018 Speech Prosody Conference, Poznan, showing that there is a growing interest and need for a tool of this type.

This presentation reports work in progress on developing an improved and simplified algorithm for automatically coding the output of the Momel algorithm using the INTSINT alphabet, building on recent work (De Looze & Hirst 2014) which proposes the Octave-Median scale (ome = log2(Hz/Median)) as a natural scale for the representation of pitch.

Methodology

Campione et al (2000) compared different alternatives to the INTSINT algorithm and found that two versions provided closer fits to the Momel anchor points than the standard model. These versions (*Ampli3* and *Levels*), however, introduced tones, which unlike those of INTSINT, were not derived from phonological descriptions of intonation. Both models used three absolute tones (T, M, B) and six relative tones. This required optimising a total of 15 parameters. If the aim is simply to provide a close copy of the original anchor-points, it would be far simpler and more economical to code each directly in semitones, since a span of 15 semitones covers most of the pitch range of unemphatic utterances. Models like this do not provide a useful coding which could be used in a rule-based model of intonation.

In the most recent implementation of INTSINT (Hirst 2007), a Perl script is used to optimise both the automatic coding of the Momel anchor points and two parameters: *key* and *span,* which are used to interpret the coding.

In this presentation, I explore the possibility of coding INTSINT tones using the Octave-Median scale. Each tone is here defined by a formula, as in (1), using as variables only the median value of the pitch curve and/or the value of the preceding anchor-point (P). The tone *t*, for example[1], is defined as half an octave above the median, while the tone *h* is defined as the geometric mean of the preceding anchor point and the value of *t*. New values *t*+ and *b*- are introduced as extreme values more than 2 semitones above/below the values for *t* and *b*. This is motivated by the fact that there seems to be much greater variability in the value of the tones coded by *t* in the standard system than for the other tones.

m = median;	t = m * sqrt(2)	b = m / sqrt(2)
h = sqrt(P * t)	s = P	l = sqrt(P * b)
u = sqrt(P * h)	-	d = sqrt(P * l)
t+ = t * 2^(1/6)		-b- = b / 2^(1/6)

(1) Formulas for calculating pitch values corresponding to INTSINT tone labels using the new algorithm. P represents the value in Hz of the preceding anchor point.

Coding each Momel anchor-point is then simply a question of comparing the value of the current point to the current values of each tone and choosing the closest value. This represents a considerable simplification of the standard algorithm which used an iteration of 400 different codings to optimise the two parameters *key* and *span*. In the new algorithm, the value of *key* is calculated directly as the median of the pitch values and the value of *span* is taken as fixed at 1 octave.

Evaluation of this new coding algorithm is currently being carried out on recordings from the OMProDat database (Hirst et al 2013) and compared to that of the standard INTSINT coding. Results reported here are based on the analysis of the two corpora *omprodat-eng01* and *omprodat-cmn01,* each of which contain recordings of 40 5-sentence passages read by 10 speakers (5 male and 5 female).

Results and discussion

Preliminary results on two corpora of read speech, one in English and one in Mandarin Chinese show that, besides being much simpler to implement, the new algorithm gives results which are closer to the output of the Momel algorithm than with the standard version of INTSINT, when measured as RMSD (the square root of the average of squared errors in semitones) or as the number of anchor points less than one semitone from the Momel output. The results are slightly worse when measured as the number of anchor points less than 2 semitones from the Momel output for the Chinese data but not for the English data where the values were not significantly different between the two implementations.

Results for the corpus OMProDat Cmn01 (Mandarin Chinese): 10 speakers, 40 5-sentence passages.

	old version	new version	paired t	p
% < 1 st	71.35	73.85	-6.191	1.489 e-09
% < 2 st	93.88	93.03	4.323	1.949 e-05
RMSD (sts)	1.008	0.984	3.6033	3.541 e-05

Results for the corpus OMProDat Eng01 (British English): 10 speakers, 40 5-sentence passages.

	old version	new version	paired t	p
% < 1 st	83.95	87.44	-12.296	< 2.2e-16
% < 2 st	97.40	97.13	1.767	0.078 (ns)
RMSD (sts)	0.756	0.732	3.226	2.685 e-07

Notes

[1]In recent publications, I have taken the step of using lower case letters for the INTSINT symbols to distinguish them from the symbols used by other more abstract coding systems such as ToBI.

References

Beckman, M.E., Hirschberg J., Shattuck-Hufnagel, S. 2005. The original ToBI system and the evolution of the ToBI framework. In Jun, S.-A. (Ed.), Prosodic models and transcription: Towards prosodic typology, 9-54. London & New York: OUP.

Campione, E., Hirst, D.J., Véronis, J. 2000. Automatic stylisation and modelling of french and italian intonation. In Botinis, A. (Ed.), Intonation, 185-208. Dordrecht: KAP.

De Looze, C., Hirst, D.J. 2014. The OMe (Octave-Median) scale. In Campbell, N, Gibbon, D., Hirst, D.J. (Eds.), Proc. SP7, Dublin, Ireland.

Hirst, D.J. 2007. A Praat plugin for Momel and INTSINT with improved algorithms for modelling and coding intonation. Proc. XVIth ICPhS 1233-36, Saarbrucken, Germany.

Hirst, D.J. 1987. La représentation linguistique des systèmes prosodiques: une approche cognitive. Thèse de Doctorat d'Etat, Université de Provence.

Hirst, D.J., Bigi, B., Cho, H.-S., Ding, H., Herment, S., Wang, T. 2013. Building OMProDat, an open multilingual prosodic database. In Proc. TRASP, Tools and Resources for the Analysis of Speech Prosody 11-14, Aix-en-Provence, France.

Hualde, J.I., Prieto, P. 2016. Towards an International Prosodic Alphabet (IPrA). Laboratory Phonology: Journ. Association for Laboratory Phonology, 7(1), 5.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg. J. 1992. TOBI: A Standard for Labeling English Prosody. In Second ICSLP 867-870, Banff, Alberta, Canada.

ToBI website (consulted 2018-07-28) https://www.ling.ohio-state.edu/~tobi/.

Gender differences in respiratory muscular movements in reading Japanese and English texts by JL1 and JEFL

Toshiko Isei-Jaakkola[1], Keiko Ochi[2]
[1]Department of English Language and Culture, Chubu University, Japan
[2]School of Media Science, Tokyo University of Technology, Japan
https://doi.org/10.36505/ExLing-2018/09/0013/000346

Abstract

We conducted physiological experiments to examine the gender differences in (A) the respiratory muscles used, (B) stories read, and (C) language. For this purpose, we used respiratory strain-gauge transducers to measure chest and abdominal respiratory muscle movements and two short stories in Japanese (JL1) and English (JEFL), which were read by Japanese female and male subjects. We clarified that (1) there was a gender difference in the controlling respiratory muscles; (2) the Japanese males used the upper and lower chest muscles, and upper abdominal muscles more than did the females; (3) the language difference was not significant; and (4) the story difference was feasible.
Key words: Gender difference, Japanese speakers, Reading, English

Introduction

Abdominal and chest respirations in speech are common states. We hypothesized that females tend to use chest respiration. With regard to biological aspects in speech, Williams (1995) claimed that speech mechanisms are essentially linked to expiratory mechanisms. However, in speech production, expiration does not occur without inspiration. Saida (2015) suggested that respiratory muscular movements depend on inspiration and expiration in speech, respectively, but are inseparable.

Thus, we examined whether females use more respiratory chest muscles than respiratory abdominal muscles during reading texts than males by conducting physiological experiments, and particularly clarify the gender differences in (A) respiratory muscles used, (B) stories read, and (C) languages. For this purpose, we used the respiratory strain-gauge transducers (RST) to measure chest and abdominal respiratory muscle movements and provided two kinds of Japanese texts (short stories: fables) in Japanese and English, which were read by the participants.

Methodology

The reading materials consisted of fables, namely "The North Wind and the Sun" (= NW) and "Momotaro" ("A Peach Boy") in Japanese and English. Five

male and five female university students (20–25 years old) who majored in English read these two short stories five times both in Japanese and English, wearing four thin-wired RST on the upper (Chest1) and lower chests (Chest2), and the upper (Abdomen1) and lower abdomens (Abdomen2) in the sitting position (see for the detail in Isei-Jaakkola, et al., 2018). The sampling rates of the RST and speech signals were 100 Hz and 40 kHz, respectively. Then, we calculated six kinds of cross correlation (CC) between the following:

(a) CC1: Chest1 vs. Chest2
(b) CC2: Chest1 vs. Abdomen1
(c) CC3: Chest1 vs. Abdomen2
(d) CC4: Chest2 vs. Abdomen1
(e) CC5: Chest2 vs. Abdomen2
(f) CC6: Abdomen1 vs. Abdomen2.

The values were then used to measure the similarity of two paired muscular movements from 800-trial (3,200 signal) data by using the following procedures: In the first step, we normalized the signal waveforms by setting the mean values to 0 and the standard deviations to 1. In the second step, we smoothed the detrended signals by using the 10-point moving average filters. Thereafter, we calculated the cross correlation between two signals at lag 0. To measure the similarity of each pair of the RST signals, we conducted three-way analysis of variance (ANOVA; gender: male and female; languages: English and Japanese; texts: NW and, "A Peach Boy").

Results

Figure 1 shows the mean cross correlations between each pair of RST signals. The result of the three-way ANOVA revealed the significant main effects on gender in the pairs of CC1, CC2, CC4, and CC5 ($p < 0.05$), but no significant interactions between gender, language, and texts. The cross correlations of CC1 and CC4, which are adjacent pairs of RST, were higher in the males than in the females. The cross correlations of CC2 and CC5, which were non-adjacent pairs, were slightly higher in the females than in the males. The result of the three-way ANOVA of the cross correlation of CC6 showed no significant main effects but showed significant interactions between gender and language, and among gender, language, and text. Post hoc analyses revealed a significant difference between genders only in the Japanese "A Peach Boy". Figure 2 shows the RST and speech signals when a male and a female subject read NW in English. The peaks of the Chest1 and Chest2 signals co-occurred in the male speaker, whereas Chest2 of the female speaker had noticeable peaks that were not observed in Chest1.

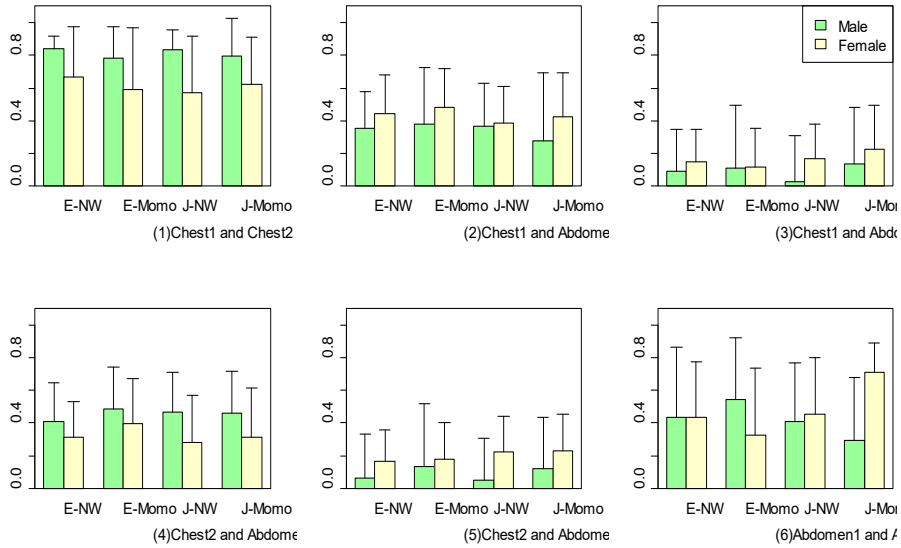


Figure 1. Six kinds of cross correlation between RST signals by the male and female subjects in English (E) and Japanese (J) according to the story (NW and Momo ("A Peach Boy"). The error bars indicate the standard deviations.

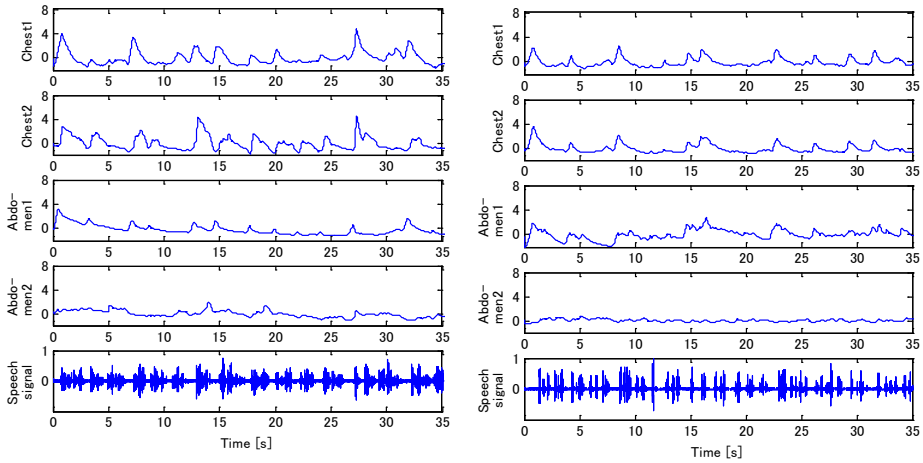


Figure 2. The RST and speech signals when a male subject (left) and a female subject (right) read NW in English.

These results indicate that the upper and lower chest muscles were more coordinated than the upper and lower abdominal muscles for the Japanese speakers, so were those for the male subjects than for the female subjects. On

the other hand, the upper and lower chest and lower abdomen were not coordinated.

Conclusions

We clarified that (1) a gender difference in the controlling respiratory muscles existed as far as the Japanese speakers were concerned under these experimental conditions and that they control respiratory muscles according to language and story, although the significance was not necessarily high. The Japanese males used upper and lower chest muscles and upper abdominal muscles more than did the females during reading. Both the upper and lower chest were highly coordinated but the lower abdomen was not coordinated with the chest and upper abdominal muscles. This implies that the upper and lower chest, and the upper abdomen are biologically inevitably more linked with respiration during speech, but the lower abdominal muscle could be controlled by more arbitrary efforts (cf., Isei-Jaakkola, 2015). (2) In addition, the difference between stories affected only the females. (3) Furthermore, no significant difference in language was observed. This may prove that JEFL used the respiratory muscle movements of JL1 in reading a targeted language to learn. Our hypothesis must be further investigated with other language speakers to reconfirm our present results.

Acknowledgements

This study is partly supported by Grant-in-Aid KAKENHI for Scientific Research C (NO. 17K02698) of Japan.

References

Isei-Jaakkola, T, Ochi, K., Hirose, K. 2018. Respiratory and Respiratory Muscular in JL1's and JL2's Text Reading Utilizing 4-RSTs and a Soft Respiratory Mask with a Two-Way Bulb. Proc. of INTERSPEECH 2018. (Accepted).

Saida, H. 2016. Medical Voice Designer Assists You In Your Vocalization Using Newly Developed Voice Maps (in Japanese). Ongaku No Tomo Sha.

Williams, P. L. 1995. Gray's Anatomy: The Anatomical Basis of Medicine and Surgery, 38e (Gray's Anatomy: the Anatomical Basis of Clinical Practice), Churchhill Livingstone.

Segmental duration in nuclear and post-nuclear syllables in Russian

Tatiana Kachkovskaia, Mayya Nurislamova
Department of Phonetics, Saint Petersburg State University, Russia
https://doi.org/10.36505/ExLing-2018/09/0014/000347

Abstract

In this paper we compared durational patterns of two types of nuclei, HL* (used to signal contrast, in addresses, exclamations etc.) and H*L (used in yes/no questions and non-final phrases), for which the role of post-nucleus is different. In a laboratory experiment with the target word "Natasha" placed in different contexts we found that the two types of nuclei differ in the duration of both stressed and post-stressed syllable. The stressed syllable is longer in HL*, while the post-stressed vowel is significantly longer for H*L, where post-nucleus plays a greater role. However, this only holds true when the target word occurs phrase-finally. If the nucleus occurs phrase-medially, the temporal difference between HL* and H*L is much weaker.

Key words: prosody, segmental duration, Russian, general question, contrastive stress

Introduction

In Russian intonation system, there are two common types of nuclei which differ in the role of post-nuclear part. They are:

> A rise-fall with an early peak, used in addresses, demands or phrases with contrast, logical stress etc. (in terms of ToRI, HL* (Odé 2008)). Most of the pitch curve is within the nucleus, and the post-nuclear part does not play a significant role.
>
> A rise-fall with a late peak, used in yes/no questions or non-final phrases (in terms of ToRI, H*L). Within the system, this type of nuclei differs from a "rise + level high" (H*H), which is used in exclamations, often with admiring or dreaming connotation. The key difference between H*L and H*H is concentrated within the post-nucleus. Hence the large role of post-nucleus for H*L (see Figure 1).

We hypothesized that if the post-nucleus plays different roles in these two contours, they might differ in durational patterns as well – in particular, we will look for differences in the stressed and post-stressed syllables.

We have to bear in mind that when the nucleus occurs phrase-finally, another factor comes into play – pre-boundary lengthening. The experiment should be designed so as to take this factor into account.

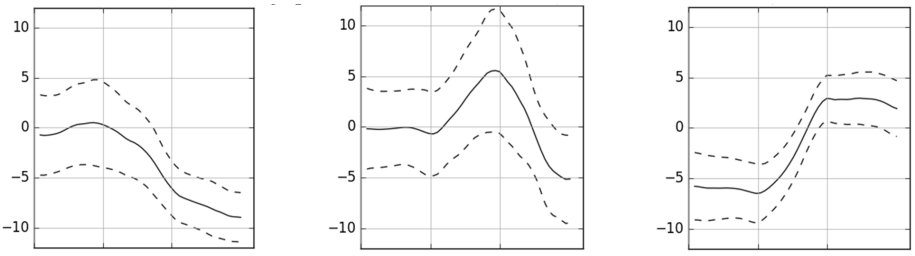


Figure 1. Average pitch curves of the 3-syllable word with penultimate stress in semitones based on CORPRES (Skrelin et al. 2010). Dashed lines show standard deviation. Vertical grey lines mark syllable boundaries.

Experimental method and results

In order to test this hypothesis, we performed a production experiment, where speakers were asked to read a set of utterances containing the target word "Natasha" (/na'taʃa/, a female name with penultimate stress) in five different contexts. In order to elicit the intended melodic contours, the speakers were instructed to put emphasis on the words given in bold.

Neutral context. "Сегодня Наташа на рисовании." (Today Natasha [is] at [the] drawing class.)

Contrastive stress (reassuring) in phrase-*final* position. "Да нет же, приходила Наташа!" (Not at all, came Natasha!)

Contrastive stress (reassuring) in phrase-*medial* position. "Вчера Наташа приходила, а не Марина." (Yesterday Natasha came, not Marina.)

Yes/no question (with a hint of surprise) in phrase-*final* position. "Что? Приходила Наташа?" (What? Came Natasha?)

Yes/no question (with a hint of surprise) in phrase-*medial* position. "Правда? Сегодня Наташа приходила?" (Really? Today Natasha came?)

The material was recorded from six speakers aged 22-50, 3 males and 3 females. All the speakers pronounced the utterances as expected: HL* for contexts 2 and 3, H*L for contexts 4 and 5, no nuclear stress for context 1. The target words were manually segmented into sounds. The utterance-final vowels were judged to end with the end of formant structure.

The duration data for all 6 speakers are shown if Figure 2.

Phrase-final contexts

The stressed syllable, /ta/, was longer in context 2 (contrast) compared with context 4 (yes/no question) for all speakers. The difference ranged between 29 to 46 ms, and most of it was due to the vowel duration.

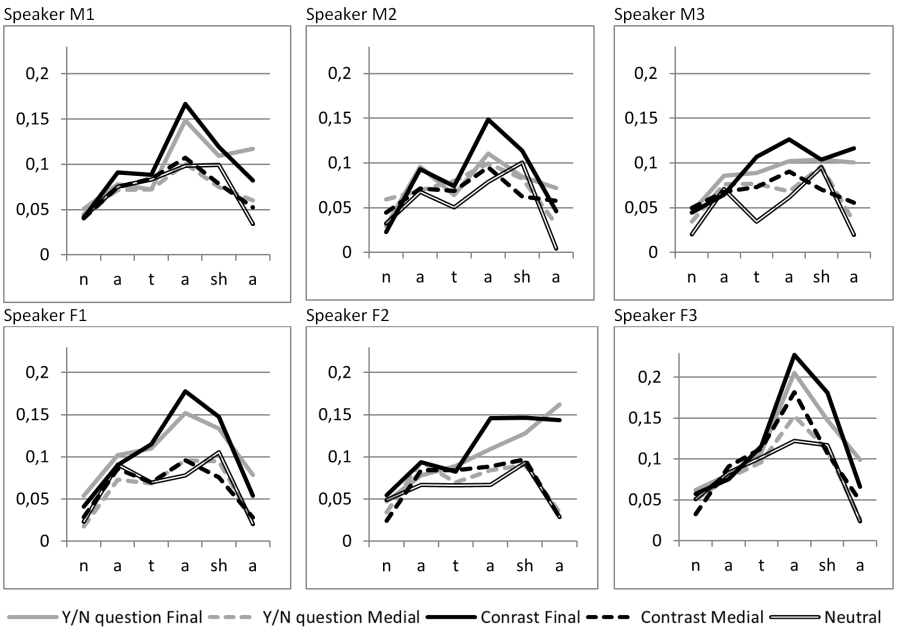


Figure 2. Duration (in ms) of the segments within the target word "Natasha" for 6 speakers and 5 different contexts (see legend)

The post-stressed syllable, /ʃa/, for context 2 was either of the same duration or shorter than for context 4 for all speakers but M3. If shorter, the difference ranged between 10 and 25 ms. Interestingly, the vowel of the post-stressed syllable was significantly shorter for context 2 for these 5 speakers. The difference ranged between 18 and 35 ms. The consonant of the syllable /ʃa/, on the other hand, was in most cases longer for context 2 than for context 4, and thus seems to be compensating for vowel lengthening. However, in our example the consonant is unvoiced, so this might not be the case with sonorants or voiced obstruents.

Thus, for context 4 (yes/no question), where post-nucleus plays a larger role, post-stressed vowel has higher duration. This might be explained by the fact that speakers need more time to show that pitch is going down; otherwise this tune might be confused with another one – "rise + level high" – which in its turn would led to misunderstanding, as the latter contour never marks a question.

The exception, speaker M3, might use a different strategy for temporal marking of these two types of nuclei. More research is needed to find out how frequent this strategy is among Russian speakers.

Phrase-medial contexts

The duration data for phrase-medial show much more variation across speakers. The stressed syllable, /ta/, was longer in context 3 (contrast) compared with context 5 (yes/no question) for 4 speakers; the difference ranged between 18 and 46 ms. Speaker M2 showed an opposite effect, and speaker F1 – almost no difference (2 ms).

For the post-stressed syllable, /ʃa/, the difference between contexts 3 and 5 was in most cases rather small (7 ms and less; 4 speakers). For speaker F1 the syllable /ʃa/ is 16 ms longer in context 5, while for speaker F3 it is 19 ms shorter. As for the final /a/, there is no evidence that it is longer in context 5 than in context 3.

Conclusion

Our data have shown that the two types of nuclei differ in temporal organization, but the latter is highly dependent on whether the nucleus is phrase-final or not. In terms of segmental durations, phrase-final words under nuclear stress are pronounced more carefully, and therefore the melodic difference is supported by difference in duration, while in phrase-medial context the types of nuclei might differ only in terms of melody.

Acknowledgements

The research is supported by the government of Russia (President's Grant # MK-2194.2017.6).

References

Odé, C. 2008. Transcription of Russian Intonation, ToRI, An Interactive Research Tool and Learning Module on the Internet. In Dutch Contributions to the Fourteenth International Congress of Slavists, Ohrid: Linguistics (SSGL 34), 431-449. Amsterdam-New York: Rodopi.

Skrelin, P., Volskaya, N., Kocharov, D., Evgrafova, K., Glotova, O., Evdokimova, V. 2010. CORPRES - Corpus of Russian Professionally Read Speech. In Sojka, P., Horak, A., Kopeček, I., Pala, K. (Eds.) Text, Speech and Dialogue. TSD 2010. LNCS, vol. 6231, 392–399. Berlin, Heidelberg: Springer.

Development of reading and writing skills of heritage Russian speakers in Cyprus

Sviatlana Karpava
University of Central Lancashire, Cyprus
https://doi.org/10.36505/ExLing-2018/09/0015/000348

Abstract

The present study is focused on language proficiency and literacy skills of Russian–Cypriot Greek bilingual children, Russian heritage speakers, children of the first generation immigrants living in Cyprus. Both cross-sectional and longitudinal methodology was implemented to investigate developmental trajectory, dominant language transfer, divergent attainment and attrition of L1 by Russian heritage speakers in Cyprus. Heritage speakers were measured on their reading and writing skills in Russian every month for a period of one year. Longitudinal data consists of the written corpus of dictations and oral corpus of reading aloud recordings. Overall, heritage children were better at reading than writing, comprehension than production. Their spelling and stress assignment errors are due to L1 transfer from Cypriot Greek.
Key words: Russian heritage speakers, reading, writing skills.

Introduction

Heritage speakers are bilinguals in home and dominant languages, they have more family or cultural motivation and connection to the former, minority or immigrant language, and are more proficient in the latter, society language (Polinsky and Kagan, 2007). The present study is focused on language proficiency and literacy skills of Russian–Cypriot Greek bilingual children, Russian heritage speakers, children of the first generation immigrants living in Cyprus. Both cross-sectional and longitudinal research methodology was implemented in order to investigate developmental trajectory, dominant language transfer, divergent attainment and attrition of L1 by Russian heritage speakers in Cyprus. The aim of this study is to examine whether Russian–CG children are balanced bilinguals, whether there is a difference between their perceptive and productive skills in both languages, Russian and CG, and which factors affect the development of their reading and writing skills.

Study

The participants were 39 simultaneous bilingual children (Russian–Cypriot Greek), 17 boys and 22 girls, born in Cyprus (father CG and mother Russian). Their dominant society language is Cypriot Greek, while their home (weak/minority) language is Russian. They have limited exposure to Russian, only at home, and low level of schooling in Russian, only 1-2 hours of Russian lessons per week, see Table 1.

Table 1. Bilingual children: Age, gender and school grade.

N	School grade	Female	Male	Mean age	Range	SD
11	2nd	5	6	8;8	7;7-9;8	0.7
9	3rd	5	4	9;9	8;11-12;5	1.4
8	4th	4	4	10;8	9;4-12;0	0.8
11	5th	8	3	12;12	9;5-14;1	1.3

Bilingual children, heritage speakers of Russian were measured regarding their reading and writing skills in Russian every month for a period of one year. Longitudinal data consists of the written corpus of dictations and oral corpus of reading aloud recordings. The participants were tested on a large battery of tests. Their language proficiency in Greek/CG and Russian were measured with the Developmental Verbal IQ Test (DVIQ), slightly adapted to CG from Stavrakaki and Tsimpli's (2000) SMG original and the Russian Proficiency Test for Multilingual Children (RPTMC) (Gagarina et al., 2010) respectively. Besides the tests, a detailed questionnaire (filled by parents) on language input situation, linguistic and extra-linguistic development of a child was used (Gagarina et al., 2010). Elicited and spontaneous oral production, in Russian and CG, obtained via elicited story-telling while describing eight sets of pictures (Tsimpli et al., 2005) was analysed in terms of speech rate (number of words per minute).

Results and discussion

The analysis of the data showed that bilingual children, heritage speakers of Russian had slightly better overall scores for DVIQ (70%) than for RPTMC (68%), perceptive skills than productive skills, nouns than verbs, see Tables 2-3.

Table 2. DVIQ Results.

DVIQ Greek: measures	Target production %	Mean	SD
Total scores	73.17	110.89	23.81
Lexicon production	62.09	16.91	4.38
Morphosyntax production	59.55	13.95	5.26
Morphosyntax comprehension	78.17	26.26	7.44
Comprehension of metalinguistic concepts	74.11	19.30	4.49
Sentence repetition	82.72	43.70	5.64

The results of the speech rate analysis (number of words per minute), based on oral production revealed that bilingual Russian–CG children had a higher speech rate in CG (*Mean 53.39, range 19-84, SD 15.93*) in comparison to Russian (*Mean 38.3, range 25-65, SD 15.12*).

Table 3. Russian Language Proficiency Test Results.

RLPTMC: measures	Target production %	Mean	SD
Production Lexicon: nouns	64.38%	16.73	5.53
Production Lexicon: verbs	54.68%	14.21	5.02
Production : case	55.79%	3.34	2.20
Perception: grammatical constructions	71.54%	15.73	2.91
Production: verbal inflection	81.88%	9.82	2.16
Perception lexicon: nouns	80.43%	8.04	1.55
Perception lexicon: verbs	76.95%	7.69	1.63

According to Pearson correlation statistical analysis (Sig. 2-tailed), age is correlated with RPTMC *(.007)*; DVIQ *(.007)* and speech rate *(.004)*. School grade — with DVIQ *(.006)*, RPTMC *(.003)* and speech rate *(.012)*. Regarding the analysis of reading skills development of bilingual children, in particular reading speed: words per minute (WPM): a measure of words processed in a minute, it was revealed that there is an overall increase in their reading speed (mean scores) from the 2nd to the 5th grade, see Figure 1.

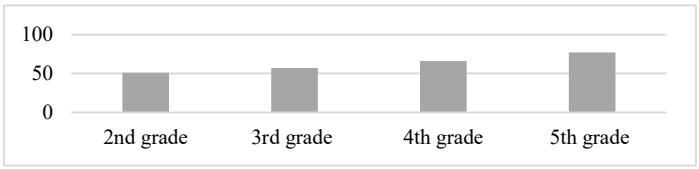


Figure 1. Reading speed: word per minute (WPM).

Stress errors of bilingual Russian–CG children are only within final-penultimate-antepenultimate syllables range (e.g., *но́жи* 'knives' instead of *ножи́*; *о́зеро* 'lake' instead of *о́зера*; *вы́соко* 'high' instead of *высо́ко; гля́дит* 'looks' instead of *гляди́т; пе́ски* 'sands' instead of *пески́*). This could be due to L1 transfer from CG. The analysis of the dictations with respect to orthography and spelling errors showed that there is a developmental pattern: bilingual children produce fewer errors with more input and training, see Figure 2.

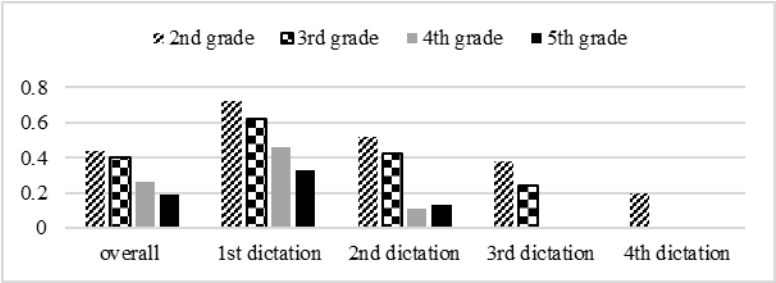


Figure 2. Dictations: Number of errors/number of words ratio.

Bilingual children had spelling errors mainly in the stem (58%) or in the ending (42%) of the word. They had more substitution errors (82%) than omission errors (18%); vowels (63%) than consonants (37%), mainly of phonetic nature (85%) (e.g., *ёш* 'hedgehog' instead of *ёж*; *вдрук* 'suddenly' instead of *вдруг*; *испугалса* 'he got afraid' instead of *испугался*; *листия* 'leaves' instead of *листья*; *решыла* 'she decided' instead of *решила*). According to Pearson correlation statistical analysis (Sig. 2-tailed), reading speed is correlated with overall dictation errors *(.000)*. School grade — with reading speed *(.005)* and overall dictation errors *(.000)*. Age — with reading speed *(.003)* and overall dictation errors *(.000)*. The results of the study showed that bilingual children in Cyprus have higher scores for perceptive skills than productive skills. They show a developmental pattern with age and school exposure for production and receptive skills, reading and writing skills. The gap between comprehension and production in bilingual children can be due to the bilingualism effect. More research is needed to inform the parents and the authorities about the importance of a balanced bilingual development of a child, without forgetting a heritage or a minority language.

References

Gagarina, N., Klassert, A., Topaj, N. 2010. Sprachstandstest Russisch für mehrsprachige Kinder [Russian language proficiency test for multilingual children]. ZAS Papers in Linguistics 54.

Polinsky, M., Kagan, O. 2007. Heritage Languages: In the "wild" and in the classroom. Language and Linguistics Compass 1, 368-395.

Stavrakaki, S., Tsimpli, I. 2000. Diagnostic verbal IQ test for Greek preschool and school age children: Standardization, statistical analysis, psychometric properties. Proceedings of the 8th Symposium of the Panhellenic Association of Logopedists, 95-106. Athens: Ellinika Grammata.

Tsimpli, I., Roussou, A., Fotiadou, G., Dimitrakopoulou, M. 2005. The syntax/morphology Interface: Agree Relations in L1 Slavic/L2 Greek. Proceedings of the 7th International Conference on Greek Linguistics 1-15. University of York.

Manners of rhotic articulation in French lyric singing

Uliana Kochetkova
Department of Phonetics, Saint Petersburg State University, Russia
https://doi.org/10.36505/ExLing-2018/09/0016/000349

Abstract

This study deals with the analysis of the uvular /r/ articulation in singing. Traditionally operatic singers have been avoiding this sound because of its disturbing effect on the vowel production. However some of the modern French lyric singers pronounce this consonant in Art songs, as well as in opera. The aim of the study was to examine the manners of rhotic articulation in singing. Commercial recordings of Art songs of Gabriel Fauré performed by two French singers (a countertenor and a soprano) were analysed. The results of the study showed that both a uvular approximant and a uvular trill were produced, frequently co-occurring with an epenthetic vowel.

Key words: French phonetics, singing, uvular rhotics, consonant clusters, vowel

Introduction

Vocal speech has been considered in numerous studies, its articulatory and acoustic characteristics being of special interest for both singers and scientists. Nevertheless most of contemporary works are focused on vowels, their formant structure, notably on "singer formant" or "singer formant cluster" analysis, as well as on the voice quality and vibrato acoustic features. Consonants have not received a detailed study, whereas consonant articulation may be a challenging task for lyric singers, especially the French /r/ pronunciation.

Singers have been advised to avoid it in formal singing and to produce an alveolar /r/. The uvular /r/ has been considered as interrupting and disturbing the correct airflow in singing since the appearance of Bel Canto style onwards. However such a pronunciation is occurring in vocal speech with increasing frequency nowadays, as it was shown in the previous analysis (Kochetkova 2016). Thus the following question arises: how can contemporary French operatic singers produce a uvular consonant without embarrassing their *portamento*?

The hypothesis is that in the middle of the phrase an approximant may be produced in order to facilitate the articulation and avoid constriction, and at the end of the phrase a so-called "mute e" (French shwa) may appear. The aim of the current study is therefore to observe possible manners of the uvular /r/ articulation in French Art Songs.

ExLing 2018: Proceedings of 9th Tutorial and Research Workshop on Experimental Linguistics, 28-30 August, Paris, France

Material and method

Variants of the uvular /r/ were analysed in the singing of two French operatic singers: one male singer (countertenor) and one female singer (soprano). For this purpose commercial recordings of their interpretations of Gabriel Fauré's Art Songs were chosen: "Clair de lune" (singers C, S), "Automne" (singer C), "Après un rêve" (singer S). Audio files were then analysed and annotated manually using PRAAT software.

Results

In the studied material two variants of the uvular /r/ were observed with almost the same frequency. The uvular approximant [ʁ̞] (Fig. 1) occurred in 47% of cases. The voiced uvular trill [ʀ] (Fig. 2) was produced in 50 % of cases. Other variants occurred very rarely. In one case the unvoiced fricative

[χ] was produced in a coda position. In two cases /r/ was realized as a vowel.

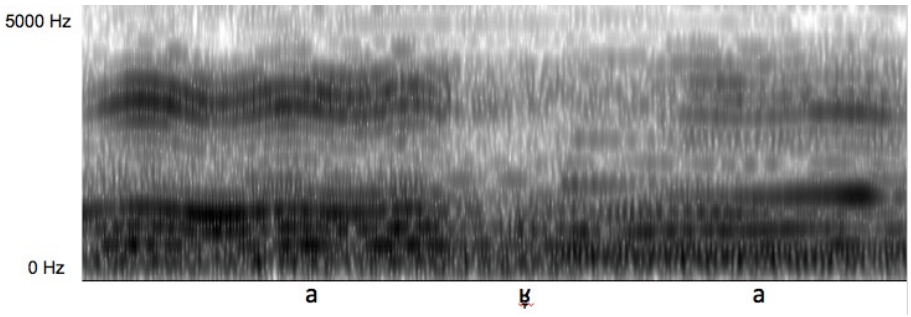


Figure 1. Uvular approximant [ʁ̞] in intervocalic position from ***croire à*** /krw**ara**/, countertenor *("Clair de lune").*

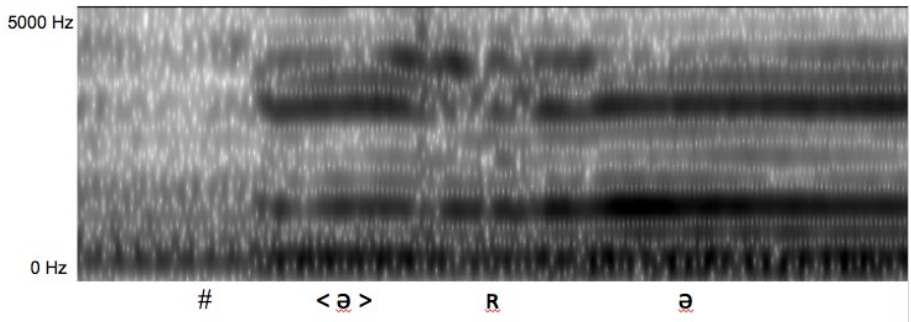


Figure 2. Uvular trill [ʀ] preceded by a vowel (vocoid) from ***# reviens*** /**#ʀə**vjɛ̃/ (initial position), soprano ("*Après un rêve*").

Uvular /r/s in various phonetic contexts

In the texts of the chosen Art songs the following /r/-contexts occurred: VrV, VCrV, VCCrV, VCrSV (between a consonant and a semi-vowel), VrCrV, VrCV, Vr# and #Vr. Hereafter the contexts with /r/s in consonant clusters were grouped into two context types: with /r/ after a consonant (CrV) and /r/ before a consonant (VrC).

As can be seen from Table 1, there was no similarity in the singers' realisations of the uvular consonant in different positions. In intervocalic positions the soprano produced mostly uvular trills (72%), while in the countertenor's singing both uvular trill and uvular approximant occurred with almost equal frequency (see table 1).

Table 1. Occurrence of two variants of the uvular /r/ in different contexts.

Context	Singer	Uvular approximant [ʁ̞]	Uvular trill [ʀ]
VrV	countertenor	53%	47%
	soprano	28%	72%
CrV	countertenor	71%	29%
	soprano	64%	36%
rCV	countertenor	25%	75%
	soprano	55%	45%

It is only after a consonant that most of /r/s were produced as an approximant [ʁ̞] by both singers. In the position where /r/ preceded a consonant, the countertenor preferred the uvular trill [ʀ] (75%), whereas the difference in occurrence of the two variants in soprano's performances was insignificant. In final and initial positions /r/s occurred very rarely. Summarizing both singers' realizations, the approximant was observed in 5 cases out of 7, and the trill appeared in 2 cases out of 4.

Co-occurrence of an epenthetic vowel with the uvular /r/

An epenthetic vowel or, more precisely, a vocoid according to K. Pike's terminology, proved to be a champion in co-occurrence with /r/-segments. This vocoid was observed not only in consonant clusters, but also in the initial position (see Fig. 1). This vocalic segment occurred at a coda as well. In this position it could possibly be treated as a vowel, because it functioned as a "mute e".

In the studied material 90% of /r/s in the soprano's singing and 79% of uvular consonants in the countertenor's performances were followed or preceded by a vocoid (excepting an intervocalic position). The analysis of /r/s in consonant clusters showed that both singers produced more vocoids before a consonant than in cases when /r/ followed the consonant. In the countertenor's performances 95% of VrC contexts and only 58% of CrV contexts contained an epenthesis. In the soprano's singing a vocoid occurred in

91% of pre-consonant positions (VrC) and in 82% of post-consonant positions (CrV).

Conclusion

The results of the current study show that uvular /r/s in singing are produced not only as an approximant, but also as a trill. It was also observed that some contexts might favor the production of one or another variant. Thus CrV context seems to be more favorable for the approximant realization than VrC. The fact that no uvular fricative [ʁ] was observed in the examined material may be explained by the Aerodynamic Voicing principle formulated by J.J.Ohala (1983) and cited in the recent works on French /r/ (Gendrot 2017). According to this principle, constriction would generally lead to the unvoicedness. It seems that singers try to avoid this unvoicedness.

The most salient feature of the uvular /r/ realization in the studied vocal speech is the production of a vocoid preceding or following the uvular consonant (a trill as well as an approximant). It seems to be the most important "technique" that enables singers to pronounce the uvular /r/. Such a realization is especially important in VrC structures. An additional syllable helps singers not only to maintain a correct airflow, but also to produce a uvular /r/ that can supposedly be perceived by listeners more easily, thus achieving a better intelligibility in singing.

References

Gendrot, C. 2017. Perception and production of word-final /r/ in French. Proc. Interspeech, 3926 – 3930, Stokcholm, Sweden.

Ladefoged, P., Maddieson, I. 1996. The Sounds of the World's Languages. Oxford: Blackwell publishers.

Kochetkova, U. 2016. Some aspects of /r/ articulation in French vocal speech. In Botinis, A. (Ed.), Proc. 7th Tutorial and Research Workshop on Experimental Linguistics, 87 – 90, Saint-Petersburg, Russia.

Ohala, J.J. 1983. The origin of sound patterns in vocal tract constraints. In MacNeilage, P. (Ed.) 1983, The Production of Speech, 189 – 216. New York: Springer -Verlag.

Spreafico, L., Vietti, A. (Eds.) 2013. Rhotics. New Data and Perspectives. Bolzano: Bozen-Bolzano University Press.

Acquiring L2 phonemes and recognition of their allophonic variances

Mariko Kondo[1], Takayuki Konishi[2]
[1]SILS & GSICCS, Waseda University, Japan
[2]GSICCS, Waseda University, Japan
https://doi.org/10.36505/ExLing-2018/09/0017/000350

Abstract

Japanese speakers have problems differentiating the English liquid consonants /l/ and /ɾ/, both in production and perception. However, recent studies have shown that Japanese speakers can identify English approximant [ɹ] because it forms a new phonetic category. We trained Japanese speakers with American accent /ɾ/ ([ɹ]) and /l/, and tested them with American and Scottish accented English; Scottish /ɾ/ is often realized as tap [ɾ]. The results showed that Japanese speakers learned to discriminate American /ɾ/ ([ɹ]) and /l/, but not Scottish /ɾ/ ([ɾ]) and /l/. The results imply that the Japanese speakers learned [ɹ], but did not acquire the English phoneme /ɾ/ and its allophones.
Key words: L2 English acquisition, English liquids, acquisition of L2 allophones

Introduction

Japanese speakers have difficulty in differentiating /l/ and /ɾ/ consonants (e.g. Takagi & Mann 1995, Flege at al. 1996, Aoyama & Flege 2011). Variations of both /l/ and /ɾ/ occur in Japanese, but they are not contrastive and so are considered as allophones of /ɾ/, with the most common realization being alveolar tap [ɾ]. However, recent studies (e.g. Flege et al. 1995, Guion et al. 2000, Aoyama et al. 2011, Hattori & Iverson 2009) found that Japanese speakers could discriminate American English and Southern British /ɾ/, because /ɾ/ in these varieties is a post-alveolar approximant [ɹ] and is quite distinct from Japanese consonants, all of which lack lip rounding and tongue retraction. It means that [ɹ] forms what the Speech Learning Model (Flege 1995) calls "a new phonetic category", and while Japanese speakers may not be able to discriminate English /l/ and /ɾ/ as separate phonemes, they can identify [ɹ] as /ɾ/ and other liquids as "not /ɾ/", hence /l/ (Figure 3a).

However, problems arise with some varieties of English, e.g. Scottish English, which has both /l/ and /ɾ/, but the realizations of /ɾ/ vary. Scottish English is rhotic and /ɾ/ is retained in all positions where it occurs and may be realized as trill [r] tap [ɾ] (e.g. Cruttenden 2014).

The issue is that if Japanese speakers have acquired an English phoneme /ɾ/ associated with the most common allophone of approximant [ɹ], can they

extend their knowledge of /r/ to different allophones in different accents, like native English speakers do? Also, can Japanese speakers differentiate tap [ɾ] from /l/ even if their logic of '[ɹ] as /r/' and 'others as /l/' cannot be applied?

In this study, we investigated acquisition of an English approximant [ɹ] by Japanese learners of English and assessed (1) if they can discriminate [ɹ] from /l/, and (2) if they can apply their phonetic knowledge to other allophonic varieties of the sound, by being trained with American English, and then tested with American and Scottish accents.

Experiments

The subjects were 18 (10 male, 8 female) 17-18 year old native Japanese high school students from near Tokyo who had studied English for about 5 years and were taking an English class taught by one of the authors. None of them had lived abroad for more than a few weeks.

The students were trained on the liquid consonants, using audio-visual materials recorded by native American English speakers, and articulatory training by one of the authors. The training lasted for approximately 10 minutes in each lesson, once a week for 5 weeks. The training emphasized lip-rounding and tongue retraction of the [ɹ] articulation, and alveolar and tongue-tip contact with lateral release for the /l/. The students took three tests during the 5 weeks: (i) a pre-test in the 1st week with a male American English speaker different from the training material speakers, and (ii) two post-tests in the 5th week, one with a female American English speaker and the other with a male Scottish English speaker. The three tests were all two-alternative forced-choice identification tests of forty-five minimal pairs of words, contrasting only by /l/ and /r/, i.e. 45 pairs x 2 words = 90 stimuli. The students had to choose the word they heard from pairs of words listed on the answer sheet: word initial (e.g. *lead-read*), consonant clusters (e.g. *fly-fry*), word medial (e.g. *collect-correct*) and word final (e.g. *tool-tour*). The 90 test words were randomly presented once, with a 5 second pause between stimuli.

Results and discussion

The students' English levels differed, so we compared the improvement ratio of their scores between the pre-test and the post-tests. The average pre-test score was 66.73%, and the average scores of the American English and Scottish English post-tests were 77.78 and 67.35, respectively (Figure 1). These data showed that the students' identification scores improved much more for American English than for Scottish English (Fig. 2).

The sensitive score (d') of the forced choice tests was calculated and the results are shown in Figure 3. One-way ANOVA (repeated) of the Pre-test and American post-test and Scottish post-test sensitivity scores showed a significant main effect [$F(2, 34) = 21.01$, $p < .001$]. Pairwise *t*-test with Bonferroni correction showed significant differences between the Pre-test and American

accent Post-test, and between the American and Scottish accent Post-tests ($p<.05$). However, the difference between Pre-test and Scottish accent Post-test was not significant.

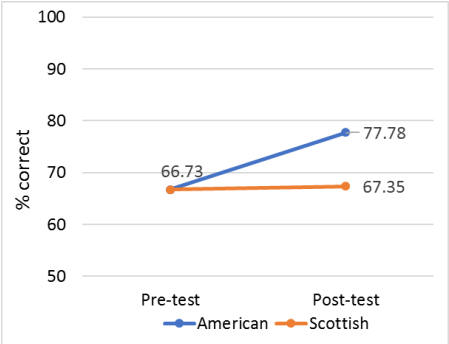


Figure 1. Average correct identification for American and Scottish accents.

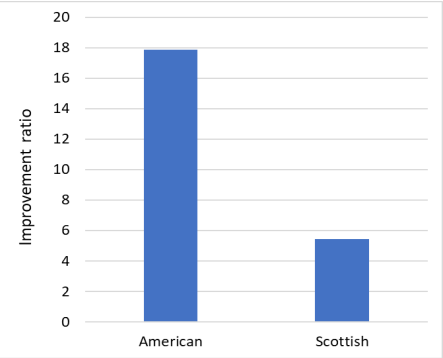


Figure 2. Average % improvement between pre- and post-test.

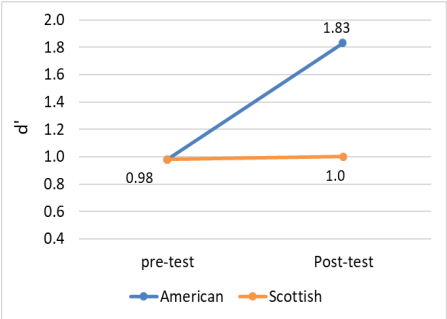


Figure 3. Sensitivity index of identification for American and Scottish accents.

(a) American English

/l/[ɹ]

↓ ↓

Japanese perception

(b) Scottish English

/l/[ɾ]

↓↙

Japanese perception

Figure 4. Perception of English liquids by Japanese speakers. |ɹ| here represent a new phonetic category.

The results showed that the Japanese speakers learned how to discriminate American English /l/ and /ɹ/, but not Scottish /l/ and /ɾ/. This may be because they could recognize the approximant [ɹ] as a new phonetic category and discriminate it from /l/ which was categorized as Japanese /ɾ/, as some previous studies have suggested (Figure 4(a)). Figure 3 shows that the sensitivity

score (d') of the American accent Post-test was much higher than that for the Scottish accent Post-test. This result suggests that the Japanese students relied on the [ɹ] as a cue to discriminate the liquids but not /l/. Our Scottish English speaker used tap [ɾ] much more than [ɹ]. Therefore, in the Scottish English speech, both /l/ and tap [ɾ], which is the typical realization of Japanese /ɾ/, were undifferentiated and were probably categorized as the same /ɾ/ sound by the Japanese speakers (Figure 4(b)).

Conclusions

The results suggest that Japanese speakers can recognize approximant [ɹ] and used it as a cue for discriminating /ɾ/ from /l/. However, they have not acquired [ɹ] as a phoneme /ɾ/, and therefore they cannot recognize its allophone [ɾ] as the sound in the same category.

References

Aoyama, K., Flege, J.E. 2011. Effects of L2 Experience on Perception of English /ɾ/ and /l/ by Native Japanese Speakers. Journ. Phonetic Society of Japan 15:3, 5-13.

Aoyama, K., Flege, J.E., Guion, S.G., Akahane-Yamada, R., Yamada, T. 2011. Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /ɾ/ and English /l/ and /ɾ/" Journal of Phonetics 32, 233-250.

Cruttenden, A. 2014. Gimson's Pronunciation of English (8th ed.). Abingdon: Routledge.

Flege, J.E. 1995. Second language speech learning: Theory, findings, and problems. In Strange, W. (Ed.). Speech perception and linguistic experience: Issues in cross-language research 233–277. Timonium, MD: York Press.

Flege, J.E., Takagi, N., Mann, V. 1995. Japanese adults can learn to produce English /ɾ/ and /l/ accurately. Language and Speech 38, 25-55.

Flege, J.E., Takagi, N., Mann, V. 1996. Lexical familiarity and English-language experience affect Japanese adults' perception of /ɾ/ and /l/. JASA 99, 1161–1173.

Guion, S.G., Flege, J.E., Akahane-Yamada, R., Pruitt, J.C. 2000. An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. JASA 107, 2711–2724.

Hattori, K., Iverson, P. 2009. English /ɾ/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. JASA 125, 469-479.

Takagi, N., Mann, V.A. 1995. The limits of extended naturalistic exposure on the perceptual mastery of English /ɾ/ and /l/ by adult Japanese learners of English. Applied Psycholinguistics 16, 379-405.

Prosodic and pragmatic values of discourse particles in French

Lou Lee[1], Katarina Bartkova[1], Mathilde Dargnat[1], Denis Jouvet[2]
[1]Université de Lorraine, CNRS, ATILF, F-54000 Nancy, France
[2]Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France
https://doi.org/10.36505/ExLing-2018/09/0018/000351

Abstract

This paper analyses prosodic properties of three discourse particles (DP) in French: '*alors*' ('so'), '*bon*' ('well') and '*donc*' ('thus'), according to their different pragmatic functions. DP occurrences of these words extracted from a large speech database have been annotated manually with pragmatic labels. For each DP, prosodic characteristics of occurrences of each pragmatic function (conclusive, introductive, etc.) are automatically extracted. Then, for each DP and each pragmatic function, the most frequent F0 forms are retained as the representative forms. Results show that a pragmatic function, common to several discourse particles, gives rise to a uniform prosodic marking, and this lead to suppose that such DPs are most of the time commutable.

Key words: discourse particles, prosody, pragmatics, computational linguistics

Introduction

Discourse is defined as intrinsically interactional, or dialogic (Bakhtine 1978; Benveniste 1958; a.o.). Spoken language contains discourse particles (DPs), that are cues for discourse or interaction interpretation (Aijmer 2013; Dostie 2004).

We investigate here whether prosodic properties of DPs in French provide information that is related to their various pragmatic values they convey ('introduction', 'conclusion', 'comment', 'emotional state', etc.). Although there are many studies on DPs, very few are dealing with their prosodic correlates.

In this paper, a systematic study is carried out on the prosodic specificities of three French DPs ('*alors*', '*donc*' and '*bon*') with respect to their pragmatic functions. About 1000 occurrences of the three words ('*alors*', '*donc*', '*bon*') have been randomly extracted from several French speech corpora. Each occurrence has been manually annotated, first as DP / non-DP (Bartkova 2016), and with pragmatic labels when DP.

Methodology and data base

The main goal in this study is to evaluate how different pragmatic functions of the three DPs are defined prosodically. Hence, F0 patterns of the pragmatic functions are extracted and their forms are studied jointly with the position of the DP occurrences inside the prosodic groups.

DPs are extracted from more than 100 hours of French speech corpora of various degree of spontaneity, coming from the ESTER2 evaluation campaign (Galliano 2009) and the ORFEO project (ORFEO).

All the data is segmented automatically into phonemes and words, using speech-text forced alignment. The automatic detection of prosodic groups is based on F0 slope values, pitch level and vowel duration. F0 values are normalized according to the speaker's pitch range.

DPs' prosodic articulation

F0 patterns reflect prosodic articulations between DPs and their immediate contexts. The DPs prosodic articulations are studied with respect to their pragmatic functions. For this, the syllable nuclei under consideration are the last syllable of the left context ('*w-1*'), the last syllable nuclei of the DP, and the first syllable of the right context ('*w+1*').

The movements of F0 between '*w-1*' and the DP, and between the DP and '*w+1*', are classified into 3 classes according to the F0 slope directions: falling, rising and plateau. Figure 1 gives examples of F0 patterns.

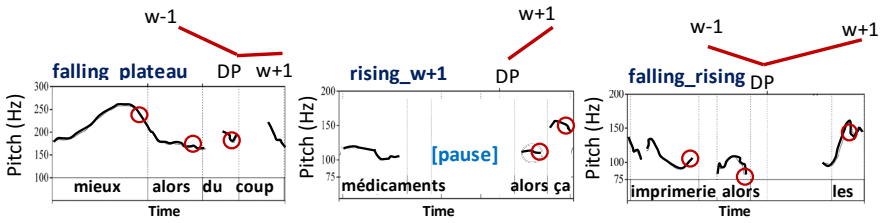


Figure 1: Example of the some frequent F0 patterns.

F0 pattern modelling

Typical F0 patterns per pragmatic function are obtained using a vector quantization procedure. The representative F0 patterns correspond to the centroid of the class. Some F0 patterns are very similar from one function to another; yet, some others reflect prosodic differences among the pragmatic functions. For the DP '*alors*', for example, Figure 2 shows a 'falling-plateau' pattern with lower values for 'conclusion' function than for 'introduction' or 're-introduction' functions.

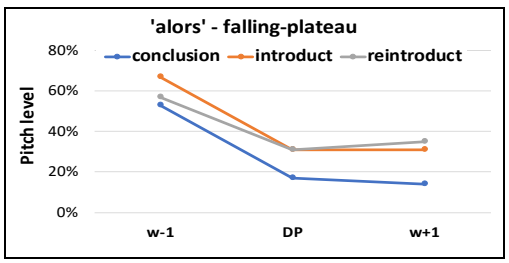


Figure 2: Comparison of stylized F0 patterns for 'alors'.

Representative patterns of pragmatic functions

Table 1 displays the most representative F0 pattern(s) for each pragmatic value of each DP. It is observed that the 'falling' pattern followed by either a 'plateau' or a 'rising' slope, is favoured by 'conclusion', 'confirmation' and 'incident pragmatic functions. 'Conclusion' and 'confirmation' functions express a 'look-back' semantic action of "finality" marked prosodically by a 'falling-rising' pattern highlighting a strong semantic break. 'Incident' represented mainly by a 'falling-plateau' pattern introduces a parenthetical comment. 'Introduction' and 're-introduction' functions have a 'look-ahead' semantic action, prosodically confirmed by a 'rising', and a 'rising-plateau' patterns. In fact, the prosodic patterns of the same pragmatic values are also very similar, and no major differences exist among the 'falling-rising' patterns of the DPs. Therefore, one can suggest that these DPs in this pragmatic function are commutable and their only distinctive mark is not lexical but a prosodic one.

Table 1: Representative F0 patterns.

DP	Pragmatic value	Representative F0 pattern
'alors'	Conclusion	falling-rising & falling-plateau
	Introduction	rising & rising-plateau
	Reintroduction	falling-plateau & plateau
'donc'	Conclusion	falling-plateau & plateau
	Reintroduction	rising-plateau & plateau
	Addition	falling-plateau & plateau
'bon'	Conclusion	falling-rising & falling-plateau
	Interruption	plateau
	Confirmation	falling-rising & plateau
	Incident	falling-plateau

Position in prosodic groups (PG)

Though semantically related to sentences, DPs present a relative syntactic and prosodic independence in the sentence. Most of the time, DPs occur alone in prosodic groups. That is, even in absence of pauses, DPs are prosodically separated from their left and right contexts. 'Alors' is found in single word PGs 82% when 'introduction' and 90% when 'conclusion'. 'Bon' is encountered alone in PGs from 77% to 89% of the cases depending on pragmatic functions. 'Donc' occurs in single word GPs in more than 80% of cases.

Conclusion

This prosodic analysis show that particle pragmatic functions have prosodic specificities and these prosodic marks are related more to the pragmatic function than to the lexical content of the words studied here. An F0 modelling procedure using F0 levels allowed extracting the most prominent F0 patterns

for each pragmatic function. In fact, for a given pragmatic value, the prosodic patterns are very similar. Therefore, it can be supposed that DPs having the same pragmatic function can be interchangeable in the speech chain, and further work is on-going to investigate this hypothesis.

Acknowledgements

This work has been partially supported by the CPER LCHN (Contrat Plan Etat Région "Langues, Connaissances et Humanités Numériques").

References

Bakhtine M. 1978. Esthétique et théorie du roman. Paris: Gallimard.

Benveniste E. 1958. De la subjectivité dans le langage, in Problèmes de linguistique générale, tome 1, 258-266. Paris: Gallimard.

Aijmer K. 2013. Understanding Pragmatic Markers: A Variational Pragmatic Approach. Edinburgh: Edinburgh University Press.

Dostie G. 2004. Pragmaticalisation et marqueurs discursifs. Analyse sémantique et traitement lexicographique. Bruxelles: DeBoeck/Duculot.

Bartkova K., Bastien A., Dargnat M. 2016. How to be a Discourse Particle?, Proceedings of Speech Prosody 2016, 858-863, Barnes, J., Brugos, A., Shattuck-Hufnagel, S., Veilleux, N. (Eds), Boston, USA.

Galliano S., Gravier G., Chaubard L. 2009. The ESTER 2 evaluation campaign for rich transcription of French broadcasts, Proc. Interspeech 2009, 2583-2586, 10th Annual Conf. of the Int. Speech Communication Association, Brighton.

ORFEO project: http://www.projet-orfeo.fr/

A comprehensive word difficulty index for L2 listening

Kourosh Meshgi, Maryam Sadat Mirzaei
Graduate School of Informatics, Kyoto University, Japan
https://doi.org/10.36505/ExLing-2018/09/0019/000352

Abstract

Word difficulty in the listening task is considered challenging because of high subjectivity, high dimensionality, and low generalizability. We propose a word listening difficulty score, as a linear combination of several complementary features. A dataset of expert-annotated partial and synchronized captions for TED talks is prepared for a target language proficiency, in which only the difficult words are shown. A linear SVM was trained on this dataset, and the learned parameters of the SVM were transferred to the proposed score. This data-driven score demonstrates higher accuracy on the annotated dataset and facilitates model and feature expansion.

Key words: Word listening difficulty score, partial and synchronized caption

Introduction

We have developed a tool, partial and synchronized caption (PSC), to foster L2 listening skill by providing an adequate amount of text in the caption to provide assistance to the listeners only when they encounter difficulties. PSC uses automatic speech recognition (ASR) technology to realize word-level text-to-speech alignment and makes a principled selection of words for inclusion in the caption (Fig. 1) based on word frequency, speech rate, and specificity. Our earlier studies using these features revealed that PSC leads to the same level of comprehension as the full-caption, but better prepares learners for listening without using any textual clues (Mirzaei et al., 2017).

To cover various sources of listening difficulties (Bloomfield et al., 2010), it is possible to incorporate more features to the PSC system: lexical features, acoustic/speaker features, content attributes, and perceptual hindering factors. However, these features may be correlated with each other and finding their footprint in listening difficulty is not straightforward. Furthermore, not all the features are equally useful for detecting difficult words in the speech. To gauge the difficulty of a word, readability scores (e.g., Dale-Chall score) consider the frequency of the words in corpus but discard the speaker/acoustic information. Yoon et al. (2016) proposed a listenability score, by fitting a regression model on the average of Likert-scale difficulty scores for each word, given by several non-native speakers with different proficiencies. Their model cannot generalize well on out-of-vocabulary data, and their process was subjective and dependent

on data obtained from non-experts. Kotani & Yoshimi (2017) further explored this idea by using learner proficiency as a feature in the score.

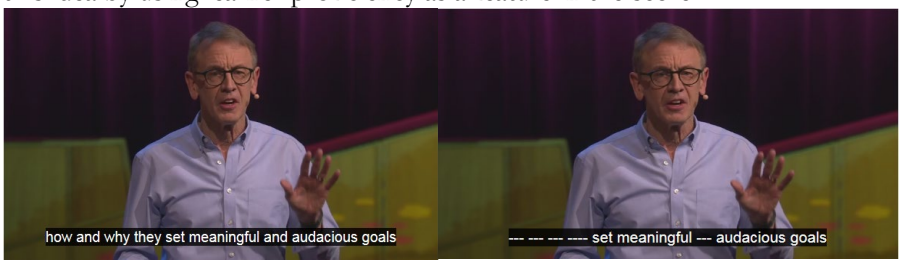


Figure 1. Screen shot for a TED talk with word-level synchronized caption (left) and partial and synchronized caption (right).

This study approaches this problem by extracting various features from speech and find the degree by which they contribute to listening difficulty of each word for a certain proficiency of L2 learners. Using this importance weights, a data-driven word difficulty index is proposed to predict the difficulty of each word in an L2 listening task for a target proficiency.

Methodology

Selecting the difficult words for a target language proficiency can be reformulated as a binary classification problem where, a classifier defined by model θ tries to score word w_i through the use of classifier confidence function $h(w_i|\theta)$: {vocabulary}→[0,1]. This study strives to define such function as a linear combination of given features f_j:

$$h(w_i \mid \theta) = \sum_{j=1}^{n} \lambda_j K\left(f_j(w_{i-\Delta:i+\Delta'}), \theta_j\right) \quad (1)$$

Where $K(.)$ is a kernel function and λ_j denotes the weight of the j-th feature. Such a scoring function may measure word listening difficulty and can be incorporated in PSC. In this view, the words of a speech's transcript are scored, and sorted based on their difficulty. Then the desired number of difficult words are selected from this list to be shown in the caption, which enables adjusting the amount of words flexibly.

Dataset: We prepared a dataset of several TED talks (~90 min including ~10,200 words) given by American native speakers, forced aligned them using Kaldi ASR and Gentle forced aligner (*v0.10.1*), and provided them to two experts to be labeled for intermediate English proficiency. The labels are *show* (for difficult words) and *hide* (for others). Both annotators had linguistic backgrounds and received a set of clear instructions and objectives, that resulted in a high inter-annotation agreement (Cohen's $\kappa = 0.83$).

Features: Figure 2 shows a RadViz chart of most informative features (explained later) extracted from the annotated dataset. RadViz projects a high-

dimensional feature set into a 2D space where the influence of each feature can is shown as a balance between all of the features.

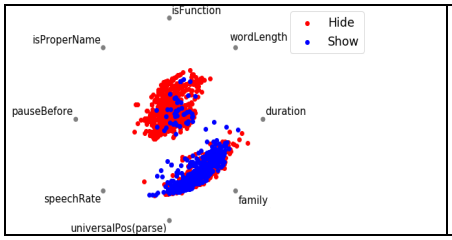


Figure 2. RadViz graph of most informative features extracted from the dataset.

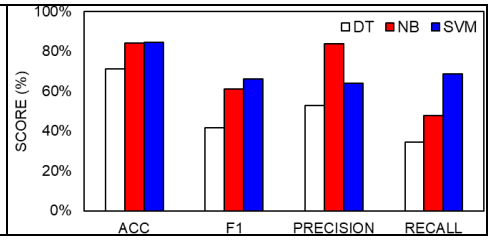


Figure 3. Accuracy of different classifyers, decision tree (DT), naïve Bayes (NB) and linear support vector machine (SVM).

Classifiers: We conducted an experiment to assess the performance of different classifiers on this classification task given the obtained human annotation as ground truth. To this end, a J48 decision tree (*DT*), a naïve Bayes classifier (*NB*), and a linear support vector machine (*SVM*) is trained on the annotated data via a 5-fold cross-validation. Figure 3 shows that while NB and SVM demonstrate comparable accuracy and F1-score. When generating PSC, however, recall is more important than the precision. Precision indicates the portion of shown words that are actually difficult, whereas recall indicates the portion of the difficult words that are shown. Therefore, SVM that shows more of difficult words in the PSC is preferred. Nevertheless, NB helped to find the most informative features (Fig. 2).

Scoring Function: The superior performance of linear SVM in the given classification task, inspired us to create a word listening difficulty score as a linear combination of features where the coefficient of different features is transferred from the slope of the SVM's decision boundary ($\lambda_j^{(\mathrm{SVM})}$, and the magnitude of the feature is calculated as its distance from the decision boundary in that specific dimension ($\theta_j^{(\mathrm{SVM})}$).

$$h(w_i \mid \theta) = \sum_{j=1}^{n} \lambda_j^{(\mathrm{SVM})} \left| f_j(w_{i-\Delta:i+\Delta'}) - \theta_j^{(\mathrm{SVM})} \right| \quad (2)$$

It should be noted that any arbitrary kernels may be considered to calculate this score, as it will be shown in the next section.

Evaluation

We generated partial and synchronized captions for the annotated TED talks using the proposed score function (*Score-L*), an improved version of the score using Gaussian kernel (*Score-K*) and expert rules (*PSC1*) described in Mirzaei et al. (2017) for intermediate L2 learners. The number of shown words in PSCs using proposed scores are kept consistent with PSC1. Figure 4 shows that

proposed scores have better accuracy and F1-score in comparison to PSC1. It also reveals that using Gaussian kernel improves the accuracy of the proposed score in the cost of a few extra parameters.

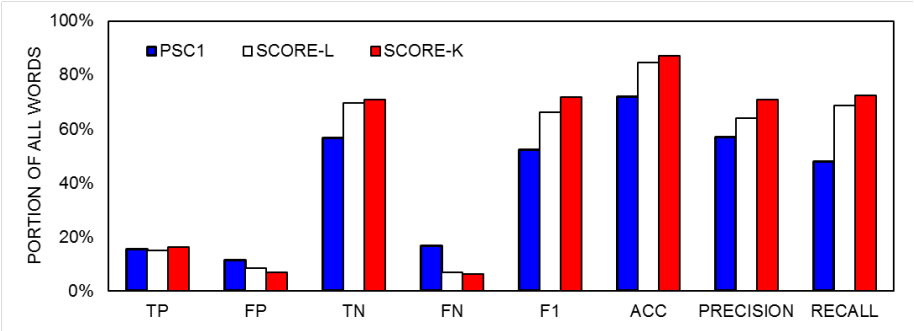


Figure 4. Three different generated PSCs against ground truth.

Conclusion

The proposed index can predict the listening difficulty of L2 learners for a given word, allows learner adaptation (by learning from learner's self-annotations), and automatically generate PSC by observing a few instances of expert annotations for a target L2 proficiency.

Acknowledgments

This research was supported by Japan's MEXT Kakenhi grant #17K02925.

References

Bloomfield, A., Wayland, S.C., Rhoades, E., Blodgett, A., Linck, J. Ross, S. 2010. What makes listening difficult? Maryland University College Park.

Kotani, K., Yoshimi, T. 2017. Effectiveness of Linguistic and Learner Features for Listenability Measurement Using a Decision Tree Classifier. The Journal of Information and Systems in Education 16(1), 7-11.

Mirzaei, M.S., Meshgi, K., Akita, Y., Kawahara, T. 2017. Partial and synchronized captioning. ReCALL 29(2), 178-199.

Yoon, S. Y., Cho, Y., Napolitano, D. 2016. Spoken text difficulty estimation using linguistic features. In Proceedings of 11th NLP-BEA Workshop, 267-276.

The importance of folk-linguistic approaches in the study of dialectal phenomena

Cameron Morin
Department of English, Ecole Normale Supérieure Paris-Saclay, France
https://doi.org/10.36505/ExLing-2018/09/0020/000353

Abstract

As the traditional distinction between the studies of linguistic competence and performance (Chomsky 1965) seems increasingly outdated, so is the separation between theory and practice in disciplines of linguistics equally undesirable, especially in the field of dialectology. As an illustration thereof, this paper aims to present the virtues of the alternative, underutilised approach of *folk linguistics* (Niezielski & Preston 2010), in the shape of a questionnaire for judgment data elicitation about various implications surrounding double modals (DM) in Borders Scots. By detailing the methodology and results of this questionnaire, carried out in January 2018, it will be shown that many dialectal phenomena in English, including multiple modality (MM), absolutely require such sources of evidence to reach a convincing state of analysis.

Key words: folk-linguistics, dialectology, judgment data, questionnaire, modality

Introduction

For a tremendous amount of specific linguistic features, the foremost approach to their study is the search for patterns of occurrence in a corpus, conceived as "a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language" (Sinclair 1996: 4). Recent progress in informatics has led to the primacy of "computer corpora", "encoded in a standardised and homogenous way for open-ended retrieval tasks", and it has resolved the obvious shortcomings of 'toy' systems in theoretical linguistics. Yet it is apparent that a vast range of linguistic phenomena cannot be grasped by way of corpus-based methodologies, for simple reasons including that they may be **i)** as of yet improperly recorded and transcribed if they are rare, **ii)** part of an essentially oral nonstandard language variety, or **iii)** lacking a consensual written form altogether.

This paper aims to show the advantages of an alternative, underrated approach labelled *folk-linguistic*, the core principles of which are the speaker's competence in metalinguistic judgments and the elicitation of said judgments by the researcher in the speech communities concerned. Many folk-linguistic assumptions are usually shared and applied in the various branches of sociolinguistics; the novelty of this paper is to also show how an appropriate usage can help resolve deep theoretical issues raised by features licensed in the

distant periphery of a given language, viz. in dialects. The case study in support of these arguments is that of multiple modality (MM) in Modern Scots, through the example of a questionnaire for judgment data elicitation vis-à-vis this phenomenon.

What is folk linguistics?

Hoenigswald (1966) was the first to formulate the basic principles underlying the experimental school of folk linguistics, namely that we should add to descriptions of "what goes on", i.e language, analyses of "(b) how people react to what goes on and (...) (c) what people say goes on (talk concerning language)" (20). Native speakers of a language variety may indeed have much more linguistic knowledge than we sometimes give them credit for: for instance, they are aware of notions such as reference, morphological segmentability, pragmatic presupposition for anaphora, meanings of linguistic items independent of context and inference of direct speech acts (Niedzielski and Preston 2000: 10-16; Silverstein 1981). The researcher must compensate with their own expertise the disadvantages inherent to judgment data, esp. that they should be subjective and potentially impoverished by inaccuracy or lack of terminology speakers have at their disposal to enrich them. Methods of elicitation are therefore confronted with the challenge of being sufficiently clear and precise to capture the kind of judgment needed for a study's purposes without orienting the speaker's reasoning too much and risking denaturation of the subsequent judgment. This balance, to be respected in every encounter between the researcher and the speaker, is sometimes a delicate affair.

A typical means of folk-linguistic experimentation

It naturally follows that folk linguistics, and more generally sociolinguistics, heavily rely on fieldwork to elicit satisfactory amounts of qualitative judgment data. Data collection is to be done *in vivo,* i.e through interaction with a sample of speakers representing a linguistic community using a set of objective and reliable tools. The most common of these tools is the questionnaire, as the systematic fashion in which it collects data allows for quantitative analysis afterwards (Calvet & Dumont 1999; Johnstone 2000). While *fact questions* are supposed to relate to empirically verifiable phenomena, *psychological questions* relate to opinions, motivations and attitudes. Formally, questions can be closed (yes/no), semi-closed (multiple choice) or open: this determines whether the questionnaire overall is *structured* or *non-structured.* Multiplying the subtypes of collectable judgment is of much importance to give a fuller, pluridimensional approach to the analysis of the linguistic feature at hand. Other methodological guidelines include the systematicity of the questionnaire, i.e that the same form should be given to each and every sub-group of the sample; the brevity of the questionnaire, so that it should not take more than fifteen minutes to complete in order to keep the speaker focused and invested; and the simplicity of the

questions, which implies one idea per sub-task formulated in an easily understandable manner (e.g by avoiding jargon). Finally, questions are required to be neutral, in that they should not withhold any kind of prejudice on the part of the researcher with respect to the research topic concerned and its conclusions.

Folk linguistics and double modals in Modern Scots

Being an essentially oral and basilectal feature also found in other non-standard varieties of American and British English, Scottish DMs (typically *might could,* or *will can*) are impossible to capture by way of contemporary digital corpora (there are next to 0 hits in the corpora available at the Angus McIntosh Centre for Historical Linguistics (Edinburgh)), and previous abstracted accounts of their syntax, semantics and pragmatics have either failed or not been properly undertaken. One solution to this problem is the elicitation of judgment data. A field mission was carried out over three days and nights from the 10th to the 13th January 2018 in the town of Hawick, one of the larger towns in the Borders where MM in Scots has been most extensively studied (Brown 1990; Bour 2014). Its main tasks were the following:

(0) Knowledge of the age, gender, activity and living area of the subject
(1) Knowledge of how the subject represents the geography of their language variety
(2) (2) Recognition and usage of typical DM structures
(3) Syntactic manipulation of DM structures into negatives and questions,
(4) Pragmatic and sociolinguistic information about the current usage of DMs
(5) Recognition and usage of all DMs attested previously in the literature (see Morin 2018 for a full version of the form)

Prior to the field mission, 120 questionnaires had been printed and divided into four stacks of 30 each, three of which were dropped off and regularly checked in central institutions or locations of Hawick while the last stack was personally brought to hosts, proprietors of local shops and businesses during the day, and inhabitants met in public places such as parks, benches on the roadside, cafes, etc. 61 questionnaires were completed and compiled.

Results

The small scale of the experiment is in stark contrast with the wealth of data it provided, relatively to previous research: it showed that *contra* many preconceptions, DMs are still an active component of Borders Scots; it has also led to several hypotheses on a number of (voluntarily ideal-typical) levels:

(a) The syntax of DMs may rely on a context-dependent speaker's choice
(b) DMs may have more assertive strength than standard modal structures, and their semantics map onto those of American DMs
(c) DMs are informal, local and familial features that have adapted to the revolutions of the Internet and social networks
(d) they may partake in the construction of a distinct sociolinguistic identity, for instance through the medium of humour.

Conclusions

The principles of folk linguistics and their experimental methods need to be encouraged in dialectology, sociolinguistics, and as has been shown in the previous section, even beyond. It appears as a highly valuable, even necessary resource in the study of non-written linguistic phenomena; and it may very well be a precious complement to more theoretically-driven disciplines of linguistics, for instance in studies of grammatical core and periphery in which MM needs to be more deeply seated.

References

Bour, A. 2014. Description of Multiple Modality in Contemporary Scotland: Double and Triple Modals in the Scottish Borders. PhD dissertation, University of Freiburg.

Brown, K. 1991. Double modals in Hawick Scots. In Trudgill, P., Chambers, J.K. (Eds.), Dialects of English, 74-103. London: Longman.

Chomsky, N. 1965. Aspects of the Theory of Syntax. Cambridge: MIT Press.

Hoenigswald, H. 1966. A proposal for the study of folk-linguistics. In Bright, W. (Ed.), Sociolinguistics 16–26. The Hague: Mouton.

Johnstone, B. 1999. Qualitative Methods in Sociolinguistics. Oxford University Press.

Morin, C. 2018. Questionnaire for eliciting judgment data on the recognition and usage of double modals in Hawick (Scotland), TulQuest, Paris: CNRS.

Niedzielski, N.A., Preston, D. 2010. Folk Linguistics. Berlin & New York: Mouton de Gruyter.

Silverstein, M. 1981. The limits of awareness. Sociolinguistic Working Papers 84, Austin, Texas: Southwest Educational Development Laboratory.

Sinclair, J. 1996. EAGLES: Preliminary recommendations on Corpus Typology, EAGLES Document EAG TCWG-CTYP/P.

Applying critical discourse analysis in the translation of Maghrebian literature

Hassan Ou-hssata
English Department, Sultan Moulay Slimane University, Morocco
https://doi.org/10.36505/ExLing-2018/09/0021/000354

Abstract

Within the frames of Descriptive Translation Studies (DTS) and Critical Discourse Analysis (CDA), this paper examines the several factors that exert influence upon translating texts both as a process and a product. More precisely, it investigates the notion of ideology with particular use of critical discourse analysis. The purpose is to see the degree to which the translator's socio-cultural and ideological backgrounds have impacts on translations. It also aims to shed light on the potential relationship between language (as a discourse) and ideology in translated texts. This work is a mixed research method study whose corpus is a combination of two literary Maghribi texts along with their translated counterparts. Through a two-level analysis (the macro-level and the micro-level), data analysis aims to find out the dissimilarity between the proportions of the information obtained from the target texts (TTs) and their equivalent at the source text (STs). The results obtained in this research proved that the application of CDA of the STs and TTs helps becoming aware of the genre conventions, social and situational context of the ST and TT, and outlines the formation of power and ideological relations on the text-linguistic level.

Key words: Critical discourse analysis, ideology, translation

Introduction

This paper is based on the integration of Critical Discourse Analysis (CDA) in Translation Studies (TS). CDA has become an independent field within linguistics and it is continuously adapted to new phenomena, one of them being TS. The existing research in the respective field consists of a cluster of different approaches and does not provide an applicable framework that may be used as an auxiliary tool in the translation process for the analysis of source texts (ST) and target texts (TT). Thus, the main aim of this thesis is to create a set of CDA guidelines, combining the CDA framework by Norman Fairclough (1989) with the existing approaches of CDA within TS created by Basil Hatim and Ian Mason (1990; 1997) and Christina Schäffner (1997; 2002; 2003; 2004), as well as to prove that CDA may be a useful tool in the determination of the social and situational context, power relations and ideological struggle during the translation process of political texts.

Translation-oriented source text analysis

The application of CDA in TS is usually performed in two ways: a) the framework may be applied to the analysis of the ST and b) the framework may be applied to the analysis of the TT.

In both cases the aim of the approach remains the same, i.e. to determine what ideological or power relations are reflected in lexical, grammatical and structural elements of the text, how they contribute to the overall rhetorical purpose of the text as well as whether the respective information can be useful to the translator during the translation process. Because of the conference constraints, the focus -as a first leg- is on the analysis of the ST.

Social context

The rule of thumb when CDA is applied for the analysis of texts is the to determine the social and situational contexts within which communication takes place and which determine the roles of the participants. There is no doubt that in order to produce a successful translation, translators must be able to get the intended meaning of the ST producer. Hatim and Mason (1990:224) describe translators as "privilege readers" of the ST, because the translator reads in order to produce and decodes in order to re-encode.

For a translator to be able to interpret the situational and social context within which the communicative event takes place, it is necessary to have at least basic understanding of the topics in the ST. Without the necessary background knowledge, the translator would not be able to uncover the underlying motivations behind the choice of linguistic or grammatical elements and their contribution to the creation of unequal power relationships, if any, between the participants of the communicative event.

The Maghreb literature of French language is a literary production, born under the French colonial period, in the three countries of the Maghreb: Morocco, Algeria and Tunisia. This literature was born mainly around the years 1945-1950 in the Arab Maghreb countries. The authors of this literature are indigenous, that is native to the country. Maghreb literature will become a recognized form of expression after the Second World War. This literature attracts the sympathy of the French-speaking peoples to make them adhere to their cause: To obtain independence and later to resist power structures left after independence.

This information immediately signals a clash between two world views – the Western societies and Third World countries in Northern Africa.

Situational context

The social context determines the situational context, i.e. the relationship between participants, power and distance aspects and language use in the communicative event. The translator must be aware of the type of genre and what constraints the respective genre imposes on the texture and structure of the text in both the ST and the TT. The source texts are considered to be

political due to following factors: the characters and interlocutors represent opposite social groupings of the global society;

In 'Praise of Defeat', there are two characters "A" and "B", both of them forming the same narrator. This can be seen as a medium of two contradictory speeches. The issues raised in the source texts concern framing two distinct yet complementary approaches to Moroccan modern literature, culture and politics.

Indeed, while Abdellatif Laâbi was committed to overt political activism in the Moroccan Left during the Years of Lead, Khatibi embodied the figure of a prolific postcolonial thinker.

the writers are themselves politicians and are interested in politics.

In brief, with regards to the types of power exercised through the choice of textual elements in the source texts, the translators' tasks are to reproduce the unequal power relations by spotting the respective power relations in the ST and then choosing the correct linguistic elements in the TL which would thus create an equivalent effect in the TL.

Text-linguistic analysis of source text

According to CDA, the choice of textual elements reflects the text producer's intentions and linguistic, social and political background. Translators must be aware of the fact that the social and situational context determines the vocabulary, syntax and the overall organization of the text of the political discourse in particular.

Vocabulary and Grammatical Structures

Ideological struggle and power relations may be exercised implicitly through lexis and grammatical structures. The CDA research within TS has been mainly based on the analysis of lexical and grammatical structures and their role in the translation process. Intertextuality and interdiscursivity are closely connected with the analysis of lexis, because language users recognize the meaning on the basis of their background knowledge and experience dealing with other texts and discourses. Thus, the translator must be able to detect the underlying ideological patterns and possible positive or negative connotations of lexical elements in the ST in order to establish equivalence between the ST and the TT by choosing appropriate textual elements on word level as well.

In his foreword to the translation of Laâbi's volume of selected poetry, Pierre Joris aptly notes that the Moroccan poet "writes with a quiet, unassuming elegance that holds and hides the violence any act of creation proposes" (p.iii). This contrast between quietness and violence, elegance and resistance, love and revolt, is one of the most central features of Laâbi's poetry. 'Praise of Defeat' offers an impressive view of how this contrast has evolved throughout Laâbi's career. Starting with his early poems from the period of Souffles, marked by a disrupted and disorientating syntax, the poet is already aware of the social and political impact of his creation as he writes: "now I know what power inhabits me peoples run through my language" (p.5). A space

of collective subversion, awakening and renewal, poetry serves the revolutionary project of "building a kingdom / of insubordination" (p.11).

Conclusion

The ideological struggle in the texts is between two world views:. The writers explicitly and inexplicitly represent and defend the opinion expressed by the Maghrebian culture. This position and the genre of the text allow them to acquire an implicit position of resistance and revolt against any type of power abuse, so that they become what Fairclough (1989) defines as a "gatekeeper" of a cross-cultural encounter. The gatekeeper phenomenon results in the fact that the writers belong to the submissive yet resistant culture of the Maghrebian society.

References

Chilton, P.A., Schäffner, C. 2002 Amsterdam: John Benjamins Publishers.

Fairclough, N. 1989. Language and Power. London: Longman.

Fairclough, N. 1999. Critical Discourse Analysis: The Critical Study of Language. London: Longman.

Hatim, B. Mason, I. 1997. The Translator as Communicator. London and New York: Routledge.

Hatim, B., Mason, I. 1990. The Discourse and the Translator. London: Longman.

Schäffner, C. 2002. The Role of Discourse Analysis for Translation and in Translation Training. Clevedon: Multilingual Matters.

Schäffner, C. 1997. Strategies of translating political texts. In Trosborg, A. (Ed.), Text typology and translation, 19-143. Amsterdam: John Benjamins.

Schäffner, C. 2003. Third Ways and new centres - ideological unity or difference? In Calzada-Pérez, M. (Ed.), Apropos of ideology: Translation studies on ideology - ideologies in translation studies, 23-41. Manchester: St Jerome Publishing.

Criteria for the assessment of visual word processing

Carina Pinto[1], Alina Villalva[1,2]
[1]Linguistic Center of University of Lisbon, Portugal
[2]Department of General and Romance Linguistics, University of Lisbon, Portugal
https://doi.org/10.36505/ExLing-2018/09/0022/000355

Abstract

Our paper aims to verify the role of different criteria, namely, number of syllables; number of morphological constituents; type of morphological structure and word frequency, in the visual word processing. We used a priming paradigm with a lexical decision task. The subjects were exposed to a verb prime (e.g. *doar* 'to donate') for 50 ms, immediately followed by the deverbal derivative in *–ção* (e.g. *doação* 'donation'). The results show that (i) there are no significant differences related to the number of morphological constituents; (ii) there are significant differences between pairs with 5-syllable primes and 2-syllable or 4-syllable primes; (iii) the morphological structure of the verb yields a significant difference between identity and lexicalized pairs; and (iv) frequency triggers significant differences.

Key words: Visual word processing; morphological priming, morphological complexity, frequency of use.

Introduction

Our current research aims to compare different criteria for the assessment of visual word processing (VWP). We are interested in morphological processing and lexical access, but in order to develop any experiment in those fields, we need to ensure that the adopted methodology will produce results that are related to morphological issues and not to phonological or phonetic features, nor to semantic properties or word frequency effects.

A commonly reported result in studies that use the morphological priming paradigm is the facilitation of word recognition when the target is preceded by a morphologically related word. In order to confirm this hypothesis, researchers build word lists that include several control factors that will allow them to claim that the output results are due to morphological, orthographic, or semantic properties, and not to properties such as frequency, word length, age of acquisition, among others.

Balota et al. (2004) report that there are several problems when we try to control the various psycholinguistic variables. The authors state that it is difficult to select words that vary only in one dimension, since these variables are highly correlated. The smallest words tend to be the most frequent, to have a simple morphological structure and they are those that are acquired earlier. Another problem rises from the tendency to choose items that are located in

extreme positions in a particular variable. For example, a given word is typically compared to both a highly frequent and a non-frequent one. Consequently, the behaviour of the target item is distinct at both ends. The authors thus claim that variables such as frequency must be continuous rather than dichotomous.

Many of the visual recognition models, as well as a large majority of the studies associated with them, have been developed on data from English, a language that has a poor morphological system. Our study analyses Portuguese data that has a richer morphological system. A global and systematic approach to the knowledge of the Portuguese morphological system of Portuguese seems to be essential for the observation of morphological processing of Portuguese words. Therefore, we have adopted Villalva (1994, 2000, 2008) as the framework for the definition of the morphological conditions that we want to analyse. Furthermore, the experiment presented below was designed in the framework of the *mixed models* of VWP that argue that the path chosen to access complex words depends on frequency, size and familiarity among other word properties (cf. Baayen at al., 1997; Domínguez et al., 2000).

Method

We have used a priming paradigm with a lexical decision task and we have tested 34 healthy college students (M_{age}= 21,74 ± 5,1). The subjects were exposed to a verb prime (e.g. *doar* 'to donate') for 50 ms, immediately followed by the deverbal derivative in *–ção* (e.g. *doação* 'donation'), available until the lexical decision was made.

We have set four dependent variables:

a) number of syllables (prime): from 2 to 5 (e.g. *do-ar*, *a-cu-sar*, *u-ti-li-zar*, *so-cia-a-li-zar*);
b) number of morphological constituents (prime), disregarding morphological specifiers: from 1 to 3 (e.g. [*do*]*ar*; [[*util*][*iz*]]*ar* 'to use'; [[[*soci*][*al*]][*iz*]]*ar* 'to socialize');
c) type of morphological structure (prime): simple vs complex; compositional vs. complex lexicalized (e.g. *doar*, *utilizar*, *idealizar*). In this case, we have also presented an identity condition (e.g. DOAÇÃO/doação);
d) word frequency (prime and target), considered as a continuous, non-dichotomic, variable.

Priming material contains (i) 38 identity pairs; (ii) 34 pairs formed by prime words that exhibit a different number of syllables; (iii) 38 pairs with a different number of morphological constituents in the prime; and (iv) 33 pairs with different morphological structure of the prime. We have also introduced 77 pairs of WORD/pseudo word as fillers.

Results

We have performed an exclusion of the data that had reaction time values above 2000 ms (3,19%). Preliminary results show an error percentage of 2,5%. Considering the dependent variable number of prime syllables, these errors occur mainly in pairs where the prime was 3 syllables (50%), followed by the pairs with 5 syllables (20,8%), and pairs with 4 syllables (16,7%). The pairs with fewer errors were those that have a 2-syllable prime (12,5%). Regarding the number of morphological constituents, the pairs that yield more wrong replies (68,8%) include a prime that has only one morphological constituent (10,4% of wrong replies with 2 morphological constituents; and 4,2% with 3 constituents). Finally, derivatives primed by a simplex base have a 60,4% of wrong replies, while compositional derived bases have 16,7% and lexicalised derivatives have 6,3%. The identity condition behaves like compositional derived bases.

The remaining results regard exclusively correct replies. The most prominent results considering the reactions times (RTs) are the following:

- There are no significant differences between pairs with different number of morphological constituents ($F = 1,269$, $p = 0,281$);
- Considering the morphological structure of the prime, there are significant differences between conditions ($F = 2,815$, $p = 0,038$), being that difference between identity pairs and lexicalized pairs ($t = 2,415$, $p = 0,016$);
- There are significant differences in the between the pairs with different number of syllables ($F = 5,509$, $p = 0,004$), but that difference it's only between pairs were the prime has 5 syllables and the other pairs (difference from 2 syllables: $t = -2,799$, $p = 0,005$; and from 4 syllables: $t = -2980$; $p = 0,003$);
- Finally, the results show that the frequency of the base produces significant differences in the RTs ($F = 2,814$, $p = 0,000$) and the same for the frequency of the target ($F = 2,730$, $p = 0,000$). These results need further analysis that may consider several stages of frequencies (lower, median and higher, for example).

These results correspond to the first stage of development of the experiment and they will help us to detect methodological issues and to stage some hypothesis about the role of morphology in VWP.

Discussion

We have concluded that (i) there are no significant differences related to the number of morphological constituents; (ii) there are significant differences

between pairs with 5-syllable primes and 2-syllable or 4-syllable primes; (iii) the morphological structure of the verb yields a significant difference between identity and lexicalized pairs; and (iv) frequency triggers significant differences.

These results are still preliminary, but they allow us to hypothesize that the number of morphological constituents *per se* is not a relevant criterion in VWP. Compositionality seems to be more relevant, particularly if lexicalized derivatives are considered. Finally, frequency is a very relevant criterion, but we need to recalibrate our data, taking degrees into account (low, medium and high, for instance). In sum, the word length seems to be irrelevant (unless we select 5-syllable words); the number of morphological constituents is equally unimportant; but the familiarity and the word semantic transparency (or opacity) need to be considered. These results may legitimate further experiments that neglect contrasts between words with 2, 3 and 4 syllables, and 1, 2 or 3 morphological constituents. Conversely, they confirm that the contrasts between familiar and unfamiliar words and lexicalized derived words and compositional derivatives cannot be disregarded.

References

Baayen, H., Dijkstra, T., Schreuder, R. 1997. Singulars and Plurals in Dutch: Evidence for a Parallel Dual- Rout Model. Journal of Memory and Language 37, 94-117. doi: 10.1006/jmla.1997.2509.

Domínguez, A., Cuetos, F., Segui, J. 2000. Morphological processing in word recognition: A review with particular reference to Spanish data. Psicológica 21, 375-401. doi: 10.1.1.16.8654

Balota, D.A., Cortese, M.J., Sergent-Marshall, S.D., Spieler, D.H., Yap, M.J. 2004. Visual word recognition of single-syllable words. Journal of Experimental Psychology: General, 133, 283-316. doi: 10.1037/0096-3445.133.2.283.

Villalva, A. 1994. Estruturas Morfológicas. Unidades e Hierarquias nas Palavras do Português. Universidade de Lisboa: Dissertação de Doutoramento.

Villalva, A. 2000. Estruturas Morfológicas. Unidades e Hierarquias nas Palavras do Português. Lisboa: FCT-FCG. ISBN 9789723108743.

Villalva, A. 2003. Aspectos morfológicos da gramática do Português. In Mateus, M.H., Brito, A.M., Duarte, I., Faria, I.H. (Eds), Gramática do Português. Lisboa: Caminho.

Villalva, A. 2008. Morfologia do Português. Lisboa: Universidade Aberta.

Contrast as bearer of implicit meaning

Lioudmila Savinitch
Institute for information transmission problems RAS, Russia
https://doi.org/10.36505/ExLing-2018/09/0023/000356

Abstract

This paper analyzes contrast, modifier of communicative meaning, its accent structure, intonation contour, sound intensity, and use for conveying implicit meanings. We argue that contrastive highlighting of one of the utterances components in the given example is made by the speaker strategically, in order to convey occasional implicit meaning. All examples are illustrated with graphs displaying tone fluctuations and sound intensity.
Key words: prosody, intonation construction, contrast, implication

Introduction

While examining judicial discourse we focused on the speaker's different prosodic intention in two identical components of the utterance**:** first without contrastive highlighting and then with contrast. The question arose**:** was the accentual highlighting made accidentally or was it strategically realized, that is in order to express a definite communicative intention of the speaker? If strategically realized, then with what aim? Thus, the objectives of the present investigation were as follows: in the examples analyzed to identify the speaker's communicative strategies, semantics of the accents, prosodic characteristics of the accent bearers' word forms, and accent structure of the sentence.

Before proceeding to the analysis of our example, we shall give the definition of communicative strategy in Speech Acts theory:

> The communicative strategy of a speaker consists of the choice of communicative intentions, the distribution of quanta of information on communicative components, and of the choice of the order of communicative components in a sentence" (Yanko 2001: 38).

The communicative strategies of a speaker are implemented in the structures of bearers of communicative meanings and can express intentions to make a statement, ask a question, make a request, give an order, etc. (Austin 1962; Searle 1976: 1–23).

Modifying communicative meanings are those meanings which do not belong to any category of basic illocutionary meanings, such as a statement, a question, a request, or an entreaty, but only modify the main types of illocution and their communicative components. These modifiers include contrast, verifying, or yes/no meaning, and emphasis.

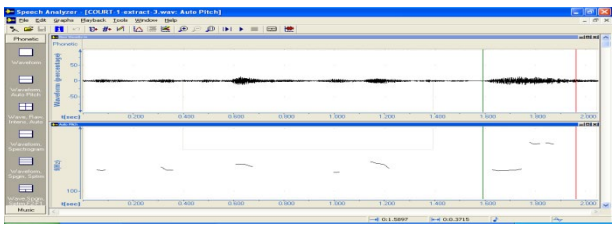
Characteristics of accented word forms

Our example relates to the field of law and was recorded on audio media during a speech by a state prosecutor in court:

(1) **Toktosunov↘ Islambek↗** is accused of applying violence, not dangerous to life and **health↗**, against a government **representative↘↗** in the fulfillment of his official **duties↘**.

In this example the word forms highlighted in bold are accented bearers of communicative meanings; changes in fundamental frequency of the speech are indicated by arrows, which are located after the accented word forms. We will only select all accent bearers in the analyzed example and will specify some of their prosodic characteristics. Explicating the idea of 'intonation pattern' the present article proposes to introduce a concept of 'intonation construction' (Bryzgunova 1980: 96–122) that includes specific aspects of prosody, which will be considered in what follows.

At the beginning of the sentence two components of the proper name *Toktosunov↘ Islambek↗* serve as bearers of accents. The first accent bearer of this group is the word form *Toktosunov↘* with descending intonation of the 2nd Intonation Construction type (IC-2). This is characterized by a more intensive falling of pitch on a stressed syllable than in IC-1 type and a rising pitch on pre-tonic syllables, if any, which provides abrupt falling on a stressed syllable. The second accent is on the word form *Islambek↗* and is pronounced with a rising pitch on the final stressed syllable of the IC-3 type, marking the topic of the utterance. Next, the third accented word form *health↗*, is a member of the attributive group *not dangerous to life and health↗*. It is pronounced with a rising pitch on the stressed syllable and a falling pitch on the post-tonic of the type IC-3 (Figure 1).



not dangerous to life and **health↗**

Figure 1. Intonation contour of the attributive group.

The ascending intonation IC–3 in this case is an ascending accent of incompleteness. That is, it does not carry a local function, nor a function relating to the formation of a separate speech act, but a discursive function. In other words, the rising intonation on this accent bearer does not mark one of

the communicative components, such as topic, as in the word form *Islambek*↗, but provides connectivity of discourse, it indicates that this fragment is not the last piece of text, and is followed by a sequel.

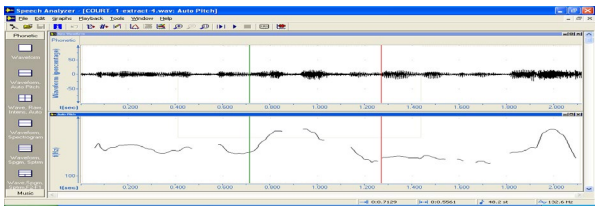
We note that the whole part preceding the accented word form, as is clearly seen in the graph, is spoken practically on one level tone, without sharp frequency fluctuations. The fourth accented word form *government* ***representative***↘↗ is pronounced with a falling tone on the stressed syllable and a rising tone on the post-tonic syllables of the IC-4 type, being an accent of the incompleteness of the text. And the last, the fifth accent bearer in this sentence, the word form ***duties***↘, is pronounced with a smooth falling tone on the stressed syllable and the subsequent post-tonic syllables of the IC-1 type, marking the end of the sentence.

Contrast as a modifier of communicative meaning and implicature of non-obvious sense

While reading the narrative of the committed crime, the state prosecutor repeated again the previously uttered attributive group:

(2) violence, not **dangerous**↘ to life and **health**↗**,**

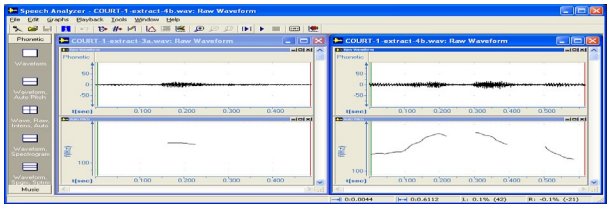
but placed the communicatively relevant accents differently: with a distinct accent on the post-positive adjective *not dangerous*, with a rising tone on its pre-tonic syllables and a falling tone on the stressed and post-stressed syllables of the IC-2 type (Figure 2 between the cursors).



not ***dangerous***↘ *to life and* ***health***↗

Figure 2. Rising and falling tone on the adjective.

In the last variant the new accent modifies the meaning of the communicative focus component and gains a new meaning – contrast. The semantics of contrast is related to a mental procedure of selecting from a variety of options associated with a component chosen by intonation and known to interlocutors (Yanko 2001:47). In our case this set may be limited, for example, to the options: not dangerous vs. dangerous. The prosodic expression of a contrastive focus with the contrast on an adjective, as is shown in the graphics of the right bottom panel of Figure 3, differs significantly from the non-contrastive version of the left bottom panel.



not dangerous not **dangerous↘**

Figure 3. Intonation contours of the non-contrastive and contrastive versions.

Measurements of intensity (an acoustic correlate of the volume) of both variants confirm highlighting of one of the utterances components. Thus, the intensity on the non-contrastive component is equal to -19,4 decibels (dB), on the contrastive one is -17 dB, so the contrastive option sounds louder.

Conclusion

Thus, in the second example, physical measurements confirm the appearance of a new accent. Thereby there is the strategically intelligent singling out of the communicatively significant component of the sentence with the purpose, as stated above, of indicating by means of intonation the choice of that component. This choice has been made from the set of the possible variants associated with the selected component. (In our case it is two opposed variants: not dangerous vs. dangerous.) Meanwhile, in the use of this communicative strategy there is also another, probably more essential aspect: the prosecutor implicitly assesses, implicitly evaluates the committed crime according to the current legislation. The last hypothesis primarily may be confirmed by comparison with the current Penal Code, from which the above mentioned attributive construction is quoted and according to which the law provides for appropriate sanctions.

References

Austin, J. 1962. How to Do Things with Words. Oxford: Oxford University Press.

Bryzgunova, E. 1980. Intonation. Russian grammar 1, 96–122.

Searle, J. 1976. A classification of illocutionary acts. Language in Society 5, 1–23.

Yanko, T. 2001. Russian Speech Communicative Strategies. Moscow: Slavic cultures languages.

A corpus-based study of metadiscourse markers in English and Urdu

Haroon Shafique
University of Lahore, Gujrat Pakistan
https://doi.org/10.36505/ExLing-2018/09/0024/000357

Abstract

Interactional metadiscourse markers are the self-reflective linguistic expressions that make the writers more powerful in interaction (Hyland, 2004). In this study, a large corpus has been compiled from English and Urdu newspapers. The compiled corpus is analyzed quantitatively as well as qualitatively to draw the results by using Hyland's (2005) model of interaction. The quantitative results exhibit that the news writers of English and well as Urdu prefer to manipulate the viewpoint of their readers by their judgments when they use stance markers. The contrastive analysis of interaction markers in both corpora reveals that Urdu journalistic discourse is more persuasive, convincing and influential as it consists more interaction markers as compared to English journalistic discourse.

Keywords: Metadiscourse, interactional markers, corpus

Introduction

Interactional metadiscourse has already been highlighted by so many linguists in the past and highlighted interactional metadiscourse in different genres as well as in different languages but there is hardly any research found on interactional metadisocurse in Urdu. Moreover, no contrastive study has yet been done to analyze interactional markers in English and Urdu in journalistic discourse.

The term metadiscourse is widely used in current discourse analysis and language education, referring to an interesting and relatively new approach to conceptualize interactions between text producers and their texts and between producers and users (Hyland, 2010).

Metadiscourse is specifically defined as "the linguistic resources used to organize the author's stance towards either its content or to the reader" (Hyland, 2000; p.109). On the other hand, metadiscourse is more generally seen as writer's linguistic and rhetoric manifestation in a text so as to bracket the discourse organization and the expressive implications of what is being said (Schiffrin 1980).

Research questions

1. How differently interaction markers are used in English and Urdu journalistic discourse?
2. What kind of interaction markers are preferred in English and Urdu journalistic discourse; stance markers or engagement markers?

Methodology

This is a corpus-based study where mixed method approach is applied to find the results. The data is first quantified and then analyzed qualitatively. Hyland's (2005) model of interaction is applied to English and Urdu corpus to find out interaction markers in journalistic discourse.

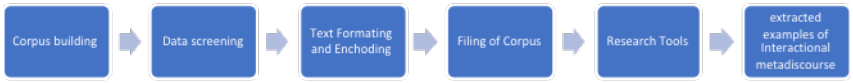


Figure 4. The process of corpus building and analysis

Theoretical framework

The theoretical framework applied in this research is Hyland's (2005) model of interaction which analyzes the interactional features of discourse. Stance is called the textual voice which refers to the ways in which a writer projects himself in a text and conveys his judgments, opinions and commitments. On the other hand, engagement is the reader-oriented approach in which a writer recognizes the presence of others, pulls them along with his arguments (Hyland, 2005).

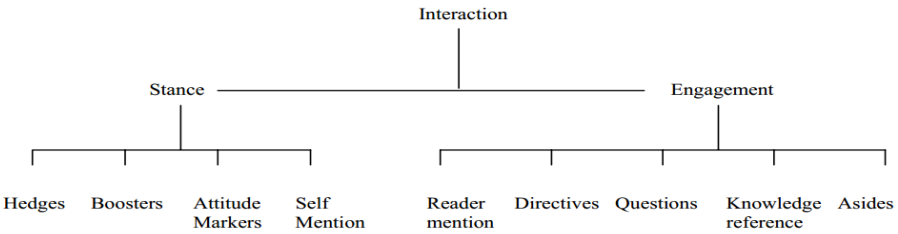


Figure 5. Hyland's model of Interaction (2005).

Data analysis

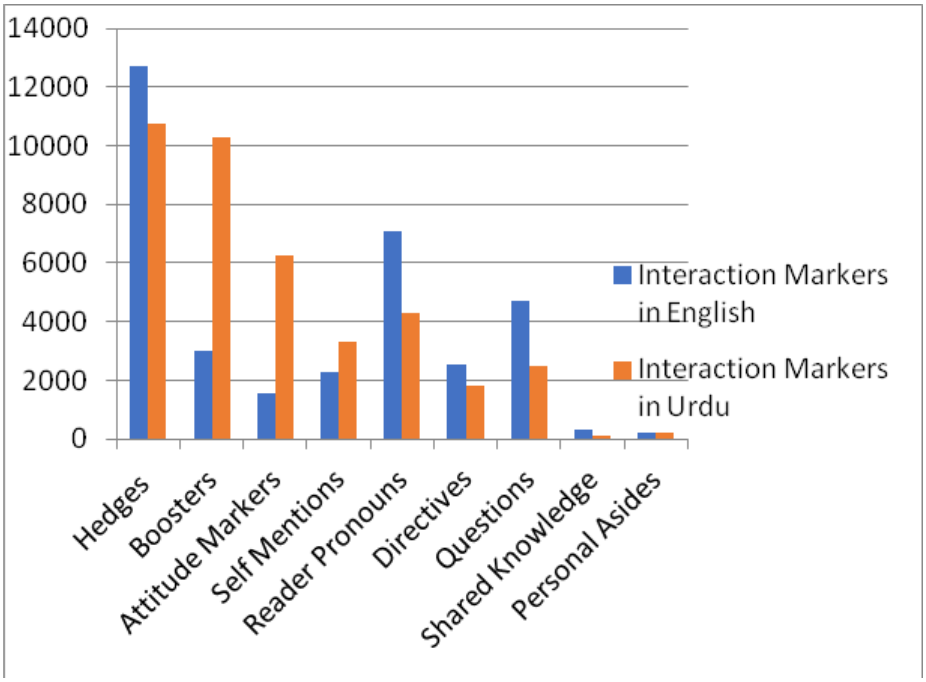


Figure 6. Interaction markers in English and Urdu

The graph above given exhibits the contrastive analysis of interaction markers i.e. stance markers and engagement markers in English and Urdu journalistic discourse. The graph shows, hedges are the only stance markers with high frequency in English discourse as compared to Urdu. The other three stance markers have high frequency in Urdu journalistic discourse than English. It implies that the English Journalistic writers prefer to use hedges more to project their stance while the rest of three stance devices are given less priority in English newspaper writing. As far as engagement markers are concerned, all the engagement markers have higher frequency in English newspaper writing than Urdu. It suggests that the journalistic writers of English prefer to use more engagement markers as compared to Urdu journalistic writers.

Conclusion

The findings of the study reveal many important facts. The quantitative analysis of the study shows that hedges are the most occurring Interaction markers in English and Urdu corpus which depicts that the uncertain statements are the core feature of Journalistic Discourse. The results exhibit that stance is the dominating category of metadiscourse markers in both journalistic discourses. The quantification of the corpus also highlights that the news writers of the

Urdu language use more interaction markers as compared to the English news writers which implies that Urdu journalistic text is more persuasive, influential and convincing for the readers.

References

Hyland, K. 2000. Disciplinary discourse: Social interactions in academic writing. London: Longman.

Hyland, K. 2004. A convincing argument: Corpus analysis and academic persuasion. In Connor, U., Upton, T.A. (Eds.), Discourse in the professions: Perspectives from corpus linguistics 87-112. Amsterdam: John Benjamins.

Hyland, K 2005. Metadiscourse: Mapping Interactions in Academic Writing. University of London, UK

Schiffrin, D. 1980. Meta-talk: Organizational and evaluative brackets in discourse. Sociological Inquiry 50(3-4), 199-236.

Exploring the potential of visual shadowing as an L2 listening pedagogy at universities in Japan

Fuyu Shimomura
Department of English Studies, Kyoto Women's University, Japan
https://doi.org/10.36505/ExLing-2018/09/0025/000358

Abstract

Since MEXT announced the "Action plan for educating global citizens with high English competence" in 2003, universities in Japan are expected to improve student English communication competence. However, many Japanese students tend to struggle with listening to English sentences. Given this inclination, some scholars pointed out the effectiveness of repeating, visual shadowing and shadowing practices as a SLA listening pedagogy. This paper explores how visual shadowing activity helps students listen to English better with using student TOEIC listening scores, and interview data with case study students. (85 words)

Key words: shadowing, listening pedagogy, EFL, university, Japan

Background

MEXT (Ministry of Education, Culture, Sports, Science, and Technology) announced two English education policy changes in 1998 and 2003, regarding what higher educational institutions should be responsible for teaching to college students. The "Action Plan for educating global citizens with high English competence" announced in 2003 specifies that universities should develop student English skills to the level where they feel comfortable communicating (MEXT, 2003). In addition, MEXT (2003) also states that they should use English proficiency test such as TOEIC, TOEFL and IELTS as valid measures of student English competence.

However, Richards (1983) points out that Japanese EFL learners tend to be challenged with listening, particularly at the perception stage (bottom-up processing). Osuka (2008) further clarifies that particularly their inabilities to perceive 1) fast speeches and 2) particular English sounds or phonemes are the two major impairing factors for Japanese learners. In other words, if Japanese EFL learners have access to the pedagogies that help them overcome these two impairing factors, their listening perception skills should significantly improve. To achieve this end, shadowing has potentials to help learners increase their articulation rates and contributes to developing better L2 listening abilities by helping working memory more effectively hold and process the auditory information entering the phonological loop (Hamada, 2017; Kadota, 2012 & 2015; Tamai, 2002 & 2005). Taken together, it is possible to claim that shadowing has potentials to improve learners' phoneme perception skills as

well as the articulation/subvocal rehearsal rate, which determines the speed and amount of information learners could process within the time limitation of working memory. Therefore, this research explores how differences in articulation rates influence learners' L2 lisning skills improvement.

Research methods

Purpose. Given the claim that shadowing and similar outputting activities including visual shadowing are helpful in terms of improving listening skills in L2, and that inability to understand fast speech and particular English sounds or phonemes are the two major impairing factors for Japanese EFL students, this paper aims to explore how difference in articulation rate for fast-paced visual shadowing influences the improvement of student listening skills in English. An independent variable for this research inquiry is differences in articulation rates (Group A: slower, Group B: faster) for shadowing activities with texts (=visual shadowing), and two dependent variables are differences in: 1) TOEIC listening score increase, and 2) the increases in WPM (=articulation rate) in visual shadowing TOEIC listening scripts between pre- and post-tests.

Participants. The participants were 65 non-English major freshmen enrolled in a private women's college, taking mandatory TOEIC preparation classes with the author. Both classes include repeating and visual shadowing (shadowing with texts) activities for the same time duration and same frequency.

Materials for visual shadowing. As the classes the author conducted researches in were mainly designed to prepare students for TOEIC tests, both slower and faster visual shadowing groups used TOEIC preparation materials as textbooks. Group A (slower visual shadowing group) used an unofficial TOEIC preparation material, while Group B (faster visual shadowing group) used one of the official TOEIC preparation materials published by ETS. Group A's textbook covered only the conversation and narration parts (Parts 3 and 4), while Group B's listening textbook covered all listening parts of the TOEIC exam (Part 1 through Part 4).

Procedures. First, students took TOEIC IP tests (pre-test). To examine if the articulation rates of visual shadowing influenced the development of listening skills or not, students were divided into two groups (Groups A & B) based on their majors. Both groups had repeating practices at the same articulation rate and fast-paced visual shadowing at different articulation rates (Group A: slower, Group B: faster). After working on these activities for ten consecutive weekly class sessions, students took TOEIC IP test (post-test) again. The author also conducted pre- and post- interviews with randomly chosen six case study students from each group to explore: 1) if they perceived any changes in the way they perceive or understand English speech after working on rapid visual shadowing for the consecutive 10 weeks, and 2) if there were any differences observed in the articulation rates between pre- and post-interviews when they read aloud the exact same TOEIC listening scripts.

Results

An independent variable for this research inquiry is difference in articulation rate (Group A: slower, Group B: faster) for visual shadowing, and two dependent variables are differences in: 1) the increase in TOEIC listening score, and 2) the increase in WPM (=articulation rate) in visual shadowing TOEIC listening scripts between pre- and post-tests. Comparing listening score increases in Groups A and B highlighted that there are minor differences in the score increase (approximately by five points) between Groups A and B. (See Table 1). Although, Group B - the faster articulation group - showed a slightly bigger increase in mean score than Group A, this difference (5 points) in mean score increases should not be considered significant.

Table 1: Improvement of listening scores after shadowing for 10 weeks.

	Group A (slower)	Group B (faster)
Pre-test listening score	232.79	238.5
Mean L score increase (dependent variable)	39.3	44.1
Standard Deviation (of L score increases)	43.5	51.1
L-score increase range	-55 to 95	-60 to 165

Cross analysis of interview data sets with listening test score also highlighted the following two points: 1) those who worked hard to increase their articulation rates tend to increase their listening scores much bigger than those who worked hard to master the correct prosody, and 2) shadowing is also helpful for higher proficient learners. These research results indicate that when working on shadowing activities for improving listening skills, learners need to focus more on their articulation rates than their prosody, and also shadowing has a potential to help higher proficient learners improve their L2 listening skills as well.

Conclusion

Given these research findings, it became clear that 1) visual shadowing is helpful for learners of both higher and of lower proficiency - the difference to note is that lower proficient learners were less likely and higher proficient learners were more likely to feel their improvement in listening, and 2) students' fast articulation rates matter more to their listening skill improvement than mastering correct prosody, including sound changes or reductions in fast speeches. For providing more effective EFL listening pedagogies, exploring: 1) how much students should care about prosody while visual shadowing to improve their listening comprehension, 2) how fast students should be able to articulate English words or sentences to feel comfortable in listening to the English speech at the natural speed, and 3) whether TOEIC should be

considered as a valid measure for listening comprehension, are the three major further avenues for investigation.

References

Hamada, Y. 2017. Teaching EFL learners shadowing for listening: Developing learners' bottom-up skills. New York: Routledge.

Kadota, S. 2012. Shadoingu to ondoku to eigoshutoku no kagaku [Science of shadowing, oral reading, and English acquisition]. Tokyo: Cosmopier Publishing Company.

Kadota, S. 2015. Shadowing, ondoku to eigo communication no kagaku. [Shadowing, repeating, and the mechanism of English Communication]. Tokyo: Cosmo Pier.

MEXT. 2003. Eigo ga tsukaeru Nihon-jin ikusei no tame no koudou keikaku [Action plan for educating global citizens with high English competence]. http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/004/siryo/04031601/005.pdf

Osuka, N. 2008. What factors affect Japanese EFL learners' listening comprehension? In K. Bradford Watts, T. Muller, M. Swanson (Eds.), JALT2007 Conference Proceedings. Tokyo: JALT.

Richards, J.C. 1983. Listening comprehension: Approach, design, procedure. TESOL Quaretrly, 17(2), 219-240.

Tamai, K. 2002. Listening ryoku kojo ni okeru shadowing no koka nit tsuite [On the effects of shadowing on listening comprehension]. Keynote lecture at the 3rd Annual Conference of JAIS. Interpretation Studies 2, 178-192.

Tamai, K. 2005. Listening shidoho to shite no shadowing no koka ni kansuru kenkyu [Research on the effect of shadowing as a listening instruction method]. Japan: Kazama.

Focal vs. global ways of motion event processing and the role of language: Evidence from categorization tasks and eye tracking

Efstathia Soroli
Department of Linguistics, University of Lille, France
https://doi.org/10.36505/ExLing-2018/09/0026/000359

Abstract

Past research has shown that cultural experience and language specificities affect how people process spatial information cognitively. Holistic cognition is generally associated with East-Asian cultures while analytic processing to the Western ones. Similarly, the languages of the world vary greatly: some (verb-framed) invite to encode mainly general, core spatial components (Path) avoiding encodings related to the process of motion (Manner), while others (satellite-framed) focus on Manner systematically adding information about Path. The current study asked whether speakers who share the same cultural background (Western) but speak different languages (English (satellite-framed), French (verb-framed) and Greek (parallel)) show differences in categorization and visual processing of motion events. The findings show that holistic processing is not exclusive to Eastern cultures: speakers of indo-european are also influenced by the degree of focality of their language.

Key words: holistic vs. analytic cognition, focal vs. global strategies, spatial language, categorization, eye movements.

Introduction

There is a widespread idea that the human cognitive system functions the same, at least for all normal humans. This assumption of universality was recently strengthened by genetic theories that support the idea of a common genetic basis of the cognitive system (e.g., Kovas & Plomin, 2006) and by the Minimalist Program (Chomsky, 2014) according to which humans are all equipped with the same set of general conceptual categories that allows for processing of core features irrespective of linguistic or cultural background.

A growing number of studies, however, show that experience and more specifically experience with language may be one of the formative or even transformative aspects of human cognition. Work by Nisbett et al. (2001) and Soroli et al. (2018) reveal pervasive effects of culture, learning and language on reasoning and other non-verbal cognitive mechanisms such as categorization and attentional processing (e.g., Pannasch & Velichkovsky, 2009). More specifically, from a cross-cultural perspective it has been argued that social and cultural differences strongly affect how people process their environment and

more specifically events. Nisbett et al. (2001), as well as Ji and Hohestein (2017) suggest two ways of event processing: one holistic, experience-based way of thinking associated with the East Asian (i.e. Chinese-speaking) cultures, and one analytic, object-based way of thinking, associated with the Western (i.e. English-speaking) societies.

From a cross-linguistic perspective, similar distinctions have been formulated by Tai (2003): while English-speaking participants have been found more interested in the sequences of a motion event, focusing on processes, agents, objects and action-based constructions, Chinese speakers are more field-focused and their encoding patterns tend to be result-based. This action-based vs. result-based distinction is inspired by the typological distinctions Talmy (2000) formulated in his seminal work on cognitive semantics. According to this work, which goes beyond the East-West distinction, the languages of the world present great differences in the way spatial information is mapped onto lexical and syntactic structures: the so-called *Verb-framed* languages (e.g. French), allow mostly for lexicalization of the Path component, information about *where* the event is performed or what is the general result of the displacement without entering into the details often avoiding any specific focal reference to Manner information; the *Satellite-framed languages,* (e.g. English), mostly allow Path to be expressed in constituents that stand in a sister position to the main verb, lexicalizing Manner of motion, thus focusing on the process and on *how* motion is actually performed. Additionally, it has been argued that some languages, such as French, allow also some *hybrid* spatial encodings, with fused Path+Manner verbs, while others, such as Spanish, Chinese, or Greek may present *mixed*, *equipollently-framed or parallel (verb-* and *satellite-framed)* systems of conflation, probably depending on the *degree of focality* (salience of focal details of the event) within each spatial configuration.

Such differences across and within systems led researchers to address some fundamental questions about the relationship between language/culture and the cognitive mechanisms underlying event representation. Are cognitive processes (e.g., categorization, visual attention etc.) part of an innate independent system or they rather have connections and interact with other systems of processing such as language and culture? And how people, who share the same cultural background (e.g. French, English and Greeks), perceive, categorize and encode motion events? If language only affects language-related experiences, then it is rather unlikely that the building blocks for event representation are language-specific. If the cultural experience (western in this case) affects event processing, then no difference should be found across the three language groups. On the other hand, if language impacts language-related behavior but also the representational system, then it is unlikely that solely universal mental categories play a role in event representation. In this case it is expected to find differences even among groups that share the same cultural background with respect to the degrees of focality and sensitivity to details of the displacement.

Method

60 speakers of three typologically different languages, (English, French, Greek), 20 per language, were tested. They performed three controlled tasks involving motion events: (1) a non-verbal categorization task; (2) a verbal categorization task; and (3) a production task, all coupled with an eye-tracking paradigm for further insights on on-line cognitive processing. In experiment 1, participants saw a target-video presenting a motion event performed in a certain Manner and along a certain Path (e.g. *A woman riding a bicycle into a building*). The target was then followed by two video variants: one Manner-congruent (*A woman riding a bicycle out of a building*) and one Path-congruent (*A woman riding a scooter into a building*). Participants had to choose the variant that looked most like the target as fast as they could by pressing a key. Experiment 2 was exactly the same, except that the target video was replaced by a target sentence. In experiment 3 participants had to describe verbally the events.

The analysis was focused on how participants performed categorization, according to which criterion (Manner or Path) (experiments 1 and 2), what participants expressed, with which linguistic means and in relation to which specific events (experiment 3), and which specific areas of interest (AOI) they were looking at, how many times (numbers of fixations), for how long (visit-durations) and following which visual trajectory (gazepaths).

Results

The results show that all groups followed the typological patterns of their native language: French participants preferred to lexicalize Path in the utterances leaving Manner either unexpressed or in the periphery; English participants systematically encoded Manner within the main verb and Path in the periphery (with particles or prepositional phrases); while Greek-speaking participants alternated their verb- and satellite-framed constructions (i.e. by using lexicalized Path as well as many peripheral devices, preverb configurations, complex Manner-first patterns etc.). French participants were less focal in their non-verbal behaviour than English participants. They made more Path-choices in the categorization tasks, attended more and longer to Path components combining this preference with ballistic (from-source-to-goal) global ways of exploration of the events, as opposed to English who paid less attention to Path and followed a rather focal (linear/step-by-step) strategy for visual processing. Greek participants, depending on the context and the saliency of the components, alternated their visual strategies, showing however that when verbal input is not explicit, overt attention to specific components may differ in fixation counts but not in visit-durations.

Conclusion

Participants were largely influenced by the typological properties of their native language, not only when performing verbal descriptions but also when making their non-verbal decisions. Despite the fact that they shared the same cultural background (Western), participants categorized and shifted attention mostly based on the features of their language but in some cases, when no verbal input was explicitly involved in the task, the language effect was only superficial. These findings confirm, at least to some extent, the impact of typological constraints in the representational system and support a moderate view about the impact of focality on the Language-Thought relation: some languages (e.g. English) allow for more focal strategies of processing than others (e.g. French), while holistic ways of thinking and ambient processing are not entirely exclusive to Eastern cultures as previously claimed.

References

Chomsky, N. 2014. The Minimalist Program: 20th Anniversary edition. Cambridge, MA: MIT Press.

Ji, Y., Hohenstein, J. 2017. Conceptualising voluntary motion events beyond language use. Lingua 195, 57-71.

Kovas, Y., Plomin, R. 2006. Generalist genes. Trends in Cog. Sciences 10(5), 198-203.

Nisbett, R.E., Peng, K., Choi, I., Norenzayan, A. 2001. Culture and systems of thought: holistic vs. analytic cognition. Psychol. Rev. 108, 291-310.

Pannasch, S., Velichkovsky, B.M. 2009. Distractor effect and saccade amplitudes. Visual Cognition 17(6–7), 1109-1131.

Soroli, E., Hickmann, M., Hendriks, H. 2018. Casting an eye on motion events. In Aurnague, M., Stosic D. (Eds.), The semantics of dynamic space in French, 381-438. Amsterdam: John Benjamins.

Tai, J. 2003. Cognitive relativism: Resultative constructions in Chinese. Language and Linguistics 4(2), 301-316.

Talmy, L. 2000. Toward a cognitive semantics. Cambridge, MA: MIT Press.

Effects of Cognitive Impairment on vowel duration

Charalambos Themistocleous[1,2], Dimitrios Kokkinakis[1], Marie Eckerström[3], Kathleen Fraser[1], Kristina Lundholm Fors[1]

[1]Department of Swedish, University of Gothenburg, Sweden
[2]Department of Neurology, Johns Hopkins University, USA
[3]Department of Psychiatry and Neurochemistry, University of Gothenburg, Sweden

https://doi.org/10.36505/ExLing-2018/09/0027/000360

Abstract

Mild cognitive impairment (MCI) is a neurological condition, which is characterized by a noticeable decline of cognitive abilities, including communicative and linguistic skills. In this study, we have measured the duration of vowels produced in a reading task by 55 speakers— 30 healthy controls and 25 MCI—. The main results showed that MCI speakers differed significantly from HC in vowel duration as MCI speakers produced overall longer vowels. Also, we found that gender effects on vowel duration were different in MCI and HC. One significant aspect of this finding is that they highlight the contribution of vowel acoustic features as markers of MCI.

Key words: MCI/AD, vowel duration, language pathologies, Swedish

Introduction

Mild Cognitive Impairment is a neurodegenerative condition that causes small but noticeable and measurable decline in cognitive abilities, including memory and language skills that differ from normal ageing. Often MCI is accompanied by depression, anxiety, aggression, irritability, and apathy. People with MCI have memory difficulties—such as remembering events and situations—in decision making, planning, interpreting instructions, and in finding their way in familiar environments. There is not a single cause of MCI, thought biomedical—such as shrinkage of the hippocampus, enlargement of the ventricles, abnormal clumps of beta-amyloid protein and tau—genetic—such as the presence of the APOE-e4 gene also linked with AD—and lifestyle factors—such as smoking, high blood pressure, luck of physical exercise, and diabetes, are high risk factors that can contribute to MCI. The aim of this is study is to determine the effects of MCI vs healthy controls (HC) on vowel duration. Vowel duration is extremely sensitive to linguistic (such as vowel quality, stress, positive in the utterance, etc.), emotional, and physiological (age, medical condition etc.) factors (Haley & Overton, 2001, Themistocleous 2017). Previous research showed that vowel duration can be affected in MCI (Jarrold, et al. 2014).

Methodology

The acoustic materials of this study were collected from a subcohort of the Gothenburg MCI study (Wallin et al., 2016), which is a large longitudinal research on MCI. Specifically, 13 female and 12 male participants with MCI and 19 female and 11 male healthy control (HC) speakers participated in this study. The two groups did not differ with respect to age ($t(52.72) = -1.8178$, p = n.s.) and gender ($W = 1567.5$, p = n.s.), as it is evident by the non-significant results from a *t* test and an independent 2-group Mann-Whitney U Test respectively. All speakers were native speakers of Swedish. Speaker selection was based on certain inclusion and exclusion criteria: they should not suffer from dyslexia, deep depression, substance abuse, and other serious psychiatric, neurological or brain-related diseases, such as Parkinson's disease. Ethic approvals for the study were obtained by the local ethical committee review board (number: L09199, 1999; T479- 11, 2011); while the currently described study was approved by the local ethical committee decision 206-16, 2016 and T021-18.

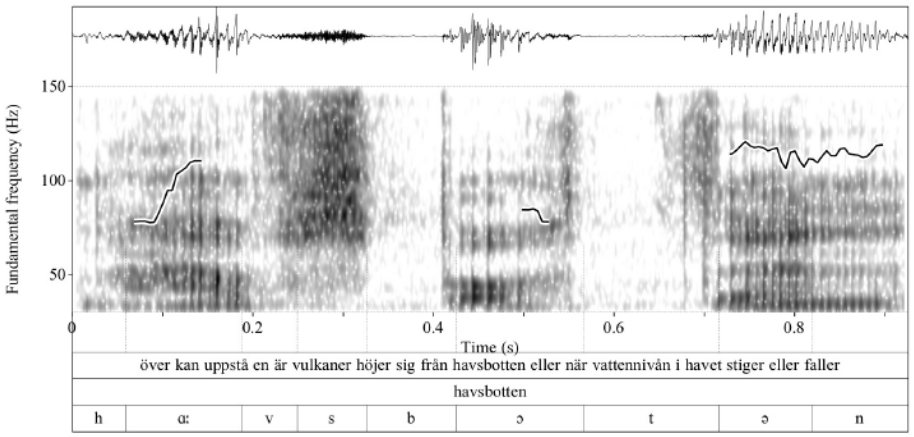


Figure 1. An example of one transcription generated using Themis-SV.

The data were collected from a reading task where participants read a short passage. Participants were instructed to read the text aloud and without interruptions. The narratives were audio-recorded and transcribed and segmented using Themis-SV a system for the automatic transcription of Swedish. The system processes these recordings and returns an output with three tiers: the utterance tier, the word tier, and the vowels and consonants tier (see an example output Figure 1). The output of the system is a fast and reliable transcription and segmentation of speech, which is very close to transcriptions and segmentations performed manually. The automatic segmentation of speech enables targeted acoustic measurements, such as measurements of consonant spectra, formant frequencies of vowels, fundamental frequency, pauses, speech rate, etc. and other acoustic measurements that have been known to

differentiate between the different types of language disorders. All automatic transcriptions were evaluated manually by the first author. For the statistical analysis, we run a linear mixed effects model:
duration ~ condition + (1 | gender) + (1 | vowel)
where duration is a depended variable, condition (MCI vs. HC) a fixed factor, and gender and vowel are random intercepts. The duration was log-transformed to correct for normality. Linear mixed model was fit by REML t-tests using Satterthwaite approximations to degrees of freedom. All statistics were conducted in R using the lme4 and lmerTest packages.

Results

The mean values and the standard deviation (SD) are presented in Table 1. The results of the statistical model are presented in Table 2. Overall, MCI speakers produced longer vowels whereas HC produced shorter vowels. This finding was significant (see Table 2).

Table 1. Mean and SD of vowel duration in ms for HC and MCI female and male participants.

	HC		MCI	
	Mean	SD	Mean	SD
Female	99	56	101	60
Male	89	52	98	58

Table 2. Results of the linear mixed effects model for the effects of condition on the log-transformed duration.

	Estimate	SE	df	t value	Pr(>|t|)
Intercept	4.528	0.082	9	55.154	.0001
Condition MCI	0.043	0.009	13070	4.902	.0001

Table 3. Results of the linear mixed effects model for the effects of the interaction condition × gender on the log-transformed duration.

	Estimate	SE	df	t value	Pr(>|t|)
Intercept	4.59	0.0695	17	66.037	.0001
Condition-MCI	-0.0019	0.0117	13070	-0.162	.871
Gender-Male	-0.1378	0.0122	13070	-11.294	.0001
Condition-MCI: Gender-M	0.1061	0.0179	13070	5.934	.0001

However, MCI and HC participants had great variation from the mean. And there was a clear difference between the two genders on vowel duration. To this purpose, we changed our initial model and added gender as a fixed factor and explored the interaction of gender and condition on vowel duration (adding only vowel as random intercept) (see Table 3). The findings in this case

show a significant effect of the interaction of condition × gender on the log-transformed duration.

Discussion

Overall, participants with MCI produce longer vowels than healthy controls. Longer vowels can be associated with an overall slower MCI speech than HC speech, which can be attributed to slower articulatory movements, greater cognitive processing and planning time of utterances. Nevertheless, future research is required to determine the durational properties of the speech of MCI vs. HC participants. Another important finding is that men MCI participants produce longer vowels than women MCI and HC participants. This finding may be attributed to gender specific effects on speech production in MCI Swedish speakers. Future research will investigate the effects of MCI vs. healthy controls on the duration of both vowels and consonants to establish if these findings are attested in consonants as well.

Acknowledgements

Riksbankens Jubileumsfond – The Swedish Foundation for Humanities & Social Sciences, through the grant agreement no: NHS 14-1761:1.

References

Haley K.L., Overton H.B. 2001. Word length and vowel duration in apraxia of speech: The use of relative measures. Brain and Language 79, 397–406.

Fraser K., Lundholm Fors K., Eckerström M., Themistocleous C., Kokkinakis D., 2018. Improving the Sensitivity and Specificity of MCI Screening with Linguistic Information. LREC workshop: RaPID-2. Miyazaki, Japan.

Jarrold, W., Peintner, B., Wilkins, D., Vergryi, D., Richey, C., Gorno-Tempini, M.L., Ogar, J. 2014. Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. In: CLPsych 2014, 27–36.

Themistocleous Ch. 2017. Modern Greek vowels and the nature of acoustic gradience. Phonetica 74, 157–172.

Themistocleous Ch., Kokkinakis D. 2018. THEMIS-SV. Proc. 4th European Stroke Organisation Conference. Gothenburg, Sweden.

Wallin A. et al. 2016. The Gothenburg MCI study: Design and distribution of Alzheimers disease and subcortical vascular disease diagnoses from baseline to 6-year follow-up. J Cer Blood Flow Metab. 36(1):114-131.

Investigating the phonetic expression of successful motivation

Jana Voße[1,2], Petra Wagner[1,3]
[1]Phonetics and Phonology Work Group, Bielefeld University, Germany
[2]Department of Philosophy, Gothenburg University, Sweden
[3]CITEC, Bielefeld University, Germany
https://doi.org/10.36505/ExLing-2018/09/0028/000361

Abstract

The present study provides a comprehensive acoustic phonetic analysis of motivational speech by collecting, annotating and processing 50 minutes of speech data representing less and more successful degrees of motivation. The analysis shows significant differences regarding the acoustic phonetic features f_0 (median, range, variation), intensity (median, range) and speaking rate. We observe inconsistent results for the variation of intensity, pointing to the necessity of a more fine-grained analysis of this feature. This study provides first support for the existence of a specific motivational speaking style.

Key words: acoustic phonetics, motivation, speaking style, emotional speech

Introduction

The concept of motivation is a frequently observed phenomenon in everyday human-human interaction, but also in specific domains like teaching, coaching or nursing. In such interactive situations, linguistic communication is probably the most intuitive way to create a motivational impact. This paper investigates the role of acoustic phonetic parameters within motivational speech.

Although the concept of motivational speech has not been studied intensively so far, we observe research progress on the phonetic expression of related concepts of motivational speech, such as charismatic (Niebuhr et al. 2016) and volitional speech (Skutella et al 2014). These concepts correspond with respect to the characteristics of their acoustic phonetic features, e.g. f_0, intensity, and speaking rate.

In creating a motivational impact, emotions play a substantial role. Following the concept of emotional empathy, the emotional state of a recipient can be influenced by a speaker's displayed emotion. By expressing a positive emotion, a speaker can set the recipient into a positive state, which in turn influences the recipient's readiness to be motivated positively (Abele 1999). For the expression of emotions, acoustic phonetic features such as speaking rate, f_0 (Burkhardt et al. 2000), and intensity (Tao et al. 2005) are strong means.

Because of the causal relation of phonetics, emotions and motivation and the pragmatic proximity of motivational, charismatic and volitional speech, we

hypothesize motivational speech to be characterized similarly. Specifically, we expect motivational speech to be expressed by the following parameters: (1) *Speaking rate:* high number of syllables/second, (2) *f_0:* high median (logHz), range (logHz) and variation coefficient, (3) *Intensity:* high median (dB), range (dB) and variation coefficient.

Methodology

We collected, annotated and processed 50 minutes of speech data representing less and more successful degrees of motivation. Based on these, we identified and analyzed our set of acoustic phonetic features potentially relevant for motivational impact. The data consists of the audio extracted from 6 motivational YouTube videos, each presented by a different female speaker aged between 16 and 30 years. The aim of these videos is to motivate their audience to engage in sports and to be on a healthy diet. While presenters' age, gender, video topic and structure as well as upload date are homogeneous, the videos differ in their online ratings. We used these ratings to differentiate between more and less successful motivation. This left us with 3 videos of less successful (15 minutes), and 3 videos of more successful motivation (35 minutes).

The data were force-aligned with AlignTool (Schillingmann et al. 2018) both on a phone and syllable level and corrected manually. Perceptually labeled Interpausal Units (IPUs) are used as a measure of utterance segmentation (mean pause duration = 0.45s). Acoustic phonetic features were measured within IPUs using Praat scripts and served as dependent variables in the subsequent analyses. We assume that they differ significantly between less and more successful levels of motivation. Due to the non normal distribution and high correlation of the dependent variables, statistical analyses are carried out by a series of non-parametric tests (Bonferroni-corrected).

Results

All dependent variables show significant differences between *more motivational speech* (MMS) and *less motivational speech* (LMS), except for intensity coefficient of variation. We further observe higher medians in MMS than in LMS for all parameters, except for intensity coefficient of variation, which shows the opposite case. According to the Brown-Forsythe test, the intensity coefficient of variation, f_0 median and f_0 range show homogeneous variances between MMS and LMS. Regarding the form of distribution (tested with Kolmogorov-Smirnov), all dependent variables are characterized by heterogeneous distributions of MMS and LMS.

Table 1. Summary of various test results for all dependent variables.

Dep. variable	MMS (median)	LMS (median)	Wilcoxon rank-sum ('greater')	Brown-Forsythe	Kolmogorov-Smirnov ('two.sided')
Speaking rate (sylls/s)	4.997	4.728	$W = 73383$, $p < 0.01$*	$F = 30.771$, p<0.001***	$D = 0.1671$, $p < 0.01$**
f_0 median (logHz)	2.385	2.355	W = 81209, $p < 0.001$***	$F = 3.9884$, $p > 0.05$	$D = 0.2718$, $p < 0.001$***
f_0 range (logHz)	0.376	0.248	$W = 89347$, $p < 0.001$***	$F = 1.256$, $p > 0.05$	$D = 0.32431$, $p < 0.001$***
f_0 (variation coefficient)	0.705	0.631	W =88206, $p < 0.001$***	$F = 8.7725$, $p < 0.05$*	$D = 0.30885$, $p < 0.001$***
Intensity median (dB)	73.126	60.147	$W = 114260$, $p < 0.001$***	$F = 158.12$, $p < 0.001$***	$D = 0.65703$, $p < 0.001$***
Intensity range (dB)	37.531	35.851	W =77291, $p < 0.001$***	$F = 6.9092$, $p > 0.05$	$D = 0.15694$, $p < 0.01$**
Intensity (variation coefficient)	0.112	0.136	$W = 37232$, $p > 0.05$	$F = 3.7247$, $p > 0.05$	$D = 0.35986$, $p < 0.001$***

Discussion

We observe statistically significant medians and distributions in MMS and LMS for all dependent variables except for the intensity coefficient of variation (median). Regarding variance, only half the dependent variables show significant results. Obtaining a clear differentiation of MMS and LMS in most dependent variables supports our assumption of a motivating prosodic speaking style contrasting with a less-motivating one. Future perception experiments will investigate whether these production differences are perceptually relevant.

Regarding the assumption of a motivating speaking style, it must be considered that the observed parameter shapes might be speaker-intrinsic rather than articulatorily targeted in a conscious manner, as our study follows a between-subjects design. Testing motivational stimuli in a within-subject design will provide further insight regarding this matter.

Although the results of the analysis of speaking rate, the f_0 parameters, and intensity mean support our hypothesis regarding the relation between successful motivational speech and charismatic, volitional, and positive emotional speech, we observe differing results regarding intensity variation. A more fine-grained analysis is needed to investigate the role of this parameter further.

For the interpretation, it must be also considered that the chosen unit of analysis (IPU) affects the results. Analysing the given phonetic features on a different level might result in divergent observations.

We are aware that the audio quality of the recorded videos impacts the analysed parameters, especially those of intensity. Hence, the interpretation of the intensity must be considered with reservation. Future experiments with controlled audio qualities will be carried out to examine the validity of the results of the present study. Another point of discussion is the validity of online rankings as a criterion for differentiating levels of more and less successful motivation. Perception experiments are planned to substantiate the approach taken here.

To conclude, our study indicates that successful motivational speech is characterized by a high and variable pitch as well as by a loud and fairly fast articulation, but with a potentially stable intensity within individual utterances.

References

Abele, A. 1999. Motivationale Mediatoren von Emotionseinflüssen auf die Leistung: Ein vernachlässigtes Forschungsgebiet. In Jerusalem, M. Pekrun, R. (Eds.) 1999, Emotion, Motivation und Leistung, 31-50. Göttingen, Bern, Toronto, Seattle, Hogrefe-Verlag.

Burkhardt, F., Sendlmeier, W.F. 2000. Verification of acoustical correlates of emotional speech using formant-synthesis. In SpeechEmotion-2000, 151-156, Newcastle, Northern Ireland, UK.

Niebuhr, O., Voße, J., Brem, A. 2016. What makes a charismatic speaker? A computer-based acoustic-prosodic analysis of Steve Jobs tone of voice. Computers in Human Behavior 64 366-382.

Schillingmann, L., Ernst, J., Keite, V., Wrede, B., Meyer, A. S., Belke, E. 2018. AlignTool: The automatic temporal alignment of spoken utterances in German, Dutch, and British English for psycholinguistic purposes. Behaviour Research Methods 50(2), 466-489.

Skutella, L.V., Süssenbach, L., Pitsch, K., Wagner, P. 2014. The prosody of motivation. First results from an indoor cycling scenario. In Hoffmann, R. (Ed.), Elektronische Sprachsignalverarbeitung 2014, 71, Dresden, Germany.

Tao, J., Kang, Y. 2005. Features importance analysis for emotional speech classification. In Tao, J., Tan, T., Picard, R.W. (Eds.), International Conference on Affective Computing and Intelligent Interaction, 449-457, Beijing, China.

Analysis of vocal implicit bias in SCOTUS decisions through predictive modelling

Ramya Vunikili[1], Hitesh Ochani[1], Divisha Jaiswal[1], Richa Deshmukh[1], Daniel L. Chen[2], Elliott Ash[3]

[1]Department of Computer Science, New York University, USA
[2]University Toulouse Capitole, France
[3]Center for Law and Economics, ETH Zurich, Switzerland

https://doi.org/10.36505/ExLing-2018/09/0029/000362

Abstract

Several existing pen and paper tests to measure implicit bias have been found to have discrepancies. This could be largely due to the fact that the subjects are aware of the implicit bias tests and they consciously choose to change their answers. Hence, we've leveraged machine learning techniques to detect bias in the judicial context by examining the oral arguments. The adverse implications due to the presence of implicit bias in judiciary decisions could have far-reaching consequences. This study aims to check if the vocal intonations of the Justices and lawyers at the Supreme Court of the United States could act as an indicator for predicting the case outcome.

Key words: Speech analysis, Implicit gender bias, Machine learning, SCOTUS, FAVE

Introduction

Supreme Court of the United States (SCOTUS) is the highest federal court of the country. The cases heard by it are of utmost importance. This gives us the major motivation to employ machine learning techniques to check for the presence of any implicit bias. The SCOTUS comprises of the Chief Justice of United States and eight associate judges. There lies a huge responsibility in their hands to make rational decisions. However, it would be unrealistic to assume that all decisions are rational and unbiased.

This study aims to explore the relationship between implicit gender bias in the oral arguments and the final outcome of the case. In other words, we analyze if the features related to vocal intonations and masculine/feminine style of the speaker is an indicator of their implicit gender bias.

Related work

According to a study done by Chen *et al.* (2016), it is observed that the perceived masculinity has a negative correlation with the winning of a case. Further, studies done by Klofstad *et al.* (2012) and Tigue *et al.* (2012), claim that individuals with lower-pitched male voices are often associated with higher competence and trustworthiness.

Dataset

The SCOTUS oral arguments have been recorded since October 1995. These recordings along with their transcriptions are available on the Oyez website (See: https://www.oyez.org/). This study uses 1246 cases from the SCOTUS collected during the years 1998, 1999 and 2003-2012. In addition to the recordings and transcriptions of these cases, we also gathered information about the Justices (gender, year of birth, party of appointing President, Segal-cover score etc), lawyers (gender, total number of cases involved, number of cases that involves him\her as a petitioner, number of cases that involves him\her as a respondent etc) along with the case specific information like the issue date, name of the case and the winner etc. In total there are about 2,137 hours of lawyers' recordings and 502 hours of the Justices' recordings.

Based on the type of speakers and their order of speech there are two types of pre-processed datasets – ABA and AxByA. In the AxByA dataset, A and B refer to two different Justices while x and y can represent same or different lawyers. Similarly, in the ABA format, A is always a Justice and B can represent both lawyers or Justice.

Methodology

Data Pre-processing

Using a list of 135 masculine words (such as uncle, man etc) and 135 feminine gendered words (such as sister, waitress etc), we classified all the relevant words spoken by the Justices and lawyers from ABA and AxByA datasets into three classes – masculine, feminine and neutral (neither masculine nor feminine). Among these, 60% were used as the training set, 20% as validation set and the remaining 20% as the test set.

In order to perform hard classification on each dataset, we trained a random forest classifier with hyperparameters optimized based on the validation set. With the number of estimators for the model fixed as 100, we achieved an accuracy of 78.9% on the AxByA test set and an accuracy of 83.3% on the ABA test set.

Further, we added features related to the interruption of a speaker based on the timestamps in the transcriptions i.e., if a Justice has interrupted a lawyer or a Justice has been interrupted by another Justice.

Modelling

In order to predict if the vote of a particular Justice is going to be in favour of or against a lawyer, we've trained two models. They are Extreme Gradient Boosting (XGBoost – Baseline) and Linear Support Vector Machine (SVM – Enhanced Model). For each case, we've extracted features such as the number of masculine and feminine words spoken by the Justice and the lawyer, the number of neutral words spoken by each of them that are classified into

masculine and feminine words, the number of times a Justice was interrupted by male/female lawyers and the number of times a Justice interrupts a male/female lawyer, gender of the lawyer and the Justice, the ratio of neutral words that are classified into masculine words for a Justice and the ratio of neutral words that are classified into feminine words for the same Justice. These features were then normalized before training the models.

The best hyperparameters for each model are retrieved by tuning the models on the validation set. These hyperparameters were then used for prediction on the test set. Table 1 and Table 2 give the list of hyperparameter for each model.

Table 1. XGBoost Hyperparameters.

XGBoost Parameter	Value
learning_rate	0.03
max_depth	10
n_estimators	50
objective	binary:logistic

Table 2. SVM Hperparameters.

SVM Parameter	Value
C	0.03
loss	hinge
penalty	L2
tol	0.0001

Results

From Table 3, it can be observed that SVM performs better than XGBoost in predicting the vote of a Justice. While the accuracy of XGBoost is only about 46.85% on the test set, the SVM has a slightly better accuracy of 51.13%.

Table 3. Accuracy of the models on training and test sets.

Model	Train Accuracy	Test accuracy
XGBoost	60.03%	46.85%
SVM	51.57%	51.13%

Though the accuracy of prediction in either case isn't outstanding the most important features that contributed to the vote prediction such as the number of masculine/feminine words spoken by a Judge and their ratio with the neutral words are found to be at the top in both the models. This can be observed in Figure 1.

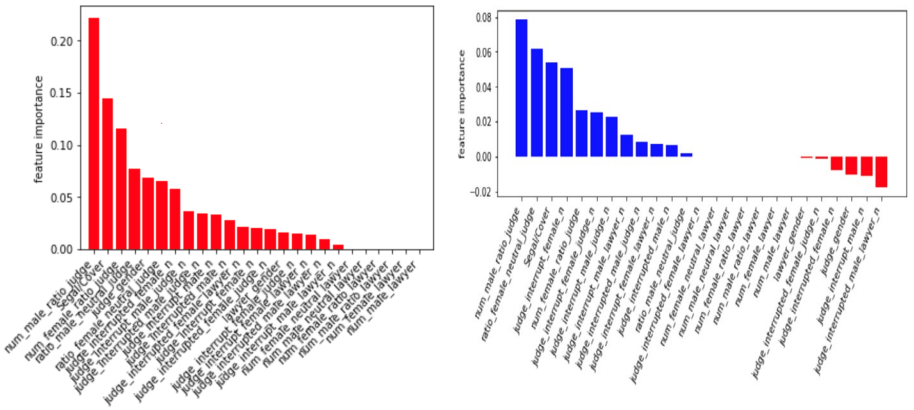


Figure 1. Feature importance of XGBoost (on the left) and SVM (on the right).

Acknowledgements

We've developed this project based on the technical assistance provided by Mufei Li, Kailin Luo, Jiayun Yu, Xuan Wang.

References

Chen, D.L., Halberstam, Y, Yu, A. 2016. Covering: Mutable Characteristics and Perceptions of Voice in the US Supreme Court, No. 16-180. Toulouse School of Economics (TSE).

Klofstad, C.A., Anderson, R.C. Peters, S. 2012. Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. Proceedings of the Royal Society B: Biological Sciences, 279 (1738), 2698–2704.

Tigue, C.C., Borak, D.J., O'Connor, J.J.M, Schandl, Ch., Feinberg, D.R. 2012. Voice pitch influences voting behavior. Evolution and Human Behavior 33, 210–216.

Index of names

ExLing 2018
Ebook ISSN: 2529-1092
Ebook ISBN: 978-960-466-198-5