



International Speech Communication Association

ExLing 2010

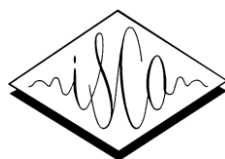
Proceedings of 3rd Tutorial and Research Workshop on Experimental Linguistics

25-27 August 2010, Athens, Greece

Edited by Antonis Botinis



National and Kapodistrian
University of Athens



International Speech Communication Association

ExLing 2010

Proceedings of 3rd Tutorial and Research Workshop on
Experimental Linguistics

25-27 August 2010, Athens, Greece

Edited by Antonis Botinis



National and Kapodistrian
University of Athens

ExLing 2010

Proceedings of 3rd Tutorial and Research Workshop on Experimental Linguistics

Ebook ISSN: 2529-1092

Ebook ISBN: 978-960-466-157-2

Copyright © 2010 Antonis Botinis

The present volume presents the proceedings of ExLing 2010, the 3rd Tutorial and Research Workshop on Experimental Linguistics, in Athens, Greece, 25-27 August 2010. This conference is regularly organised in Athens, since 2006.

We have thus gathered in Athens to discuss current issues and methodological approaches in linguistics. Our main interest is the development of theory based on experimental evidence as well as the unification of different aspects of linguistics research. Phonetics, as a well-established experimental discipline, is in the center of our discussion, along with diverse other disciplines related to the study of language.

Our colleagues are coming from different parts of the world and we wish them a fruitful scientific exchange. ISCA Workshops are indeed an ideal environment for the promotion of new ideas as well as the development of research collaborations.

We would like to thank all participants for their contributions and the International Advisory Committee as well as ISCA for its support in organising this conference. Special thanks to the University of Athens for the publication of the proceedings as well as our students from the University of Athens for their assistance.

Antonis Botinis

Dedicated to the memory of Robert Bannert

On Friday 16th 2010, our colleague and friend Robert Bannert passed away in his home in Freiburg, Germany, at the age of 68. Robert was watching the evening news on Thursday with his beloved wife, Brita, when he was hit by a severe stroke. Although taken immediately to the University hospital, Robert left us for ever the next day, Friday evening.

Robert and I first met at a small party at the Linguistics Department, Lund University, in the early seventies. Some time later, Robert was showing me some intonation curves of a Greek phrase at the Phonetics Department, arguing about intonation similarities in Greek and Swedish. During this time, Robert was working with phonetic typology and some dozens of diverse languages. Later, upon the conclusion of my undergraduate studies, Robert urged me to continue and pursue postgraduate studies.

After studying, teaching and carrying out research for many years at Lund University, obtaining his PhD in 1976, Robert moved to the Department of Phonetics at Stockholm University in the mid-eighties and, in 1992, he was appointed professor of phonetics at Umeå University in northern Sweden. During the last years, Robert had been collaborating with the University of Freiburg, where he had been declared Honoured Professor. Recently, Robert was retired and lived permanently with his wife Brita in a quite neighbourhood on the outskirts of Freiburg.

Much is due to Robert for the present Workshop. In the mid-nineties, Robert and I were sitting and discussing at my home in Athens about the organization of an ISCA event in Athens. Since both of us were prosodists, we agreed on a Workshop on prosody and Robert used all his influence for a positive outcome, which turned out as the Workshop on prosody in 1997. After some years, in Stockholm, we were discussing the prospects of an ISCA Workshop on Experimental Linguistics, which led to the setup of the first one in 2006 and its repetition every two years thereafter.

Robert has been in all our earlier Workshops but he wouldn't come this time. He was telling me that the summer heat in Athens was becoming intolerable, something that was unlike Robert, as he was always coming at any time, especially in summer. This was our last conversation this summer.

Robert and I have been good friends for nearly forty years. We have been in every opportunity at each other's home all these years, wherever we happened to live, and in many places in different countries. Many times, when driving through Germany, Robert and Brita were just waiting for me.

I am very lucky and privileged that I came across to meet and become friends with Robert. An honest, straightforward and friendly gentleman with positive attitude, strong will and determination in every step of his life.

Antonis Botinis

Contents

<i>Language difficulties in children with Attention Deficit Hyperactivity Disorder</i> Paraskevi Agapitou, Ioannis Galantomos, Georgia Andreou	1
<i>Human-computer interaction systems and elderly users: Linguistic specifications for prosodic modeling</i> Christina Alexandris	5
<i>Importing mental patterns of coherence as a comprehension strategy: a case study of errors in reference assignment</i> Pilar Alonso	9
<i>The impact of order and aspect in processing of causally linked sentences</i> Joanna Blochowiak, Jacques Moeschler, Thomas Castelain	13
<i>The prosody of the TV news speaking style in Brazilian Portuguese</i> Luciana Castro, Ben Serridge, João Moraes, Myrian Freitas	17
<i>Perceptual cues for sentence intonation in Greek</i> Anthi Chaida	21
<i>Text punctuation and prosody in Greek</i> Anthi Chaida, Elina Nirgianaki, Eleni Panaretou	25
<i>Coarticulation of consonant clusters in Italian learners of French</i> Sonia d'Apollito	29
<i>The acoustic analysis of speech samples designed for the Voice Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS): long-term f0 and intensity measures</i> Zuleica Antonia de Camargo, Sandra Madureira	33
<i>Algerian Arabic rhythm classification</i> Ghania Droua-Hamdani, Sid Ahmed Selouani, Malika Boudraa, Wladyslaw Cichocki	37
<i>Acoustic correlates of nasopharyngeal resonance</i> Marios Fourakis, Heather Karlsson, Christie Tilkens, Lawrence Shriberg	41
<i>Dialect and generational differences in vowel space areas</i> Robert Allen Fox, Ewa Jacewicz	45
<i>How to incorporate qualitative analysis into experimental research —Why Japanese students studying in the UK did not improve their fluency? —</i> Misa Fujio	49
<i>Evaluating languages functions in a trilingual setting</i> Ioannis Galantomos, Georgia Andreou, Fotini Anastassiou	53
<i>Production of inflectional morphology in a child with moderate hearing impairment</i> Thalia Hatzivasiliou, Marianna Hatzopoulou	57
<i>How learning to read and write influences the dialectal traits in -sC clusters</i> Isabel Henriques	61

<i>Tarascan dialectal and generational differences around aspirated stops</i> Francisco Javier Bucio García	65
<i>Boundary-related durations in Modern Greek</i> Evia Kainada	69
<i>Action and object naming in mono- and bilingual children with language impairment</i> Maria Kambanaros, Kleanthes K. Grohmann, Eleni Theodorou	73
<i>Is that a bnik I see? Testing phonotactics using word-picture matching</i> Vsevolod Kapatsinski, Lamia Haddad Johnston	77
<i>Embedded aspect in L2 acquisition: Evidence from L1-Russian learners of Greek</i> Sviatlana Karpava, Kleanthes K. Grohmann	81
<i>Evidence for depression and schizophrenia in speech prosody</i> Roi Kliper, Yonatan Vaizman, Daphna Weinshall, Shirley Portuguese	85
<i>The effect of cochlear implants on phonological acquisition</i> Evaggelia Lazarou, Marianna Hatzopoulou	89
<i>Observations on the prosodic characteristics of Finnish-speaking youngsters with Asperger Syndrome</i> Mari Lehtinen	93
<i>Metalanguage or bidialectism? Acquisition of clitic placement by Hellenic Greeks, Greek Cypriots and binationalals in the diglossic context of Cyprus</i> Evelina Leivada, Paraskevi Mavroudi, Anna Epistithiou	97
<i>Contextual information in (non) native speech perception in noise</i> Bin Li, Ratrete Wayland, Caicai Zhang	101
<i>Exploring sound symbolism in the investigation of speech expressivity</i> Sandra Madureira, Zuleica Antonia de Camargo	105
<i>A dynamic view of the prosodic structure: The example of French</i> Philippe Martin	109
<i>Acquisition of lightverbhood: Evidence from Cypriot Greek</i> Paraskevi Mavroudi, Evelina Leivada, Anna Epistithiou	113
<i>Phonetic factors influencing /l/-rhoticisation in Greek</i> Daniela Müller	117
<i>Perception of “tonal focus” in Greek</i> Olga Nikolaenkova	121
<i>Acoustic structure of fricative consonants in Greek</i> Elina Nirgianaki, Anthi Chaida, Marios Fourakis	125

<i>The IPA training on the pronunciation of difficult English words for Cantonese speakers</i>	129
Sunyoung Oh	
<i>Towards an articulatory characterization of European Portuguese /l/</i>	133
Catarina Oliveira, António Teixeira, Paula Martins	
<i>A complexity science perspective on language spread</i>	137
Michał B. Paradowski	
<i>Inda mbu, nambu or ine ti pu? Acquiring complex structures in Cypriot Greek</i>	141
Natalia Pavlou	
<i>Two languages – one brain</i>	145
Maija S. Peltola, Henna Tamminen, Laura Salonen, Heidi Toivonen, Teija Kujala, Risto Näätänen	
<i>Real and acted responses of distress: an auditory & acoustic analysis of extreme stress & emotion</i>	149
Lisa Roberts	
<i>How context affects perception: judging distress & linguistic content in forensic audio recordings</i>	153
Lisa Roberts, Peter French	
<i>Intonational strategies for backchanneling in Italian Map Task dialogues</i>	157
Michelina Savino	
<i>Improving intelligibility of synthesized speech in noise with emphasized prosody</i>	161
Sunil R. Shukla	
<i>There is more to chat discourse than written conversation</i>	165
Cláudia Silva	
<i>Visual attention during L1 and L2 sounds perception: an eye-tracking study</i>	169
Bianca Sisinni, Mirko Grimaldi, Elisa Tundo, Andrea Calabrese	
<i>Pause ratio effect in emotion ascribing</i>	173
Eszter Szabó	
<i>Prosodic evidences to sentential negation in Brazilian Portuguese</i>	177
Lílian Teixeira de Sousa	
<i>Focus prominence and tonal alignment in Athenian and Cypriot Greek</i>	181
Charalampos Themistocleous, Stelios Kyriacou	
<i>Narratives in Cypriot Greek mono- and bilingual children with SLI</i>	185
Eleni Theodorou, Kleanthes K. Grohmann	
<i>English communication skills for prospective English teachers in Japan</i>	189
Yuichi Todaka	
<i>Morphological productivity in Maltese verbs</i>	193
Alina Twist	

Language difficulties in children with Attention Deficit Hyperactivity Disorder

Paraskevi Agapitou¹, Ioannis Galantomos², Georgia Andreou³

¹Department of Special Education, University of Thessaly, Greece

²Department of Special Education & Department of Language, Literature and Culture of the Black Sea Countries, University of Thessaly & Democritus University of Thrace, Greece

³Department of Special Education, University of Thessaly, Greece

Abstract

In this study we review the relevant literature regarding the connection of Attention deficit hyperactivity disorder (ADHD) with language disorders. Special emphasis is given to the evolution of language disorders in relation with ADHD, because both of them appear during preschool age and influence in a crucial manner the acquisition of basic learning readiness skills. The review of a considerable body of literature indicates that the co morbidity of ADHD with language difficulties worsens children's behaviour problems at preschool age and affects their socialization and academic performance. Hence, it is concluded that a proper and valid diagnosis of ADHD and language difficulties plus combined and suitable interventions will reduce the amount and quality of ADHD symptoms.

Key words: ADHD, language difficulties, behaviour

Introduction

ADHD is the most common psychiatric disorder in childhood, affecting approximately 3–7% of school-age children. This disorder has specific subtypes and exhibits differences in children who display predominantly inattentive behaviours, hyperactive and impulsive behaviour or a combination of both (Barkley 1998; APA 2000). Although there has been a number of studies describing the relationship between language disorders and behaviour disorders in children there have been relatively few data available that directly examine the language skills of children with a particular subtype of behaviour disorder. Four possible theoretical models have been proposed: (a) behaviour problems are a consequence of children's frustration with their reading difficulties; (b) behaviour problems are a cause of reading difficulties; (c) behaviour problems and reading difficulties are neither a cause nor a consequence but share a common cause; or (d) all of the proposed relationships could be active in some sense, with bi-directional or transactional influences best characterizing the relationship between behaviour and achievement (Spira & Fischer 2005).

Comorbidity

Language impairments are commonly observed among children referred for psychiatric services. The most frequent psychiatric diagnosis of children with language impairment (LI) is ADHD. It is not clear whether there are differences between children with ADHD and comorbid LI and children with other psychiatric disorders who are also comorbid for LI. Developmental disorders might differ in their language profiles when using parent reports (Geurts et al. 2008).

Many of the studies of the language abilities of children with ADHD have not yet distinguished between children with ADHD alone and those with comorbid disorders. Thus, it is difficult to determine whether the language impairments are specific to ADHD or whether they are more a corollary of the comorbid disorder (Purvis & Tannock 1997). Cohen et al. (2000) examined four groups: children with ADHD-LI, children with ADHD who have normally developing language, children with psychiatric diagnoses other than ADHD with language impairment or without LI. Results indicated that children with LI were the most disadvantaged regardless of the nature of the psychiatric diagnosis.

Language and academic deficits

It is clear that some children with symptoms of ADHD at preschool age are at risk for long-lasting behavioural difficulties. Additional research suggests that children who manifest behavioural difficulties associated with ADHD at preschool age may be at risk for academic underachievement as well. There is a large amount of literature on the relationship between learning problems and ADHD in elementary-school-aged children. For instance, Agapitou & Andreou (2008) found that ADHD preschoolers who demonstrate serious language deficits are more liable to develop poor reading skills in primary school.

Children diagnosed with Attention Deficit Hyperactivity Disorder (ADHD) have cognitive processing difficulties due to their disinhibition and attention deficits, which influence their scholastic performance. Cognitive processing has also an impact on the production of oral narratives, an essential skill required for academic success. According to Moonsamy et al. (2009), narrative measures are useful instruments for oral language evaluation in children with ADHD and the importance of understanding oral narratives within a therapeutic situation is important for both therapist and child.

Thus, problems with language and pragmatics seem to be associated with the typical problems with learning and social skills in children with ADHD. Children and adolescents with ADHD are slower and less efficient than matched control subjects with regard to complex sentence comprehension (Wassenberg et al. 2010).

Children with ADHD use fewer strategies of textual organization and more avoidance, tangential, and unrelated meanings and more abandoned utterances and spelling and punctuation errors (Mathers 2006).

Conclusion

To sum up, the above mentioned review of a considerable body of recent literature suggests that ADHD along with difficulties manifested in language affect children's behaviour at preschool age in a negative manner. Furthermore, the combination of ADHD with language problems has an impact on children's overall academic performance and socialization skills. Therefore, it is concluded that a valid and in time diagnosis of both ADHD and language problems combined with a multifaceted intervention will reduce the amount of evidenced symptoms and ameliorate children's quality of life.

References

- Agapitou, P., Andreou, G. 2008. Language deficits in ADHD preschoolers. *Australian Journal of Learning Difficulties* 13, 39-49.
- American Psychiatric Association (APA). 2000. *Diagnostic and statistical manual of mental disorders – 4th edition, text revision (DSM-IV-TR)*. Arlington, VA: APA.
- Barkley, R.A. 1998. *Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment*. New York, Guilford Press.
- Cohen, N.J., Vallance, D.D., Barwick, M., Im, N., Menna, R., Horodezky, N.B. 2000. The interface between ADHD and language impairment: An examination of language, achievement, and cognitive processing. *Journal of Child Psychology and Psychiatry* 41, 353–362.
- Geurts, H., Embrechts, M. 2008. Language Profiles in ASD, SLI, and ADHD. *Journal of Autism and Developmental Disorders* 38, 1931-1943
- Mathers, M. 2006. Aspects of Language in Children with ADHD. Applying Functional Analyses to Explore Language Use. *Journal of Attention Disorders* 9, 523-533.
- Moonsamy, S. Jordaan, H., Greenop, K. 2009. Cognitive processing and narrative discourse production in children with ADHD, *South African Journal of Psychology* 39, 326-335.
- Purvis, K.L., Tannock, R. 1997. Language abilities in children with ADHD, reading disabilities and normal controls. *Journal of Abnormal Child Psychology* 25, 133-144.

- Spira, E.G., Fischer, J.E. 2005. The impact of preschool inattention, hyperactivity, and impulsivity on social and academic development: a review. *Journal of Child Psychology and Psychiatry* 46, 755-773.
- Wassenberg, R., Hendriksen, J., Hurks, P., Feron, F., Vles, J., Jolles, J. 2010. Speed of Language Comprehension is Impaired in ADHD. *Journal of Attention Disorder* 13, 374-385.

Human-computer interaction systems and elderly users: Linguistic specifications for prosodic modelling

Christina Alexandris

Department of German, University of Athens, Greece

Abstract

The present linguistic specifications and language-independent parameters for prosodic modelling aim to provide a general framework for facilitating both the construction and the evaluation processes of the Dialog Interface of Human-Computer Interaction Systems in Service for the Elderly and the related prosodic modelling. The proposed specifications target to the features of Comprehensibility and User-friendliness in the spoken output and the overall safety, efficiency and reliability of the System's performance.

Key words: comprehensibility, user friendliness, prosodic emphasis, speed rate

Data and user requirements

General parameters as usable, transferable and language independent specifications are presented here for the Speech Processing Module and Dialog Interface of Human- Computer Interaction Systems in Service for the Elderly. These parameters aim to contribute to the achievement of Comprehensibility and User-friendliness adapted to the needs of the Elderly Users. With the proposed approach, we aim to limit empirical prosodic modelling and case-specific management of a System's spoken output and to provide a general framework for facilitating the construction of the Dialog Interface and the related prosodic modelling for the Elderly as the user group. Here, we focus on experience and data gained from European Union projects (European countries including the Mediterranean and Northern Europe, language and culture-specific factors), namely data from User Requirements Analysis in Work Packages from the respective projects, recorded data, questionnaires, dialog modelling corpora and studio recordings from the respective projects, namely: the SOPRANO Project (involving smart environments and services for the Elderly, the HEARCOM Project (speech technology applications for Users with hearing problems) and the CitizenShield Dialog System for consumer complaints (Greek national project), with senior citizens constituting a significant percentage of the System's users (www.polias.gr). From a psychological perspective, previous studies in the case of Systems involving Human Computer Interaction (HCI), have demonstrated that the User is more likely to use the System and oversee minor flaws if the System is user-friendly and its spoken output is characterized by naturalness and in some cases may create "an

attachment emotion” towards the system (Matsumoto et al., 2009). In this case, the Elderly User is also more likely to make the effort and gain the most benefit from the System’s existing capabilities.

Prosodic modelling constitutes one of the typical strategies used in Dialog Systems (Kellner, 2004), contributing to the task efficiency and service efficiency of the system (Moeller, 2005). In Task-related applications, for the General Public, prosodic modelling is observed to contribute both to the achievement of clarity, lack of ambiguity and user-friendly style (Alexandris, 2009). For the requirements of the Elderly as a User-Group, these targets are summarized as “Comprehensibility” (I) and “User-Friendliness” (II). Here, prosodic modelling is designed within a Speech Act framework. Prosodic modifications are related to paralinguistic elements (prosodic emphasis, pause-length) and pragmatic elements (speech act differentiation). The main features involved in the proposed prosodic modelling are (A) prosodic emphasis, (B) speed, (C) amplitude and (D) overall tone of utterance.

Prosodic emphasis

Prosodic modelling of all utterances related to the Speech Acts for Task-oriented Dialogs is based on the use of prosodic emphasis (A) on the sublanguage-specific elements constituting the most important information in the sentence’s semantic content, as well as sublanguage-independent elements, such as negations and elements expressing time, space (movement), quality and quantity (such as “now”, “over”, “red (button)” and “two (minutes)”), used for the achievement of Precision, while prosodic emphasis on sublanguage-specific expressions and terminology (such as “airconditioner”, “button” and “turned off”) is used for the achievement of the Comprehensibility (I) resulting to Directness, (Alexandris, 2009). From a prosodic aspect, it is observed that User-friendliness (II) can be achieved with prosodic emphasis on expressions related to the User-System Relationship (“Usr-Sys-Rel” expressions), especially in Speech Acts labelled here as “Non-Task-related Speech Acts” such as “Apologize” and “Thank”, or following or preceeding the Task-Related Speech Act (Heeman et al, 1998), for example, the utterance “To make sure that the house is safe, I will ask you a few more questions” (“Introduce-new-task”) following or preceeding the Task-Related Speech Act (“Y/N Question”) based on the User-System Relationship. The “Usr-Sys-Rel” expressions can be subsumed under the general category of expressions involving the System’s or User’s positive intention or cooperation. These words express the system’s intention (action) to serve the user, expressions (usually verbs) indicating the system’s apologies, failure or error in respect to a task executed to serve the user. Usr-Sys-Rel expressions are also observed to be words (usually verbs) expressing the user’s actions or intentions, or words (mainly nouns)

expressing a task related to the actual interaction involving good intentions or a task related to the system's services. Usr-Sys-Rel expressions can be categorized as (1) system-service verbs and nouns (nominalization of verbs) ("cooperation", "give", "show"), (2) system-intention verbs and nouns ("help", "assist") and (3) user-intention (or user intended action) verbs ("wish", "want"). We note that the User-System Relationship is remarkably and evidently expressed in Usr-Sys-Rel verbs in verb-framed and pro-drop languages such as Greek (or, for example, Spanish or Italian), where the emphasized finite verb as a Usr-Sys-Rel expression, contains the features of the verb's subject, in this case the System or the User. Typical examples are the expressions "doso" (=I-give/present), "deixso" (=I-show/present) and "exshypyre'tiso" (=I-give-service/assist).

Speed, amplitude and overall tone

The specific values of the speed rate of the utterances produced in the System's spoken output may vary according to language-type. For the achievement of Comprehensibility (I) and User-friendliness (II) for the Elderly as a user group, as a default parameter, the speed of the spoken output should be lower than the threshold of the average speed of a news broadcast, but higher than the average speed maintained for reading aloud to small children. Another factor contributing to the speed rate is the management of the pauses in utterances produced. A short pause ([Srt-P]) before keywords to be emphasized (and their articles and/or pronouns, if applicable) contributes to the overall speed of the utterances produced. For English, a short pause is also inserted between the components of ACTION-TYPE (Malagardi and Alexandris, 2009) constituting expressions composed of more than one word, such as the expressions "turned-off" and "switched-on", or any elements containing key-information (and receiving prosodic emphasis, here indicated in *italics*), such as in the example: "OK. The [Srt-P] *air conditioner* is *switched* [Srt-P] *on*". The absolute values of the short pauses may vary according language-type, since they are related to the overall speed rate of the utterances produced. Similarly to the feature of speed (B), the specific values of the overall amplitude (C) of the utterances produced in the System's spoken output may also vary according to language-type. For example, some languages appear to have a characteristically loud form of communication by their speakers, in relation to other languages, especially in Northern Europe, where, in most types of transactions and related Speech Acts, lower amplitude is preferred and is acceptable by most users. However, overall amplitude can be adjusted, according to the User's needs. The acceptable tone (D) of voice in the System's spoken output may also vary according to culture and language-type, constituting a substantial factor in the requirements of the Elderly as a

user group: a characteristically vivid and expressive tone is maintained in some languages (i.e. Italian) while in others languages, a tone associated with reliability and responsibility is usually preferred (i.e. German). In most cases, at least for the languages and their Native Speakers in Europe, developers in Human-Computer Interaction systems have to balance between a tone associated with reliability and responsibility, while at the same time User-friendliness and naturalness must be preserved. We note that fine-tuning in respect to appropriate nuance tone is easier to be accomplished by pre-recorded output produced by a trained speaker than by Speech Synthesis, since, unlike prosodic emphasis, amplitude and speed, tone is both highly language and culture specific and cannot be easily encoded and integrated in the System's Speech Processing Module.

References

- Alexandris, Ch. 2009. A speech-act oriented approach for user-interactive editing and regulation processes applied in written and spoken technical texts. In Jacko, J.A. (ed.), *Human-Computer Interaction, Part II, HCII 2009*, vol. 2, LNCS 5611, 645–653, Berlin Heidelberg: Springer-Verlag
- Heeman, R., Byron, D., Allen, J.F. 1998. Identifying discourse markers in spoken dialog. *Proc. of the AAAI Spring Symposium on Applying Machine Learning to Discourse Processing*, Stanford, USA.
- Kellner, A. 2004. Dialogsysteme. In Carstensen, K.U, Ebert, Ch., Endriss, C., Jekat, S., Klabunde, H. Langer, H. (eds.), *Computerlinguistik und Sprachtechnologie, Eine Einfuehrung*, 532-539, Muenchen: Spektrum Akademischer Verlag.
- Matsumoto, N., Ueda, H., Yamazaki, T., Murai, H. 2009. Life with a robot companion video analysis of 16-days of interaction with a home robot in a “Ubiquitous Home” environment, J.A. Jacko (ed.), *Human-Computer Interaction, Part II, HCII 2009*, LNCS 5611, 341–350, Berlin Heidelberg: Springer-Verlag.
- Malagardi, I., Alexandris, Ch. 2009. Verb processing in spoken commands for household security and appliances. In Stephanidis, C. (ed.), *Universal access in HCI, Part II, HCII 2009*, vol. 6, LNCS 5615, 92–99. Heidelberg: Springer-Verlag.
- Moeller, S. 2005. *Quality of telephone-based spoken dialogue systems*. New York, Springer.
- <http://www.soprano-ip.org/>
- <http://hearcom.eu/main.html>
- www.polias.gr

Importing mental patterns of coherence as a comprehension strategy: a case study of errors in reference assignment

Pilar Alonso

Departamento de Filología Inglesa, Universidad de Salamanca, Spain

Abstract

This paper addresses the interface between structural coherence, which builds on the formal and conceptual conventions of discourse, and cognitive coherence, which develops around the internal and external aspects of communication (Pasquier and Chaib-draa 2005). More specifically, the focus is on the cognitive coherence strategies set forth by discourse receivers when they encounter comprehension problems, especially those caused by non-standard uses in structural coherence. Drawing from studies on coherence and comprehension (Giora 1997; Wilson and Matsui 2000; Hobbs 2004), the paper discusses a case study which shows how error in reference assignment may encourage readers to import mental patterns of coherence, external to the discourse, as a repair strategy, thus prioritizing structural adjacency and their search for relevance over comprehension.

Key words: Coherence, comprehension, topic construction, adjacency, relevance

Statement of the problem

It is common in narrative discourse to build readers' expectations on cataphoric referential ties which open empty mental spaces to be filled later on. Thus, cataphoric reference challenges readers' expectations on topic continuity, especially when the complexity or ambiguity of the discourse makes it difficult to assign referents to some pronouns. This is precisely what happens in the passage quoted below, which is the beginning of Katherine Mansfield's short story "The Man Without a Temperament". The personal reference problem occurs at the end of the excerpt:

HE stood at the hall door turning the ring, turning the heavy signet ring upon his little finger while his glance travelled coolly, deliberately, over the round tables and basket chairs scattered about the glassed-in veranda. He pursed his lips—he might have been going to whistle—but he did not whistle—only turned the ring—turned the ring on his pink, freshly washed hands.

Over in the corner sat The Two Topknots, drinking a decoction they always drank at this hour—something whitish, greyish, in glasses, with little husks floating on the top—and rooting in a tin full of paper shavings for pieces of speckled biscuit, which they broke, dropped into the glasses and fished for with spoons. Their two coils of knitting, like two snakes, slumbered beside the tray.

The American Woman sat where she always sat against the glass wall, in the shadow of a great creeping thing with wide open purple eyes that pressed—that flattened itself against the glass, hungrily watching her. And she knoo it was there—she knoo it was looking at her just that way. She played up to it; she gave herself little airs. Sometimes she even pointed at it, crying: "Isn't that the most terrible thing you've ever seen! Isn't that ghoulish!" It was on the other side of the veranda, after all . . . and besides it couldn't touch her, could it, Klaymongso? She was an American Woman, wasn't she, Klaymongso, and she'd just go right away to her Consul. Klaymongso, curled in her lap, with her torn antique brocade bag, a grubby handkerchief, and a pile of letters from home on top of him, sneezed for reply.

The other tables were empty. A glance passed between the American and the Topknots. She gave a foreign little shrug; they waved an understanding biscuit. But he saw nothing. Now he was still, now from his eyes you saw he listened. "Hoo-e-zip-zoo-oo!" sounded the lift. The iron cage clanged open. Light dragging steps sounded across the hall, coming towards him. A hand, like a leaf, fell on his shoulder. A soft voice said: "Let's go and sit over there—where we can see the drive. The trees are so lovely." And he moved forward with the hand still on his shoulder, and the light, dragging steps beside his. He pulled out a chair and **she** sank into it, slowly, leaning **her** head against the back, **her** arms falling along the sides.

At the end of the last paragraph, there are three cases of personal reference, the pronoun “she” and the possessive determiners “her” (here highlighted in bold types), which refer to an entity whose existence, at this early stage in the story, has to be inferred from the narrative. Although the character to which these personal and possessive elements refer becomes the centre of the story from then on, readers have to wait over a thousand words to identify her by name (her first name, Jinnie, is revealed after 1041 words in a 5062 words story, as part of an elusive flashback; and her family name, Mrs. Salesby, occurs even later, after 1664 words). This absence of direct reference identification causes problems in the mental operations of anchoring and distributing characters within the narrative structure, which readers automatically perform as they receive information. Thus, the delay imposed by the literary use of cataphoric reference by Mansfield in her story tends to become an obstacle to surmount during the comprehension process.

Case study

Over many years of teaching, a great number of my graduate and undergraduate students of English have shown a tendency to recurrently identify the female figure behind the quote “**she** sank into it, slowly, leaning **her** head against the back, **her** arms falling along the sides” with “The American Woman”, the secondary character which opens the third paragraph in the story. This is a repeated, not an isolated, fact.

From a discourse-structure adjacency perspective (Hobbs 2004), the error is comprehensible, because readers' expectations make them give priority to

proximity in their search for an identity that suits the character; and it is also understandable from a relevance point of view (Wilson and Matsui 2000), because readers' expectations are more easily, and more economically, satisfied by filling the pronoun with the closest possible matching referent. Following Giora (1997), however, both operations would be subject to the rapid identification and statement of a topic, conducive to the construction of a coherent mental model. The analysis of the situation under study actually proves, however, that local interpretive errors, such as the error in reference assignment described here, may trigger patterns of coherence which do not match the author's planned topic in the receiver's mind.

In my teaching experience, there are a significant number of students who recurrently interpret Mansfield's story along the 19th century literary pattern of the "Innocent American abroad", a topic they have studied in their literature classes. According to the imported pattern, both protagonists, Mr. And Mrs. Salesby, are a mixed-nationality couple (i.e. The English Man (Without a Temperament) and the American Woman of the story), and the conflict between them, their mixture of loving nearness and ice-cold aloofness, comes from prototypical socio-cultural differences. In this way, students construct a coherent mental model drawing inferences from a pre-acquired well-structured coherent model which helps them to solve the difficulties implicit in the story. But contrary to these students' interpretive and literary inferences, the problems presented in "The Man Without a Temperament" have nothing to do with cultural shock, or extravagant social behavior. In Mansfield's story, death is central, but it does not follow improper social behavior, as occurs, for instance, in Henry James' *Daisy Miller*, a story with which these students are familiar. In "The Man Without a Temperament", death is the personal tragedy of a woman, Mrs. Salesby, who is critically ill, and of her husband, who, being young and healthy, finds himself dramatically trapped in it.

The fact that, year after year, there are students who, disregarding the flow of information received, import a whole coherent cognitive pattern external to the discourse to satisfy their interpretive purpose, seems to imply that building a coherent mental representation of the story prevails over the need to unfold and grasp the actual topic of the story. This may indicate that: first, the search for coherence is a pre-requisite of interpretation processes, not a consequence; second, a dissociation between cognitive coherence constructs and structural coherence networks appears to cause problems in comprehension; third, coherence patterns may be borrowed from previously-known cognitive domains as a comprehension strategy in an attempt to repair local structural problems; and fourth, once imported, the external coherence patterns may be superimposed on the discourse model with large-scale cognitive effects. Our research has shown that the divorce between linguistic antecedents, frame-oriented associations, and readers' assumptions

may lead to filling up empty coherence spaces with coherent patterns obtained from previous knowledge or other discourse models.

Our research confirms that comprehension is, in fact, guided by the principles of early topic identification (Giora 1997), adjacency (Hobbs 2005) and relevance (Wilson and Matsui 2000); but our evidence has also made it clear that following these steps may not be sufficient for good comprehension results. The case of error in reference assignment, which has been studied here, has revealed that the search for relevance stops when a coherent mental representation of the discourse content (be it a complex mental model or a mental image) is constructed. Our research has also shown that, when there are failures in the identification of the structural components, readers may import external patterns of coherence as a comprehension-repair strategy to satisfy their need for a coherent cognitive representation. Once established, these patterns may prevail over the information received, with large scale consequences for the global reading process. This combination of factors recommends stressing the interface between cognitive coherence and structural coherence (Pasquier and Chaib-draa 2005) and their mutual role in the comprehension process.

Notes

1. Research for this article has been financed by the Junta de Castilla y León Grant SA012A09.

References

- Giora, R. 1997. Discourse coherence and theory of relevance: Stumbling blocks in search of a unified theory. *Journal of Pragmatics* 27, 17-34.
- Hobbs, J.R. 2004. Abduction in Natural Language Understanding. In Horn, L., Ward, G. (eds.), *Handbook of Pragmatics*. Oxford: Blackwell.
- Holmes, M.S. Literature Annotations. The Man Without a Temperament. <http://litmed.med.nyu.edu/Annotation?action=view&annid=1408>. 27 April 2010.
- Mansfield, K. 1981. The Man Without a Temperament. In *The Collected Stories of Katherine Mansfield*. Harmondsworth: Penguin Books. 129-143.
- Pasquier, P.,B. Chaib-draa, 2005. Agent Communication Pragmatics: The Cognitive Coherence Approach. *Cognitive Systems Research*, Volume 6, No 4: 364-395.
- Wilson, D., Matsui, T. 2000. Recent Approaches to Bridging: Truth, Coherence, Relevance. *Lengua, Discurso, Texto*. Bustos Tovar, J.J. et al. (ed.) Madrid: Visor Libros. 103-31.

The impact of order and aspect in processing of causally linked sentences

Joanna Blochowiak¹, Jacques Moeschler¹, Thomas Castelain²

¹Department of Linguistics, University of Geneva, Switzerland

²L2C2, Institute of Cognitive Science, Lyon, France (now University of Costa Rica)

Abstract

This paper presents experimental studies that investigate two issues related to the expression of causality in French: (i) what is the impact of order (*cause-consequence* vs. *consequence-cause*) in the processing of causally linked sentences without connectives and (ii) how, if at all, the aspectual distinction influences the nature of causal relations and their processing. Our hypothesis is that the *consequence-cause* order is processed faster, as it is the order imposed by the paradigmatic causal connective *parce que* (*because*). The differences in reading time confirmed our hypothesis for *weakly* associated causes and consequences. The experiments also showed that the aspectual contrast between *events* and *states* affects the nature of causal relation itself and plays a considerable role in the processing of causality.

Key words: causality, order, aspect, event, state

Causal discourse: order and aspect

The starting point of our experimental investigation consists of a rather simple observation. Most languages describe causal relations with the help of a paradigmatic causal connective, like *parce que* (*because*) in French, in *consequence-cause* order only. For instance, when we hear the sentence “John fell because Mary pushed him” we understand that John’s falling is the consequence of Mary’s pushing him. Nevertheless, the *cause-consequence* order, which matches the occurrence of events in reality, is also available in natural language with or without connectives. The following sentence with the connective *because* “The neighbours are home because the lights are on” involves the *cause-consequence* order but, interestingly, does not have a canonical causal interpretation. It is a typical example of so called *inferential* (Moeschler 1989) or *epistemic* (Sweetser 1990) reading in which the speaker makes a causal guess on the basis of the available evidence. So, it appears that the *consequence-cause* order is predominant in association with the canonical causal uses of connectives like *because*. The question is whether it is just an arbitrary feature of the language or if there is a cognitive motivation behind it.

Also, in discourse semantics (Asher & Lascarides 2003, farther SDRT), where the speaker must infer a relation to hold between sentences without connectives in order for a discourse to make sense, the two causal orders

give rise to two distinct discourse relations. The *cause-consequence* order is identified with a cause seen as a result ('Result' in SDRT terms) whereas the *consequence-cause* order gives rise to the causal inference ('Explanation', in SDRT terms). Thus, the existence of two distinct discourse relations might indicate two different processes of understanding. One recent study (Rohde & Horton, 2010) which showed that discourse relations are subject to the expectation-driven processing, suggests that this hypothesis is on the right track.

Contrary to some earlier studies on causal reasoning that provide evidence for the subjects' preference for the reasoning from cause to effect (consequence in our terms) (Ahn & Nosek 1998, Waldmann 2000 and 2001), our hypothesis is that the *consequence-cause* order has some cognitive advantage and, thus, will be processed faster as opposed to the *cause-consequence* order.

The second point of our experimental study aims at verifying if different types of eventualities participating in causal relations (here we limit ourselves to the opposition *events* vs. *states*) manifest themselves in differences in their causal power that would be perceptible in causal processing. Theoretically, the differences between *events* and *states* can be quite well captured in terms of aspect (Vendler 1957). In their time dimension, *events* are approximately point-like entities which are dynamic and have a culmination point whereas *states* are non dynamic, homogenous (i.e. every part of a given state basically resembles to any other part) and extended in time. It seems that these aspectual differences also affect the nature of the causal relation itself. If I say that "I have a headache because the atmospheric pressure is low", we understand that I have a headache *as long as* the atmospheric pressure is low, which means that one *state* causally support the other while in a simple *event-to-event* causality one *event* works like a causal trigger for another *event* (Blochowiak 2010). The question related to this fact is whether this ontologically fundamental contrast has an impact on the nature of causal relations and whether these differences play a role in the processing of causality.

Experimental design

In order to test the two hypotheses we designed two experiments. The first one, which was conducted in collaboration with the L2C2 (ISC, Lyon), tested the *event-condition* for both orders: *cause-consequence* and *consequence-cause*. The second one, carried out at the University of Geneva, tested the *state-condition* for both orders. Three similar phases can be distinguished for both experiments.

In the first phase, for each experiment a list of about 40 sentences was created describing the situations involving *events* (e.g.: "Mary pushed John") for the *event-condition* and the situations involving *states* (e.g.: "John is

sick”) for the *state-condition*. The examples were chosen in such a way that the situation described by each sentence could be treated as a cause or as a consequence in a causal chain. For instance, “John is sick” can be a cause for “John doesn’t go to school” or it can be a consequence for “John has eaten too much”.

In the second phase, these lists of sentences were presented to participants as stimuli for answers. Each participant received a list of sentences with one of two following instructions: *Please, provide a cause for the following sentences* or *Please, provide a consequence for the following sentences*. In this way we obtained two sentence sequences corresponding to the *cause-consequence* order (e.g.: “Paul took his medicine. He is better now.”) and *consequence-cause* order (e.g.: “Paul took his medicine. He was ill.”).

In so constructed data set we observed that for some stimuli the same causes and consequences were frequently given by different participants and for some others not. To give a more fine-grained analysis of the resulting pattern we created the parameter *association force* which measures how similar the responses to one particular sentence are. The *association force* received the value *strong* if for the same sentence participants gave more than 50% of the same responses (we allowed certain variations in wording); otherwise its value was considered *weak*. For instance, if for the sentence “John is sick”, more than 50% of participants gave the response “He has eaten too much” as cause then the force between “John is sick” and “He has eaten too much” was judged *strong* within the test ‘cause’. Taking the association force parameter into account, we built up a list of sequences of two sentences *weakly* and *strongly* associated in both *cause-consequence* and *consequence-cause* orders for both *event-* and *state- conditions*.

In the third phase, the four resulting combinations of 2-sentence sequences (*consequence-cause strong*, *consequence-cause weak*, *cause-consequence strong*, *cause-consequence weak*) served as the input to reading time tests performed with E-Prime.

Results and conclusions

The ANOVA revealed a main effect of *association force* ($p < .01$) and that the factor *order* does intervene in a significant way with the *association force* ($p < .05$). For *event-condition*, there is no statistically significant difference between the reading times in two orders when the association force is *strong* whereas when the association force is *weak* the sentences are read faster in *consequence-cause* order.

A different picture arises from the preliminary results for the *state-condition*. In *weak* associations the sequences are read much more slowly in the *consequence-cause* order than in the *cause-consequence* order. Regarding the *strong* associations, there is no significant difference in reading time between the two orders.

The results for the *event-condition* partially confirm our hypothesis of faster processing of the *consequence-cause* order (the *consequence-cause* order is read faster in *weak* associations). The results for the *state-condition* pinpoint the importance of aspectual differences. *States* which have fundamentally different aspectual properties from *events* also have a different impact on the cognitive processing of causality.

Acknowledgements

The authors would like to thank all the researchers from L2C2 (ISC, Lyon) who took part in the first experiment (*event-condition*): Coralie Chevalier, Jean-Baptiste Van der Henst and Isabelle Tapiro.

References

- Ahn, W.K., Nosek B.A. 1998. Heuristics used in reasoning with multiple causes and effects. Proceedings of the 20th Annual Conference of the Cognitive Science Society, Mahwah, Erlbaum Associates, 24-29.
- Asher, N., Lascarides A. 2003. Logic of Conversation. Cambridge. Cambridge University Press.
- Blochowiak, J. 2010. La relation causale et ses relata. NCLF 29, to appear.
- Moeschler, J. 1989. Modélisation du dialogue: représentation de l'inférence argumentative, Paris, Hermès.
- Moeschler, J. 2000. L'ordre temporel est-il naturel?. In Moeschler J. and Béguelin, M.J. (eds.), Référence temporelle et nominale, Berne, Peter Lang, 71-105.
- Moeschler, J. 2003. L'expression de la causalité en français. CLF 25, 11-42.
- Moeschler, J. et al. 2006. Le raisonnement causal: de la pragmatique du discours à la pragmatique expérimentale, NCLF 27, 241-262.
- Sweetser, E. 1990. From Etymology to Pragmatics. Metaphorical and Cultural Aspects of Semantic Structure. Cambridge. Cambridge University Press.
- Vendler, Z. 1957. Verbs and Times. Reprint in Z. Vendler, Linguistics in Philosophy, Ithaca: Cornell University Press, 1967, pp. 97-121.
- Waldman, M.R. 2000. Competition among causes but not effects in predictive and diagnostic learning. Journal of Experimental Psychology: Learning, Memory and Cognition 26(1), 53-76.
- Waldman, M.R. 2001. Predictive versus diagnostic causal learning: evidence from an overshadowing paradigm. Psychonomic Bulletin & Review 8, 600-608.

The prosody of the TV news speaking style in Brazilian Portuguese

Luciana Castro, Ben Serridge, João Moraes, Myrian Freitas
Federal University of Rio de Janeiro, Brazil

Abstract

This study characterizes the prosodic structure of the TV news speaking style in Brazil and compares it to the speech of interview subjects on a television talk show. Fifteen distinct metrics, designed to characterize both temporal and melodic characteristics of speech, were evaluated on the two speaking styles. The results of the analysis show that the TV news speaking style is characterized by a lack of filled pauses, a fast speaking rate, few and short silent pauses, and a high percentage of dynamic tones. However, the only metric for which the difference in speaking styles was shown to be statistically significant was the intra-vocalic mean melodic movement.

Key words: prosody, TV news, speaking styles, professional speech.

Introduction

TV news is a professional activity that relies heavily on oral communication. Previous studies suggest the existence of specific melodic lines that characterize this professional speaking style (Fónagy, 1976; León, 1993). This study attempts to quantify the prosodic variations of the speech of Brazilian TV news anchors.

Method

The data used in this study consist of ten minutes of labelled speech, one minute each from five TV news anchors and five interview subjects. All speech was recorded directly from Brazilian television. The interview speech consists of responses given during a television talk show, and was included in this study to allow for comparison of professional read speech and non-professional spontaneous speech. The speakers, all native speakers of Brazilian Portuguese with no apparent communication disorders, are all male between the ages of 46 and 68. Details regarding the recordings, speech data captured and transcription of the speech are described in Castro et al. (2010a).

A perception test using the same corpus demonstrated that listeners are able to distinguish between these two speaking styles with 90% accuracy, even when semantic and lexical information is removed from the signal by a low-pass filter (Castro et al., 2010b). This suggests that there are sufficient

differences in the temporal characteristics and in the fundamental frequency contour to allow listeners to distinguish these speaking styles. If so, it should be possible to define metrics that capture these differences.

As a starting point for the metrics described below, the data were analyzed based on the segmentation of the corpus into phonetic sequences and pauses. Praat was used to calculate the fundamental frequency for individual frames of speech, spaced at intervals of 10 msec (Boersma & Weenik, 2008).

The metrics used to quantify the temporal characteristics of speech were: number of silent and filled pauses per minute; average pause duration; total pause time; average phonetic sequence duration; number of syllables per phonetic sequence; and speaking rate (Freitas, 1992). To characterize the fundamental frequency contour of the TV news speaking style, the metrics used in this study were: the average and standard deviation of the fundamental frequency; the interquantile range, which is the range measured after excluding the highest and lowest 5% of F0 values; the percentage of static, rising, or falling tones (calculated only for stressed vowels); and the intra-vocalic mean melodic movement, expressed in semitones per second, which measures the sum of the absolute value of changes in fundamental frequency from one frame to the next, considering only the values within vowels (Goldman et al., 2008).

Finally, an attempt was also made to quantify the occurrence of pitch reset, which was calculated, for each pause, as the difference between the average F0 of the first stressed vowel after the pause and the average F0 of the last stressed vowel before the pause (Heldner et al, 2004). Based on this definition, two metrics were calculated: the percentage of pauses that exhibit a pitch reset of more than two semitones, and the pitch reset averaged over all pauses

Results

A striking characteristic of the TV news speaking style is the complete absence of filled pauses. In the five minutes of TV news speech, there was not a single filled pause, compared to 19 in the interview speech. Regarding silent pauses, in this corpus there were 19.8 silent pauses per minute of speech, with an average duration of 327 ms, for a total pause time of 11% of the recorded speech. The average duration of the phonetic sequences in the TV news speech was 2306 milliseconds, with 15.6 syllables per phonetic sequence and a speaking rate of 6.0 syllables per second. As a point of comparison, Uglova & Shevchenko (2006) observed a slower rate of speech (5.1 sil / sec) in English TV news broadcast.

To put these results in perspective, the same temporal metrics were evaluated on the speech of interview subjects, and a Student's t-test was then performed on each metric. Table 1 shows the results of the comparison.

Surprisingly, the temporal characteristics of professional TV news speech do not differ significantly from those of spontaneous interview speech. As expected, the TV news speaking style exhibits slightly fewer silent pauses per minute, and these pauses are slightly shorter, but the speaking rate is approximately the same and the differences observed in pause time are not statistically significant.

Table 1: Results and p-value of the Student's t-test of the temporal metrics.

Metric	TV news	Interview	p-value
number of silent pauses per minute	19.8	21.70	0.49
average pause duration (ms)	327	387	0.16
total pause time (%)	11	14	0.09
average phonetic sequence dur (ms)	2306	2216	0.56
number of syl per phonetic sequence	15.6	15.8	0.76
speaking rate (syllables / second)	6.0	6.1	0.80

The results of the metrics related to the fundamental frequency contour are presented in Table 2. As before, a Student's t-test was applied to compare the two speaking styles. Note that the only measure that differs in a statistically significant way between the two speaking styles was the intra-vocalic mean melodic movement. This may indicate that the professional speakers make more use of short-range pitch movements, although other related metrics, such as the percentage of rising or falling tones, do not reinforce this conclusion with statistical significance.

Table 2: Results and p-value of the Student's t-test of the F0 metrics.

Metric	TV news	Interview	p-value
mean fundamental frequency (Hz)	139	155	0.35
standard deviation of F0 (ST)	4.3	4.0	0.41
interquantile range (ST)	14.1	12.2	0.10
mean melodic movement (ST/seg)	36.3	30.7	0.04
percentage of pauses with pitch reset > 2 ST	55	42	0.07
average pitch reset (ST)	2.6	3.1	0.72
percentage of static tones (%)	46	54	0.24
percentage of rising tones (%)	26	25	0.79
percentage of falling tones (%)	28	21	0.34

Conclusion

In summary, the results of the analysis presented here show that the TV news speaking style is most distinctly characterized by a lack of filled

pauses. It is relatively fast speech, with few and short silent pauses, and a high percentage of dynamic tones. When compared with the speech of talk show interview subjects, the metrics defined in this study are not capable of differentiating the two speaking styles. Of the 15 prosodic metrics evaluated in this study, the only one that exhibited a statistically significant difference was the intra-vocalic mean melodic movement. However, in a perception test based on the same corpus, listeners were able to distinguish the two speaking styles from low-pass filtered speech with an accuracy of 90%, suggesting that there must be prosodic differences between the two speaking styles that are not captured in the metrics proposed here. Perhaps the relevant prosodic markers are intermittent and therefore diluted in the global metrics commonly proposed in the literature? Further research is required to develop more sophisticated metrics that capture the perceptually relevant aspects of prosody that listeners depend upon to identify the characteristics of professional speaking styles.

References

- Boersma, P., Weenink, D. PRAAT: doing phonetics by computer (5.0.21). Available at: <<http://www.praat.org/>>. Accessed: April, 23rd.
- Castro, L., Serridge, B., Moraes, J., Freitas, M. 2010a. Characterizing variation in fundamental frequency contours of professional speaking styles. In: *Proceeding of Speech Prosody*, Chicago.
- Castro, L., Serridge, B., Moraes, J., Freitas, M. 2010b. Listeners' Ability to identify professional speaking styles based on prosodic cues. In: *Proceeding of Speech Prosody*, Chicago.
- Fónagy, J. 1976. Prosodie professionnelle et changements prosodiques. *Le Français Moderne*, XLIV(3): 193-228.
- Freitas, M.J. 1992. Contributo para o estudo de padrões de estruturação temporal da fala no português europeu. In: Pereira, I. (org.) *Estudos em prosódia*. Lisbon: Colibri, 74-103.
- Goldman, J.-Ph., Auchlin, A., Avanzi, M., Simon, A.C. 2008. ProsoReport: an automatic tool for prosodic description. Application to a radio style. In: *Proceedings of Speech Prosody*, Campinas, 701-704.
- Heldner, M., Edlund, J., Björkenstam, T. 2004. Automatically extracted F0 features as acoustic correlates of prosodic boundaries. *Proceedings of FONETIK*, Stockholm, 52-55.
- Léon, P. 1993. Variation situationnelle et voix professionnelles. In: *Précis de phonostylistique*. Paris: Nathan, 157-184.
- Shevchenko, T., Uglova, N. 2006. Timing in news and weather forecasts: implications for perception. *Proceedings of Speech Prosody*, Dresden, 5-8.

Perceptual cues for sentence intonation in Greek

Anthi Chaida

Phonetics Laboratory, University of Athens, Greece

Abstract

This is a study of intonational cues for the perception of statements and polar questions in Greek. A two-part experiment was conducted with synthetic stimuli with manipulated intonation targets in stress groups, from statement to question and vice versa. According to the results, the critical perceptual cue is the boundary tone, which led listeners to identify stimuli as the reverse sentence type (81-100%). However, the perception of intonational contrasts may be based on holistic contours.

Key words: perception, intonation, prosody, sentence types, Greek.

Introduction

The present study investigates the effects of prosody on sentence perception in Greek, i.e. the decisive perceptual cues for statements and polar questions, based on synthetic stimuli, on which the intonational contours of statements are partly and totally replaced by question appropriate contours, and vice-versa. In previous studies natural stimuli of all types were perceived with 99% accuracy (Chaida, 2005). In general, intonation has been shown to contribute to the production and perception of sentence type contrasts in many languages (e.g. Chaida, 2008; Makarova, 2001; Thorsen, 1980).

Materials and method

A set of utterances produced by one female speaker at her twenties was selected from a previously recorded corpus (Chaida, 2007), including the sentence [o ma'nolis ma'zevi lemɔna] (Manolis is picking lemons), produced as statement and polar question. These utterances were processed with Praat. Their pitch was stylized (frequency resolution 1 semitone), and they were manipulated in two ways: (a) the 3 stress groups of the statement were manipulated, one by one, in pairs, and in total, in order to reach the pitch targets of the question (with the same alignment), but maintaining duration and intensity, and vice versa (examples in Fig. 1-2.3); (b) the 3 stress groups of the statement were replaced unedited (carrying duration and intensity), one by one, by the corresponding stress groups of the question, and vice-versa. A set of 14 stimuli was set up by the 1st procedure and a set of 6 synthetic stimuli by the 2nd. The stylized original utterances of a statement and a polar question were included in both sets. All stimuli were randomized 4 times. These modified stimuli were presented to 30 informants

(male-female, 20-40 years old from Athens). The stimuli were presented over headphones in a quiet room, and the informants were instructed to identify each stimulus as either a statement or a question, and then to provide an indication of how certain they were of their decision (1-6 scale: 1=least, 6=most certain).

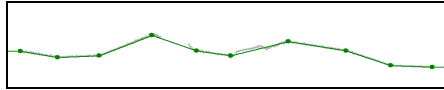


Figure 1. Intonation contour for statement with stylized pitch.

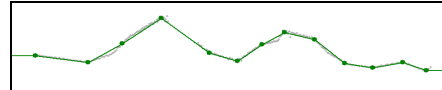


Figure 2. Intonation contour for polar question with stylized pitch.

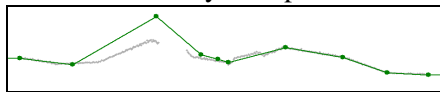


Figure 1.1. Statement with question's tonal onset (1st stress group).

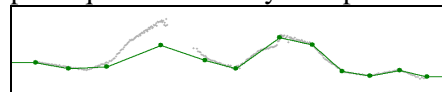


Figure 2.1. Question with statement's tonal onset (1st stress group).

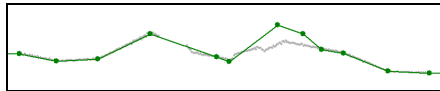


Figure 1.2. Statement with question's medial (2nd) stress group.

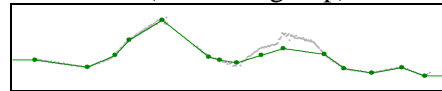


Figure 2.2. Question with statement's medial (2nd) stress group.

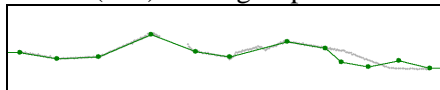


Figure 1.3. Statement with question's tonal offset (3rd stress group).

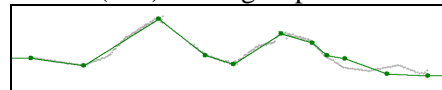


Figure 2.3. Question with statement's tonal offset (3rd stress group).

Results

The original stylized stimuli had very high correct identification rates in total: 98.35% for statements and 99.57% for questions. According to the results of the 1st part of the experiment, statements had high correct identification scores (perceived as statements) for stimuli with manipulation on the 1st stress group (99.16%), on the 1st and 2nd (100%), and on the 2nd one (96.66%). However, when the boundary tone was manipulated, statements were identified as questions: 80% for the 1st and 3rd stress group, 81.66% for the 2nd and 3rd, and 81.66% for the 3rd stress group. In the cases where the intonation contour of the statement had been totally replaced by the question, listeners mostly identified statements as questions (80%). The results for questions were similar, with high correct identification scores for stimuli with manipulation on the 1st stress group (99.16%), on the 1st and 2nd (99.16%), and on the 2nd one (99.16%). Here again, when listeners were presented stimuli with manipulation on the 3rd stress group (boundary), they identified questions as statements: 98.33% for the 1st and 3rd stress group, 98.33 % for the 2nd and 3rd, and 100% for the 3rd stress group. In

the cases where the intonation contour of the question had been totally replaced by the statement one, listeners mostly identified questions as statements (99.16 %).

Certainty ratings were fairly high: 5.21 (mean) out of 6 for statements, and 4.78 for questions. These ratings were statistically significant for questions ($F(1,872)=83.72$, $p<0.01$), but not for statements ($F(1,1044)=0.41$, $p>0.01$). Listeners seemed to be most certain when they answered for stimuli with manipulations on the whole of the intonation contour (5.32 out of 6), and least certain for stimuli with combined manipulations on the 1st and 3rd stress group (4.65).

According to the results of the 2nd part of the experiment (stimuli with replaced unedited prosodic parts), statements had high identification score (perceived as statements) in stimuli with manipulation on the 1st stress group (99.16%), and on the 2nd one (100%), while stimuli with the 3rd stress group replaced (boundary tone) were noticeably identified as questions (97.50%). Questions also had high identification scores, i.e. they were perceived as questions with manipulation on the 1st stress group (100%), and on the 2nd one (100%), and as statements with manipulation on the 3rd stress group (100%).

Certainty ratings were fairly high in general: 5.47 (mean) out of 6 for statements, and 5.38 for questions. These ratings were statistically significant for questions ($F(1,474)=30.87$, $p<0.01$), but not for statements ($F(1,482)=8.73$, $p>0.01$). Listeners seemed to be most certain when they responded to stimuli with manipulations on the 1st stress group (5.55 out of 6), and least certain for stimuli with manipulations on the 3rd stress group (5.16), but the range was very short in all cases.

In both parts of the experiment, the identification results were statistically significant (chi-square test, $p<0.01$) and the listeners' choices were not incidental, since their answers were in accordance with the initial sentence types and stress group manipulations at the 99% confidence level.

Discussion

Several studies have shown that each sentence type has an acoustically distinct tonal structure, characterized in Greek by the type and location of the tonal nucleus and the boundary tone (Chaida, 2005, 2007; Botinis Bannert, Tatham., 2000). Moreover, experiments with artificial hum stimuli without any lexical information showed that prosody is a major cue for the perception of statements and questions in Greek (Chaida, 2008).

The discrete intonational contours of sentence types are transparently distinguished by native speakers. There is a clear 1 to 1 production-perception correspondence. However, the crucial question of the present study is: what are the decisive intonational cues for the perceptual

differentiation of statements and polar questions in Greek? The results indicate that sentence type perception is mostly based on the realization of the boundary tone. In both parts of the experiment, the listeners identified statements as questions, and questions as statements. These findings agree in general with earlier studies demonstrating that in many languages low or falling boundary tones elicit 'declarative judgments', and high or rising tones lead to 'interrogative judgments' (e.g. Baltazani, 2007; Makarova, 2001; Thorsen, 1980). Moreover, it has been found that, although in general tonal cues for sentence types are distributed over the whole utterance, the last stressed syllable and the following unstressed ones are sufficient indicators for sentence type perception (Thorsen, 1980). It has been suggested that the entire intonation contour, rather than timing and scaling of turning points, can have impact on perception; i.e. perception of intonational contrasts is mostly based on holistic contours, rather than on local characteristics (Knight, 2008; Niebuhr, 2003). Hence, although the boundary tone (tonal offset) has been found to be a major perceptual cue for listeners, its effect may be overridden by earlier tonal movements, confirming the importance of the entire contour for perception.

References

- Baltazani, M. 2007. Intonation of polar questions and the location of nuclear stress in Greek. In: Gussenhoven, C., Riad, T. (eds.), *Tones and tunes-II. Experimental Studies in Word and Sentence Prosody*. Berlin: Mouton de Gruyter, 387-405.
- Botinis, A., Bannert, R., Tatham, M. 2000. Contrastive tonal analysis of focus perception in Greek and Swedish. In: Botinis, A. (ed), *Intonation: Analysis, Modelling and Technology*. Dordrecht: Kluwer Academic Publishers, 97-116.
- Chaida, A. 2005. Intonation of sentence types and focus in Greek: production and perception. MA Thesis, University of Skövde, Sweden.
- Chaida, A. 2007. Tonal structures of complex sentences in Greek. *Proc. of the 8th International Conference on Greek Linguistics*. Ioannina, Greece.
- Chaida, A. 2008. Prosodic perception of sentence types in Greek. *2nd ISCA Workshop on Experimental Linguistics*, Athens, Greece, 61-64.
- Knight, R. 2008. The Shape of nuclear falls and their effect on the perception of pitch and prominence. *Language and Speech*, 51(3): 223 - 244.
- Makarova, V. 2001. Perceptual correlates of sentence-type intonation in Russian and Japanese. *Journal of Phonetics*, 29, 137-154.
- Niebuhr, O. 2003. Perceptual study of timing variables in F0 peaks. *Proceedings of the 15th ICPHS*, Barcelona, Spain, 1225-1228.
- Thorsen N. 1980. A study of the perception of sentence intonation - evidence from Danish. *Journal of the Acoustical Society of America*, 67, 1014-1030.

Text punctuation and prosody in Greek

Anthi Chaida, Elina Nirgianaki, Eleni Panaretou
Department of Linguistics, University of Athens, Greece

Abstract

A production experiment was carried out, in order to investigate text punctuation, including standard as well as ungrammatical (communicative) punctuation marks, and prosody relations. It is shown that punctuation is directly related to the duration of pauses, leading to the following structure: question mark>exclamation mark>full stop> colon>comma> ellipsis. Pitch resetting occurs in all cases, except for commas and ellipses, and gender differences seem to be significant. Ungrammatical punctuation is more prominently marked by both temporal and tonal means.

Key words: prosody, discourse, punctuation, pauses, F0, Greek

Introduction

The present study concentrates on the relation between text punctuation and prosody in Greek. The following questions are addressed: (i) Which are the relations between prosody and text punctuation?, (ii) what are the effects of pause productions?, (iii) How does prosody interact with grammatical and ungrammatical (communicative) punctuation? The aim of the study is to provide evidence concerning the prosodic interpretation of punctuation. A variety of punctuation marks were examined (comma, full stop, colon, ellipsis, question mark, exclamation mark), as well as their ungrammatical uses, based on oral readings of a newspaper's opinion article by 10 speakers.

Given that in written language, especially in Greek, the use of punctuation is not always subject to strict rules, texts are as a rule punctuated according to two basic principles: (i) a semantico-syntactic, according to which punctuation reflects syntactic relations and affects the meaning of the sentence, and (ii) a communicative, according to which punctuation emphasizes important information included in the text. The latter function is referred to as ungrammatical punctuation, i.e. the use of a specific punctuation mark in cases where a grammatical rule prohibits it (e.g. a comma before the conjunction "and") (Panaretou, 1995, 2002).

The relation between spoken "intonation units" and "punctuation units" is often mentioned in the bibliography (Chafe, 1987; Panaretou, 1995, 2002; Nirgianaki, 2005). Significant work has also been carried out on discourse structure and its prosodic correlates (e.g. Botinis, 1991; Botinis et al., 2004, 2007; Witchmann, 2000). However, the relation between text punctuation and prosody is in general underexamined, particularly in Greek.

Methodology

A production experiment was carried out, based on a political-opinion article written in a weekly Greek newspaper. It is a text of 402 words and its duration is around 2.5 minutes. It consists of 14 paragraphs, which include: 27 full stops, 43 commas, 4 question marks, 6 exclamation marks, 2 colons and 1 ellipsis, from which 15 full stops, 1 question mark, 1 exclamation mark where ungrammatical (communicative punctuation).

Ten speakers (5 male and 5 female, 20-30 years old, native speakers of standard Athenian Greek) read aloud the text in normal tempo. The recordings were made in the recording booth of the Phonetics Laboratory of the University of Athens (with a dynamic Samson S11 microphone, 16bit mono, 22050 Hz).

Measurements were taken for pauses, intra-pause text duration, and pre-pause and post-pause pitch. Data were organized in MS Excel, the acoustic analysis was carried out with Praat, and the statistical processing with StatView.

Results

According to the results of the production experiment, 97.78% of the pauses were produced in accordance with written punctuation, whereas 2.22% of the pauses were produced within the text reading without any punctuation mark. The longest pauses (Figure 1) were produced after question marks (735 ms) and the shortest after ellipses (279 ms), whereas pauses were also produced within the text with no punctuation mark (237 ms). It should be noted that, in all cases, pauses produced after ungrammatical (stylistic) punctuation were longer (464 ms) than the ones after grammatical punctuation (527 ms), with a significant difference ($t(df)=636$, $p<0.01$). Pauses produced at the end of paragraphs, as noted by the first author, were significantly longer (686 ms) than pauses produced within the text (408 ms) ($t(df)=636$, $p<0.01$).

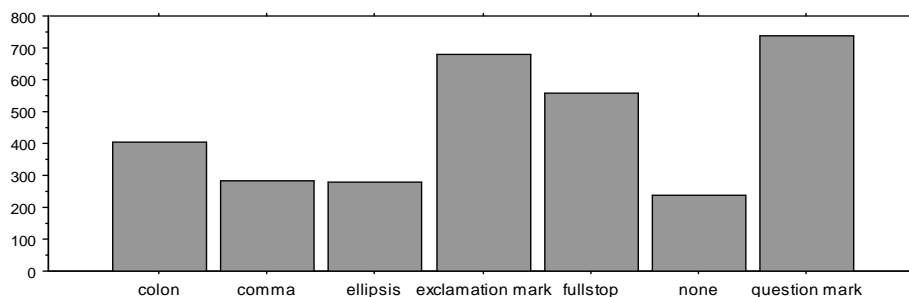


Figure 1. Mean pause durations (ms) with reference to punctuation marks.

No correlation was found between intra-pause text duration and pause durations for all punctuation marks ($r < 0.7$). The longest intra-pause text duration was noted in relation to question marks (2100 ms) and the shortest one in relation to ellipsis (543 ms).

Mean pre-pause pitch (152 Hz) was in most cases lower than post-pause pitch (171 Hz), with a significant difference ($t(df)=636$, $p < 0.01$), denoting resetting (Figure 2). However, this is not the case for commas, ellipses, and pauses produced without any punctuation mark noted, where pre-pause pitch was higher than post-pause pitch. Noticeably, ungrammatical punctuation was marked with significantly more prominent pre-post pause pitch differences ($t(df)=173$, $p < 0.01$), in comparison to grammatical punctuation ($t(df)=416$, $p < 0.01$) (Figure 3).

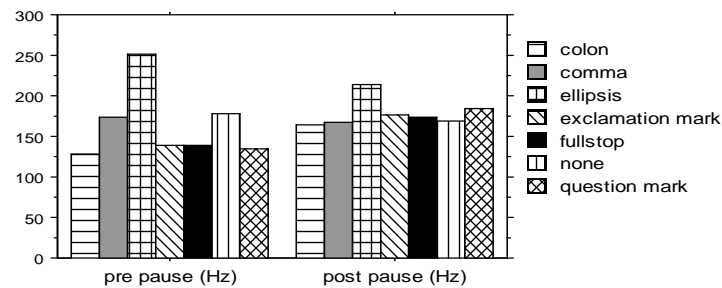


Figure 2. Mean pre-pause and post-pause pitch (Hz) with regards to punctuation marks.

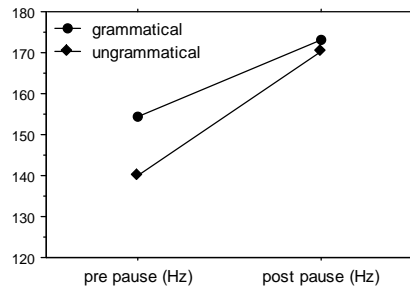


Figure 3. Mean pre-pause and post-pause pitch (Hz), as a function of grammatical and ungrammatical (stylistic) punctuation.

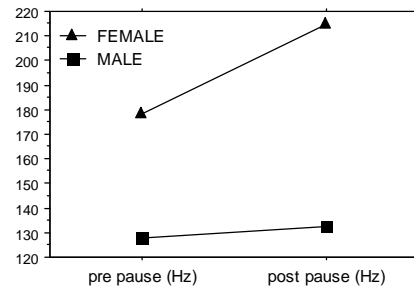


Figure 4. Mean pre-pause and post-pause pitch production (Hz), as a function of female and male gender.

Gender differences regarding pause durations in accordance with punctuation were not significant ($F(6,623)=0.562$, $p > 0.05$). However, gender differences regarding pitch resetting (pre-post pause pitch) were found to be highly significant ($t(df)=636$, $p < 0.01$), (Figure 4).

Conclusions

To recapitulate, the results of the present experiment indicate: (i) punctuation is directly related to the average duration of pauses. The structure based on pause durations is the following: question mark > exclamation mark > full stop > colon > comma > ellipsis, (ii) produced pauses were longer after ungrammatical (communicative) punctuation marks.

Pitch resetting was evident, with regards to all punctuation marks, except for commas and ellipses, whereas resetting was more evident for ungrammatical punctuation.

Ungrammatical punctuation, in general, seems to be most prominently marked by all prosodic means examined. This finding reinforces the view that ungrammatical punctuation highlights text information.

Acknowledgements

The present study has been supported by the University of Athens in the framework of the programme “Kapodistrias”, project “Text prosody in Greek”. Many thanks to Georgia Korkou and Eleni Vazoura for the collection of data, to Antonis Botinis and Marios Fourakis for their comments, and to all the participants of the experiment.

References

- Botinis, A. 1991. Intonation patterns in Greek Discourse. 12th International Congress of Phonetics Sciences, 286-289, Aix-en-Provence, France.
- Botinis, A., Bakakou-Orphanou, A., Tafroglou, H., Chistopoulou, A. 2004. Produced pauses, perceived pauses and thematic units. XVIIth Swedish Phonetics Conference, 112-115, Stockholm, Sweden.
- Botinis, A., Bakakou-Orphanou, A., Themistocleous, Ch. 2007. Multifactor analysis of discourse turn in Greek. 16th International Congress of Phonetic Sciences, 1341-44, Saarbrücken, Germany.
- Chafe, W. 1987. What good is punctuation? Occasional Paper No. 2. Berkeley, CA: Center for the Study of Writing.
- Nirgianaki, N. 2005. Prosodic production of text punctuation in Greek. 7th International Conference on Greek Linguistics, York, UK.
- Panaretou, E. 1995. Discourse functions of punctuation. A study on literary text punctuation (in Greek). PhD Thesis. University of Athens, Greece.
- Panaretou, E. 2002. Textual functions of punctuation in literary texts. Problems and questions on text editing. Athens, 130-139.
- Wichmann, A. 2000. Intonation in text and discourse: Beginnings, middles and ends. London: Longman.

Coarticulation of consonant clusters in Italian learners of French

Sonia d'Apolito

Department of Foreign Languages, University of Salento, CRIL (Centro di Ricerche Interdisciplinare sul Linguaggio), Italy.

Abstract

The present study focuses on how Italian learners of French as L2 coarticulate French heterosyllabic sibilant sequences and how speech rate influences the production of these non-native sequences. As in Italian consonant sequences are less frequent than in French and very rarely appear across word boundary, the hypothesis is that a different degree of coarticulation is found in clusters produced by French, as reported in the literature, and by Italian learners of French. First results appear to confirm our expectations. A small degree of coarticulation characterises clusters produced by Italian, especially in comparison to what has been observed for French natives even if at a fast speech rate some cases of place assimilation are observed.

Key words: place assimilation, speech rate, L2.

Introduction

Since in Italian language consonant sequences across word boundary appear in case of prepositions or loanwords Farnetani, Busà, (2004); Muliačić, (1973), it is interesting to observe which degree of coarticulation Italian learners can realize when they pronounce non-native sequences. Sibilant clusters are studied along the lines of a recent study on place assimilation in French by Niebuhr et al. (2008). In their study, the authors investigated sequences of alveolar and postalveolar sibilants produced across word boundary at a natural speech rate finding evidence for the presence of place assimilation, although French was claimed not to have it (Walker, 1982).

The study presented here investigates the same sibilant clusters in French productions of Italian learners analysing both acoustic and kinematic data. Moreover, it takes into account speech rate as an important factor since it facilitates coarticulation as a result of an increasing coproduction between successive segments (Byrd, 1996). The hypothesis is that Italian learners' production is characterised by a small degree of coarticulation because of mother tongue's influence. As described in literature Davidson (2006) and Oh (2008), a transitional schwa will be inserted within the clusters keeping them away, even if at a faster speech rate a stronger coarticulation between segments is expected. However, it seems quite improbable the realization of place assimilation by Italian learners of French.

Method

Target heterosyllabic sibilant clusters, /sB/, /Bs/, /sC/, /Cs/, /zB/, /zB/, were studied in a /a_i/ surrounding vowel context, proposed within a carrier phrase. In this study two main factors were considered: prosodic boundary and speech rate. As a prosodic boundary could interfere with the coarticulatory process, stimuli were inserted within a phonological phrase (with no boundary; e.g. “Il a dit sage syrien rapidement” *He said Syrian wise person quickly*), and an intonational phrase boundary (with a prosodic boundary; e.g. “D’abord il a dit sage, syrien l’a dit après” *Before he said Syrian, wise person has been said after*). In order to facilitate the coarticulation of clusters, subjects (two Italian learners of French, SPK1 and SPK2) repeated the corpus seven times, both at a normal and at a faster speech rate.

Acoustic and kinematic (EMMA, Ultrasound) data were collected simultaneously at CRIL laboratory. Here only acoustic data will be discussed. Auditory analysis and acoustic labelling were performed in PRAAT while MATLAB and PRAAT scripts were used for measurements. The auditory analysis aimed at verify cluster and boundary realization. The labelling were related to: 1) segment boundaries of the $V_1C_1\#C_2V_2$ sequence, including final schwa and/or pause; 2) three points at the vowel steady-state (30ms) and a point at the vowel transition in V_1 , V_2 and in the schwa, if present. Acoustic measurements carried out were: duration of all segments; for each vowel, values of first three formants; for each sibilant, mean value and range of the Centre of Gravity (CoG) in Hz, for frequencies between 1.5-15kHz. The acoustic analysis presented below focuses on the latter measurement. Indeed, the CoG represents the average value of the frequencies in the spectrum weighted by the amplitude Gordon et al. (2002). CoG gives good information about the place of articulation of fricatives and it will be useful for distinguishing alveolars from palatals both from an acoustic and perceptive point of view.

Results

Acoustic analysis reveals that, as described in literature, both speakers inserted a transitional schwa between the two segments at both speech rates. Thus the two sibilant on average are kept away even if SPK1 when talking faster realizes a schwa in few cases. Results of CoG values carried on the first/second cluster position are displayed in Figure 1. Two regions are visible in the plot, corresponding to palatals– with low mean CoG values, range between 5 and 6.5 kHz - and alveolars– with higher mean CoG values, range between 6.8 and 9 kHz. Univariate ANOVA shows that the mean CoG value (dependent variable) is significantly different depending on the type of sibilant (independent variable) ($[F(5,606)=2,711\ p=,020]$).

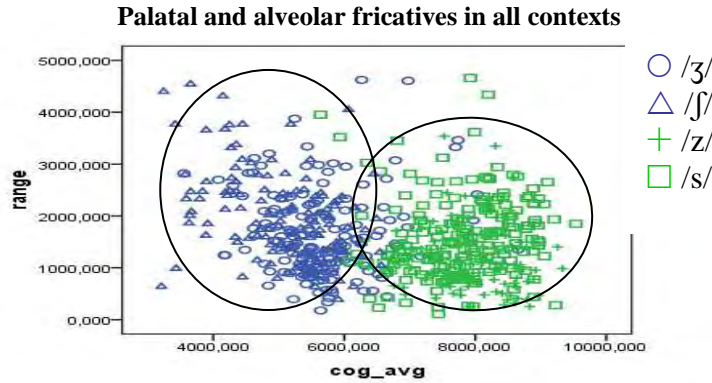


Figure 1: Scatterplot of mean CoG values for each sibilant within all contexts. On y-axis is represented CoG range while on x-axis the mean CoG. Blue indicates palatals (left circle); Green indicates alveolars (right circle).

Individual univariate ANOVAs and post-hoc tests (Tukey) with sibilant type as independent variable and mean CoG as the dependent were carried on each consonant position, prosodic boundary and speech rate. At normal speech rate, schwa insertion took place. In this case the type of consonant, in both positions, plays a significant role in affecting the mean CoG values. Values for the two speakers in both no-boundary and boundary conditions shows that: no boundary: C₁: $F(5,42)=22,93$ $p=,000$ for SPK1; $F(5,41)=64,544$, $p=,000$ for SPK2; C₂: ($F(5,34)=50,84$ $p=,000$ for SPK1; $F(5,40)=22,69$ $p=,000$ for SPK2; boundary conditions: C₁: $F(5,41)=32,96$ $p=,000$ for SPK1; $F(5,40)=112,48$ $p=,000$ for SPK2; C₂: $F(5,39)=78,98$ $p=,000$ for SPK1; $F(5,40)=39,49$ $p=,000$ for SPK2). Post-hoc tests show that mean CoG values of palatals are lower than those of alveolars meaning that alveolar and palatal fricatives are kept away at normal speech rate.

At faster speech rate, schwa insertion did not take place with the exception of some cases. In this case the type of consonant plays again a significant role in affecting CoG values for both speakers in both no-boundary and boundary conditions: no-boundary: C₁: $F(5,24)=4,78$ $p=,006$ for SPK1; $F(5,38)=16,83$ $p=,000$ for SPK2; C₂: $F(5,40)=4,13$ $p=,005$ for SPK1; $F(5,37)=30,43$ $p=,000$ for SPK2; boundary conditions: C₁: $F(5,28)=19,51$ for SPK1; $F(5,33)=106,03$ for SPK2; C₂: $F(5,28)=10,15$ $p=,000$ for SPK1; $F(5,34)=35,32$ $p=,000$. Results of post-hoc tests show that mean CoG values of palatals are lower than those of alveolars, regardless of position fricatives appeared. Thus, the two speakers keep away the two segments.

Apart from the case of /Cs/ and /Bs/ clusters for SPK1's production in which COG values show that some cases of place assimilation are realized.

This observation was consistent with the author's auditory analysis. Univariate ANOVA performed on these clusters shows that the type of consonant is non significant in affecting C2's CoG values [$F(2,61)=.680$ $p=0.51$]. This means that CoG values of alveolars is not significantly different from the mean value of palatals. Moreover, as the CoG mean values range between 5.7 and 6.8 kHz, data also show that the alveolar fricatives were palatalized. However, t-test results on speed rate (total duration of phrase/number of syllables realized ratio) shows that productions of both speakers are significantly different depending on the speech rate (with speed rate values significantly for both speakers SPK1 and SPK2 $p=,000$). Nevertheless the t-test also shows that SPK1's productions are always significantly faster than SPK2's productions $t=4,878$ $p=,000$ for normal and $t=7,369$ $p=,000$ for fast rate). The faster rate may of course have favoured the assimilatory process.

Conclusion

Italian learners' production of French sibilant clusters seems to be affected by mother tongue phonetics and phonology. The two speakers analysed so far realized cluster at normal rate with intervening schwas, both within and across prosodic constituents. On the other hand fewer schwa insertions is observed in the case of faster rate, independently of prosodic position. Preliminary acoustic (centre of gravity) results show that alveolar and palatal fricatives are kept distinct. However, at fast speech rate, some cases of place progressive assimilation are realized by one speaker, with alveolar fricatives showing a mean CoG value similar to preceding palatal fricatives.

References

- Byrd, D., Tan, C.C. 1996. Saying consonant clusters quickly. *Journal of Phonetics* 4, 263-282.
- Davidson, L. 2006. Phonology, phonetics or frequency: influences on the production of non-native sequences. *Journal of Phonetics* 34, 104-137.
- Farnetani, E., Busà, M.G. 2004. Italian clusters in continuous speech. *Proc. of the 3rd ICSLP*, vol. 1, 359-362, Yokohama, Japan.
- Gordon M., Barthmaier P., Sands K. 2002. A cross-linguistic study of voiceless fricatives. *JIPA* vol. 32, 133-139.
- Muliačić, Z. 1973. *Fonologia della lingua italiana*. Ed. Il Mulino, Bologna.
- Niebuhr O., and al.. 2008. On place assimilation in French sibilant sequences. *Proc. of the VII ISSP*, 221-224, Strasbourg, France.
- Oh, E., 2008. Coarticulation in non-native speakers of English and French: an acoustic study. *Journal of phonetics* 36, 361-384.
- Walker, D.C. 1982. On a phonological innovation in French. Ed. Cambridge University Press, vol. 12, 72-77.

The acoustic analysis of speech samples designed for the Voice Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS): long-term f0 and intensity measures

Zuleica Antonia de Camargo^{1,2}, Sandra Madureira¹

¹Integrated Acoustic Analysis and Cognition Laboratory – LIAAC, Pontifical Catholic University of São Paulo, São Paulo, Brazil

²Voice and Speech Laboratory - CEFAC Institute, São Paulo, São Paulo, Brazil

Abstract

The present study aims at describing voice quality settings by means of long-term, f0 and intensity analysis of samples which integrate the material developed with the purpose of introducing the Voice Profile Analysis Scheme-VPAS in the Brazilian Portuguese context (BP-VPAS). The voice quality database was recorded in a radio studio. The audio samples were analyzed by means of the SG Expression Evaluator, a script developed by Barbosa (2009), which automatically extracts fundamental frequency (f0) and its first derivative (df0), intensity, spectral tilt (SpTt) and Long-Term Average Spectrum (LTAS). These measures have been statistically analyzed by means of multiple and non-linear regression tests. The results contemplate the correlations between perceived voice quality settings and f0 and intensity long-term acoustic measures.

Key words: Voice Quality; Phonetics; Speech Acoustics; Auditory Perception; Speech Production Measures

Introduction

Studies focusing voice quality are rather incipient in the prosodic field in spite of its importance to the analysis of both vocal expressiveness and voice disorders.

Furthermore, there are few studies describing the long-term acoustic correlates of the laryngeal, the supralaryngeal and the tension settings which can be identified perceptually with the help of the Voice Profile Analysis Scheme (VPAS) developed by Laver et al (1981), as a phonetically-grounded evaluation of voice quality. The present study addresses these two issues for it aims at describing voice quality settings by means of long-term, f0 and intensity analysis of the audio samples which integrate the instructional material developed by a group of researchers with the purpose of introducing the Voice Profile Analysis Scheme-VPAS in the Brazilian Portuguese context: BP-VPAS (Camargo, Madureira, 2008).

Due to the importance of the susceptibility of the speech segments to the effects of voice quality settings (Laver, 1980), a team of researchers from Queen Margaret University College (QMUC- Edinburgh) engaged in

elaborating an English corpus specially designed for auditory voice quality evaluation by means of VPAS. This procedure has also been adopted by the group of researchers at LIAAC-PUCSP (Integrated Acoustic Analysis and Cognition Laboratory of the Pontifical Catholic University of São Paulo) concerned with the adaptation of VPAS to the Brazilian Portuguese context so as to improve the material aimed at instructional purposes.

The description of voice quality settings by means of long-term acoustic analysis techniques have been proved relevant (Hammarberg, Gauffin, 1991; Cukier-Blaj et al, 2008).

The detailed description of long-term acoustic correlates of the perceived voice quality settings is not an easy task, but it can provide a detailed description of events related to voice quality settings in laryngeal, supralaryngeal and tension domains, as well as for voice dynamics description. This paper addresses this demand, applying a method for speech expressiveness change detection developed by Barbosa (2009) in order to describe samples from a group of adults without any complaints about their voices.

Methods

The building of the corpus took into account the principle of susceptibility proposed by Laver (1980) and made use of the keyspeech segments (Camargo, Madureira, 2008). The sentences that constitute the corpus were recorded in a radio studio and retrieved in the voice database of LIAAC. The building of the corpus took into account the principle of susceptibility proposed by Laver (1980) and made use of the keyspeech segments (Camargo, Madureira, 2008). It was recorded by 38 subjects, aged from 20 to 58 years, being 10 male and 28 female. Besides the keyspeech segment corpus, a semi-spontaneous corpus was also recorded. The 38 subjects were asked to talk about the city where they were born. Their recordings were evaluated by two phoneticians and integrated into the instructive material of the BP-VPAS, as samples illustrating voice quality settings (laryngeal, supralaryngeal and tension settings).

For the present analysis, 509 utterances (52 semi-spontaneous speech samples and 457 repetitions of the key-sentences). The total duration of the audio samples analyzed was 114 minutes, mean of 03 minutes per speaker.

The audio samples were analyzed by means of the SG Expression Evaluator, a script developed by Barbosa (2009), running in the software PRAAT, which automatically extracts acoustic parameters: fundamental frequency (f0- median, inter-quartile semi-amplitude, skewness and 0,995 quantile) and its first derivate (df0- mean, standard-deviation and skewness), intensity (skewness), spectral tilt (SpTt – mean, standard-deviation and skewness) and Long-Term Average Spectrum (LTAS frequency standard-

deviation). Data related to intensity measures were normalized taking into account the distance between lips and microphone and the intensity level at a reference tone of 1kHz one. Some of the measures were normalized during the application of the script, such as f_0 and df_0 , SpTt and LTAS frequency.

These measures were statistically analyzed by means multiple and non-linear regression tests, in order to investigate the acoustic parameters as related to distinct perceived voice quality settings.

Results

The results contemplate the correlations, detected by means of statistical analysis, between perceived voice quality settings and acoustic measures, specially for retracted tongue tip (66,7% SD 27,2%), creaky voice (57,5% SD 7,8%), limited lips extension (28,5% SD 12,1%), laryngeal hyperfunction (23,1% SD 11,7%) and limited jaw extension (12,5% SD 11,7%).

F_0 (median, skewness and 99,5% quantil) and f_0 first derivate (mean, standard deviation and skewness) measures have correlated with laryngeal tension (both hyper and hypofunction) and laryngeal (mainly harsh voice, combined or not with whisper, breathy voice and creaky voice) settings. The 99,5% quartile measure made it possible to distinguish utterances characterized by periodicity from the ones characterized by aperiodicity.

Spectral Tilt measures have also correlated with laryngeal tension settings and the laryngeal settings which are characterized by aperiodicity. High values were related to laryngeal hypofunction and low values to hyperfunction. Besides Spectral Tilt, Intensity Skewness has also been helpful to characterize laryngeal tension settings.

Among the findings has been that of the association between supralaryngeal and the laryngeal tension settings: hyperfunction associated with tongue body settings, pharyngeal constriction and raised larynx and hypofunction with pharyngeal expansion and lowered larynx.

Another finding was related to the identification of modal voice settings. If there is little variation of f_0 and df_0 values across speech samples of the same speaker, voices are identified as modal. Furthermore, F_0 and df_0 measures were found to be useful to detect intermittent occurrences of voice quality settings such as laryngeal settings with irregular vocal fold vibration: breathy voice, harsh voice and creaky voice.

The investigation of voice quality settings based on the measures extracted by means of the SG Expressive Evaluator made it possible to correlate acoustic to perceptual dimensions and to identify instances of co-occurrence of settings. Those aspects are relevant not only to consider compensatory strategies to voice disturbances in the clinical context but also but also to foster research on speech expressivity (Madureira, 2008).

Conclusions

The results reinforce the importance of the description of voice quality based on long-term acoustic and perceptual correlations and the role of distinct key speech segments in the corpus designed for voice quality evaluation.

Acknowledgements

We acknowledge Plínio Barbosa from UNICAMP for SG Expression Evaluator Script to the database and Luiz Carlos Rusilo and Giuseppe Milone from PUCSP for the statistical support.

This study has been supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP: 06/59949-6).

References

- Barbosa, P.A. 2009. Detecting changes in speech expressiveness in participants of a radio program. In Proc. of Interspeech. v.1, 2155-2158. Brighton, Reino Unido.
- Camargo, Z., Madureira, S. 2008. Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BP-VPAS). In Proc. of the 4th Conf on Speech Prosody v.1, 57-60. Campinas, Brazil.
- Cukier-Blaj, S.C., Camargo, Z., Madureira, S. 2008. Long-term average spectrum loudness variation in speakers with asthma, paradoxical vocal fold motion and without breathing problems. In Proc. of the 4th Conf on Speech Prosody v.1, 41-44. Campinas, Brazil.
- Hammarberg, B., Gauffin, J. 1995. Perceptual and acoustics characteristics of quality differences in pathological voices as related to physiological aspects. In Fujimura O., Hirano, M (eds.). 1995. Vocal fold physiology, 283-303. San Diego, Singular Publishing.
- Laver, J. 1980. The phonetic description of voice quality. Cambridge University Press.
- Laver, J., Wirs, S., Mackenzie, J., Hiller, S.M. 1981. A perceptual protocol for the analysis of vocal profiles. Edinburgh University Department of Linguistics Work in Progress 14, 139-155.
- Madureira, S. 2008. Reciting a sonnet: production strategies and perceptual. In Proc. of the 4th Conf. on Speech Prosody v.1, 697-700. Campinas, Brazil.

Algerian Arabic rhythm classification

Ghania Droua-Hamdani¹, Sid Ahmed Selouani², Malika Boudraa³,
Wladyslaw Cichocki²

¹SPS Laboratory CRSTDLA, Algiers, Algeria

²LARIHS Lab. Moncton University, Canada

³LCPS. USTHB. Algiers, Algeria

Abstract

This paper investigates speech timing in Modern Standard Arabic (MSA) of Algerian speakers in order to situate their Arabic language between stressed-timed and syllable-timed languages. We extracted first all duration measurements from recordings taken from sound voice bank ALGASD (ALGerian Arabic Speech Database) and after computed durations with using recently interesting rhythmic approaches.

Key words: duration, rhythm, Algerian speakers, Arabic language.

Introduction

Human speech rhythm was principally classified into three big categories: syllable-timed, stress-timed or mora-timed Pike (1945), Abercrombie (1967), Blosh, B. (1950), Ladford (1975). Nowadays, different experimental studies investigate in rhythmic differences between or within languages in order to modelling them with evidences and metrics which will disclose between their prosodic characteristics Ramus (2002), Grabe (2003), Dellwo (2004), Barry (2003), White (2007).

In the rhythm class typology, Modern Standard Arabic (MSA) language has been described as belonging to stressed-timing family Abercrombie (1967). Practical rhythm studies related to MSA are relatively less numerous comparing to the large number of works dealing with other languages as: English, Dutch, Korean, French, Spanish, Portuguese, and Greek. Ramus (2002), Grabe (2003), Jang (2009), Baltazani (2007), O'Rourke (2008).

The purpose of this paper is to investigate speech rhythm of 73 Algerian speakers in order to situate Algerian MSA between languages. Traditionally Arabic is considered to be a stressed-timing language but what about the location of Algerian Arabic between languages in general and between stressed ones in particular?

Methodology

Sound material, taken from ALGASD (ALGerian Arabic Speech Database) voice bank, is composed of two sentences read by all speakers Droua-Hamdani (2010). The analysis consists on extraction and computing all

vowels and consonants duration measurements by using recently interesting rhythmic approaches which are Ramus and Grabe methods which are:

Interval measures (IM)

Ramus approach suggested computing three separate variables which are taken to be acoustic correlates of rhythm classes. These evidences or interval measures (IM) are providing from speech signal segmentation into vocalic and consonantal units. IM correlates are: %V the proportion of total utterance duration which comprises vocalic intervals; ΔV the standard deviation of vocalic intervals and ΔC , the standard deviation of consonantal intervals.

Pairwise variability indices (PVI)

PVI variables are quantitative measures which express the level of variability in successive measurements. nPVI-V parameter calculates the normalized differences of subsequent vocalic durations and rPVI-C one computes successive intervocalic differences intervals (consonants). Their respective formulas are:

$$nPVI = 100 \times \frac{\left(\sum_{k=1}^{m-1} \left| \frac{(d_k - d_{k+1})}{(d_k + d_{k+1})/2} \right| \right)}{(m-1)} \quad rPVI = \frac{\left(\sum_{k=1}^{m-1} |d_k - d_{k+1}| \right)}{(m-1)}$$

Experimentation & results

Traditionally, stressed languages such as English and Dutch are characterized by their highest ΔC and the lowest %V or by high vocalic nPVI and rPVI. However by applying different interval measures of rhythm calculation on Algerian MSA, we founded that Algerian language presents simultaneously a rising ΔC and %V. By comparing results with previous findings reported in different studies made for: stress-timed languages (English, Dutch, German, etc.); syllable-timed languages (French; Italian, Spanish, etc.); mora-timed ones (Japanese) and mixed languages (Polish and Catalan), we noticed from the projection on (ΔC , %V) plane, that the location of Algerian MSA belongs to stress-timed family of languages but set back a bit from them to pattern with syllable-timed languages as French one concerning the %V values.

As regards to Grabe's variables, Algerian rPVI-C is raising as stressed language, but its nPVI-V score seems to be similar that for syllable-timed French and appears on (nPVI-V, rPVI-C) representation patterning Catalan language which is classified as a mixed language.

Figure 1 and Figure 2 show the location of Algerian speakers among languages (stressed, timed, mora and mixed) studied by Ramus and Grabe using respectively IM and PVI correlates.

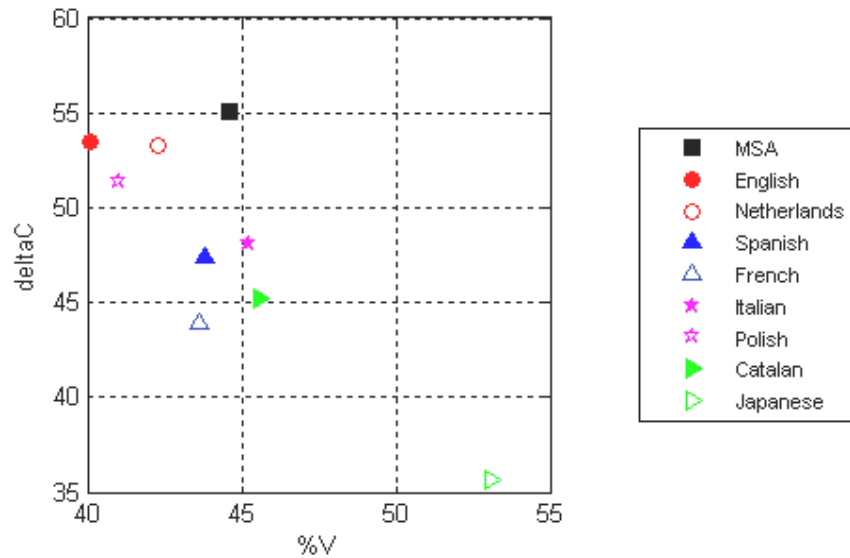


Figure 1. Algerian MSA location between languages according to Ramus measures.

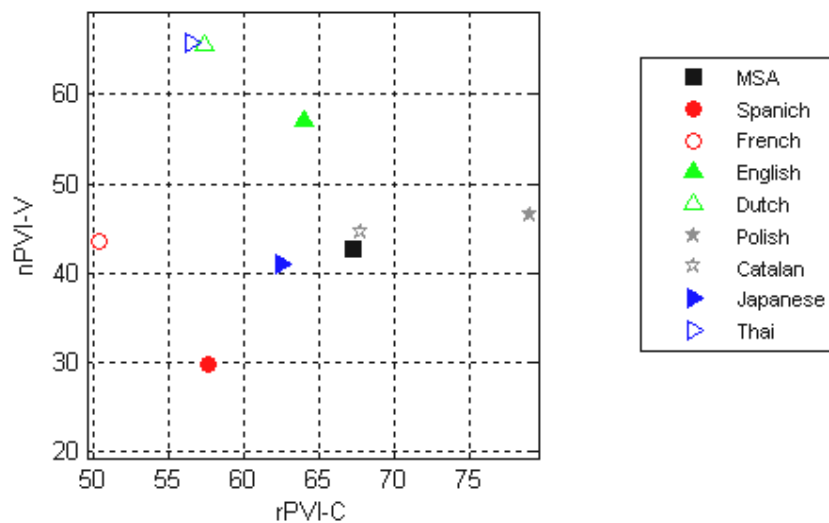


Figure 2. Algerian MSA location between languages according to Grabe measures.

From results, we conclude that although Arabic is classified as stressed language, Algerians tend to pronounce the MSA as an intermediate language (mixed) between stressed and timed languages.

References

- Baltazani, M. 2007. Prosodic rhythm and the status of vowel reduction in Greek. In *Selected Papers on Theoretical and Applied Linguistics from the 17th International Symposium on Theoretical & Applied Linguistics*, Volume 1, pp. 31-43. Department of Theoretical and Applied Linguistics, Salonica.
- Barry, W.J., Andreeva, B., Russo, M., Dimitrova, S., Kostadinov, T. 2003. Do rhythm measures tell us anything about language type? *Proc. of the 15th ICPHS*, Barcelona, Spain.
- Blosh, B. 1950. *Studies in colloquial Japanese IV: Phonemics*. *Language* 26, 86-125.
- Jang, T-Y. 2009. Automatic assessment of non-native prosody using rhythm metrics: Focusing on Korean speakers' English pronunciation. In *SFU Working Papers in Linguistics Vol. 2*. Simon Fraser University, Vancouver, Canada.
- Ladefoged, P. 1975. *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.
- Ramus, F. 2002. Acoustic correlates of linguistic rhythm: Perspectives. In *Proc. of speech prosody*, 115-120. Aix-en-Provence, France
- Abercrombie, D. 1967. *Elements of general phonetics*. Chicago: Aldine.
- Dellwo, V. 2004. Rhythm and speech rate: A variation coefficient for DC. In *Karnowski, P. and Szigeti, (ed). Language and language-processing* 231-241. Frankfurt am Main: Peter Lang.
- Droua-Hamdani, G., Selouani, S.A., Boudraa M. 2010. Algerian Arabic Speech Database (ALGASD): Description and research applications. In *Proc. conference language resources and human language technologies for Semitic languages*. Valetta, Malta.
- Grabe, E., Low, E.L. 2003. Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology* 7, 515-546.
- O'Rourke, E. 2008: Speech rhythm variation in dialects of Spanish: Applying the Pairwise Variability Index and variation Coefficients to Peruvian Spanish. *Speech Prosody*, 6-9 May, Brazil.
- Pike, K.L. 1945. *The intonation of American English*. The University of Michigan Press.
- White, L., Mattys, S. 2007. Calibrating rhythm: First language and second language studies. *Journal of Phonetics* 35, 501-522.

Acoustic correlates of nasopharyngeal resonance

Marios Fourakis, Heather Karlsson, Christie Tilkens, Lawrence Shriberg
Phonology Project, Waisman Center, University of Wisconsin-Madison

Abstract

Measurements of F1 and F2 were made for the four corner vowels produced by typically-developing adolescents and adolescents with Fragile X and Down syndromes. The two high vowels were produced with lower F2 frequencies by the Fragile X and Down syndrome groups than the group of typically developing persons. This is consistent with a hypothesized backing of the tongue.

Key words: F1, F2, corner vowels, Fragile X, Down syndrome

Introduction

Primary goals of the Phonology Project are to identify etiologic origins, risk and protective factors, and diagnostic markers of eight subtypes of speech sound disorders of currently unknown origin (Shriberg, 2010).

The diagnostic instrument used in all Phonology Project studies is termed the Speech Disorders Classification System (SDCS: Shriberg et al., in submission). Perceptual methods to code participants' prosody and voice are based on extensions to the system described in The Prosody-Voice Screening Profiles (PVSP: Shriberg et al, 1992).

Nasopharyngeal resonance

The PVSP describes PV32: Nasopharyngeal Resonance as a "muffled," "back of the throat" quality consistent with the percept of "sluggish or imprecise tongue movement" sometimes used to characterize the speech of persons with Down syndrome. Several of the audio exemplars for PV32 in the PVSP manual were obtained from speakers with Down syndrome. As indicated in Kent and Read (2002) and others, if this "back of the throat" impression is a consequence of moving the tongue further back in the mouth than appropriate for a given vowel, the acoustic effect should be a lowering of F2.

Method

To test the sensitivity of F2 lowering as a potential acoustic correlate of perceived nasopharyngeal resonance, formant measurements of the corner vowels /i/, /æ/, /a/, and /u/ were made in conversational speech samples obtained from three databases in the Phonology Project audio archives: eight 15-to 19-year-old male speakers with Fragile X syndrome, eight 15-to17-year-old speakers with Down syndrome, and 5 typically-developing 14-

year-old and 5 typically-developing 16-year-old male speakers. All tokens from each speaker came from one of their 24 CS utterances eligible for PVSP coding (i.e., were not excluded by one or more of the 32 PVSP exclusion codes). For the speakers with Fragile X syndrome, 3-12 of the 24 utterances had been coded PV32. For the speakers with Down syndrome, at least 22 of their 24 utterances had been coded PV32: Nasopharyngeal. None of the utterances from the 10 reference participants had been coded PV32.

F1 and F2 frequencies for all vowels in utterances coded PV32 were measured using the same procedures as described in the first section of this paper reporting an acoustic correlate for the percept of nasal resonance. The LPC used a number of coefficients appropriate to the sampling rate (SR \pm 4) for each person's recording, which ranged from 20 to 44.1 kHz, depending on the speaker group and the recording date.

Results

Figure 1 is a display of the means of F1 and F2 frequencies of the vowels produced by the speakers with Fragile X syndrome (triangle symbols) and the typically developing children (square symbols). In comparison to the values of the reference group, Fragile X speakers' lower F2 values for high vowels /i/ and /u/ is consistent with backing of the tongue. Effect sizes for these between-group differences are 1.18 (90% CI = 0.25-1.99) for /i/ and 1.05 (90% CI = 0.10-1.89) for /u/. Mean F1-F2 values for the low vowel /æ/ are generally similar for both groups. F1 and F2 values for the low vowel /a/ are higher for Fragile X speakers (effect size for F2 = 2.17; 90% CI = 1.10-3.03).

Figure 2 includes similar comparative findings for the speakers with Down syndrome and the typically-developing speakers. The between-group differences in vowel space are similar to those shown in Figure 1 for speakers with Fragile X syndrome. In comparison to values for the typical speakers, the high vowels for speakers with Down syndrome have lower F2 values, the vowel /e/ has higher F1 and F2 values, and the two speaker groups have similar values for the low vowel /æ/. The between-group F2 effect size for /i/ is 1.23 (90% CI = 0.32-2.01) and for /u/ is 1.15 (90% CI = 0.26-1.93). The between-group F2 effect size for /a/ is 1.06 (90% CI = 0.18-1.84).

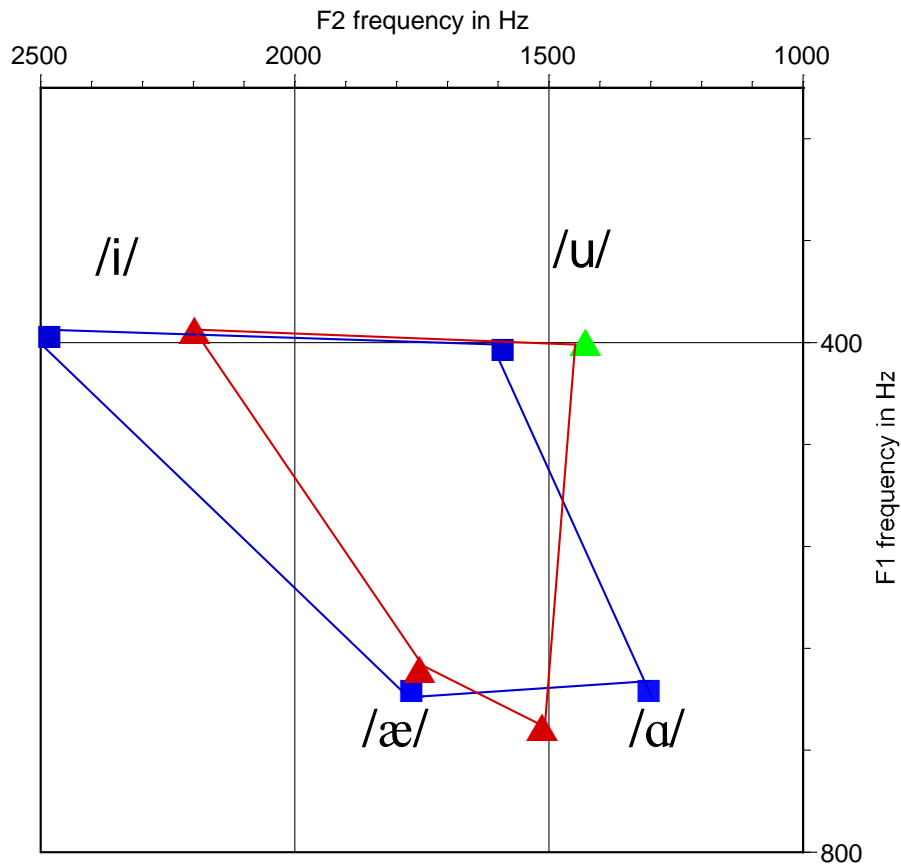


Figure 1. Mean F1 and F2 frequencies of corner vowels produced by adolescents with Fragile X syndrome (triangles) and typically-developing adolescents (squares).

Conclusion

The acoustic findings in Figures 1 and 2 are interpreted as support for lowered F2 values on high vowels as a reliable acoustic correlate of the percept of nasopharyngeal resonance. In both speaker groups with complex neurodevelopmental disorders, the percept of nasopharyngeal resonance was associated with F2 values further back in the vowel space than the corresponding vowels of typically-developing adolescents. Large, statistically significant effect sizes were obtained for both between-group comparisons. The acoustic findings emerging for all perceptual constructs in this research series, studies in progress will attempt to cross validate the present acoustic signature findings for nasal and nasopharyngeal resonance.

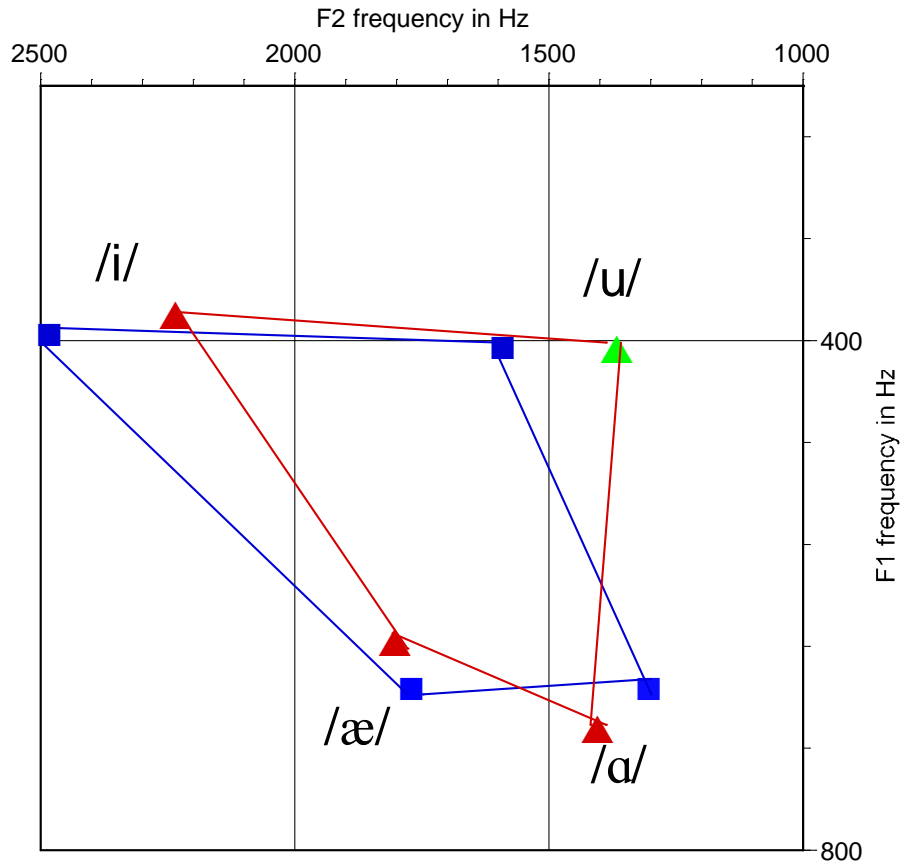


Figure 2. Mean F1 and F2 frequencies of corner vowels produced by adolescents with Down syndrome (triangles) and by typically-developing adolescents (squares).

References

- Shriberg, L. 2010. Childhood speech sound disorders: From post-behaviorism to the post-genomic era. In Paul, R., Flipsen, P. (eds.), *Speech Sound Disorders in Children*. (pp. 1-34). San diego, CA: Plural Publishing.
- Shriberg, L.D., Kwiatkowski, J., Rasmussen, C., Lof, G.L., Miller, J.F. 1992. The Prosody-Voice Screening Profile (PVSP): Psychometric data and reference information for children (Tech. Rep. No. 1). Phonology Project, Waisman Center, University of Wisconsin-Madison.

Dialect and generational differences in vowel space areas

Robert Allen Fox, Ewa Jacewicz

Department of Speech and Hearing Science, The Ohio State University, USA

Abstract

We address methodological issues in calculating the size of the working vowel space area. Although corner vowels are most commonly used in determining its size, we show that this approach may severely underestimate the actual working space. The vowel spaces determined using the large vowels set provided insight into subtle differences between dialects and among age groups. Three approaches were tested which were based on the number of vowels used, formant measurement location and the way boundary lines were drawn.

Key words: vowel, vowel space, formants, dialect.

Introduction

Different dialects of the same language vary in terms both the location and formant movement of vowels in the acoustic (and articulatory) vowel space (e.g., Fox & Jacewicz 2009). A logical question is whether this variation also leads to a significant difference in the size of the acoustic vowel space between speakers of different dialects. The area of a vowel's working space arises from the articulatory configuration of the tongue and jaw that produced vowels. The coordinated tongue and jaw movements are reflected in the formant pattern, which is the basis for measurement of the vowel space area (VSA). Systematic variation in the VSA was found as a function of speaking styles (Bradlow et al. 1996), speech disorders (Turner et al. 1995; Liu et al. 2005) and child development patterns (Vorperian & Kent 2007). The general finding from these studies is that a reduced VSA in production leads to lower intelligibility scores in perception.

However, there are several methodological issues that have not been fully addressed in determining the size of the VSA. For example, should one use a designated subset of the vowels such as the so-called "corner vowels" ([i æ a u]), only the monophthongs (ignoring the diphthongs with their significant formant changes) or all vowels? Second, should the VSA boundaries be drawn on the basis of a single measurement point from each vowel or from more than one point? Finally, which approach should be taken in drawing the boundaries: liberal or conservative? Our current research addresses these issues, and this paper presents an overview of our approach.

Methodology and results

The study examines the variation in VSA as a function of dialect in a larger corpus produced by 260 English speakers of two distinct dialects in the United States: Wisconsin (upper Midwestern dialect) and North Carolina (Southern dialect). More details about these dialect regions can be found in Jacewicz et al. (2009). Data presented here come from 114 female speakers who fall into four age groups (in years): girls (8-12), young (35-50), older (51-65) and oldest (66-91) women. The acoustic measurements used in the calculation of the VSA were taken from 13 vowels produced in citation from “words” in the hVd context: *heed*, *hid*, *heyd*, *head*, *had*, *hod*, *hawed*, *hoed*, *who’d*, *hood*, *hoyd*, *hide*, *howed*.

The boundaries of the VSA were determined on the basis of (1) either the four corner vowels [i æ a u] or all 13 vowels, (2) the F1 and F2 values at the midpoints (50%) or at the 20-35-50-65-80% points and (3) drawing the boundary lines of the VSA using a liberal or a conservative approach. The liberal approach connects the most disparate points with a straight line, ignoring the fact that this may include a portion of the speaker’s VSA in which vowels are never produced. The conservative approach determines boundaries on the basis of where vowels are actually found in the VSA as produced by that particular speaker. These differences are illustrated in Figures 1 and 2 which display three of the vowel spaces calculated for an individual North Carolina female speaker. Mean group areas are provided in Table 1.

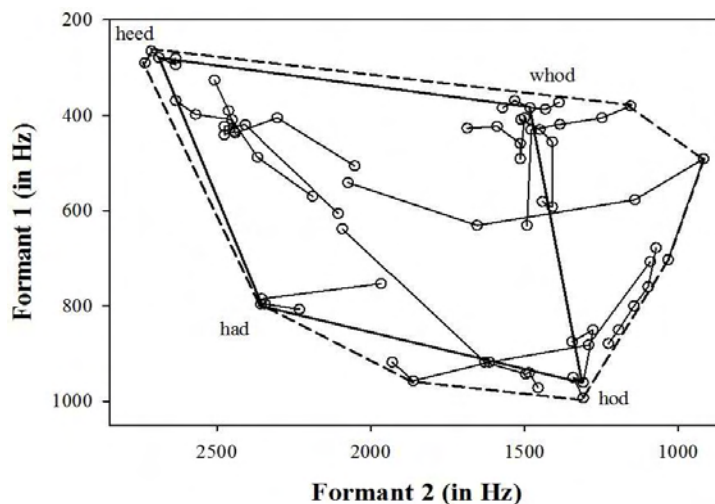


Figure 1. Vowel space areas encompassing the four corner vowels measured at the midpoint (bold solid line) and all vowels measured at all five temporal points using a liberal approach (broken line).

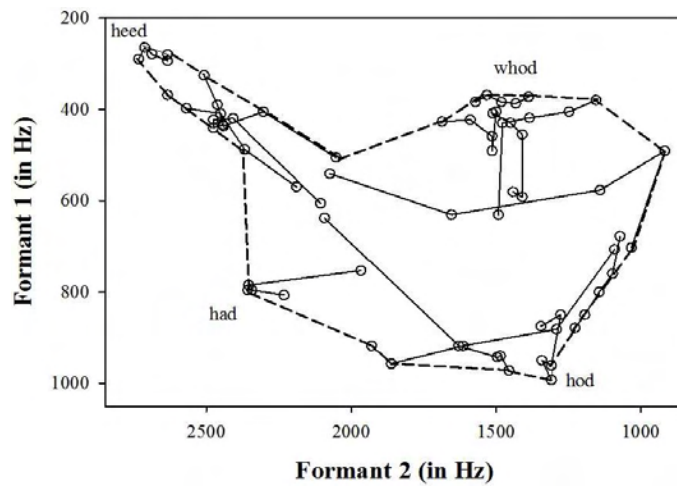


Figure 2. Vowel space area encompassing all vowels and all five measurement points using a conservative approach (broken line).

Table 1. Mean liberal and conservative vowel space areas for measurements at the midpoint and at all 5 temporal points for four groups of females.

Dialect	Age	Corner Vowels		All 13 vowels	
		Midpoint	All	Midpoint	All
Liberal Vowel Space Boundaries (kHz ²)					
Wisconsin	8-12	0.887	1.042	1.039	1.240
	35-50	0.650	0.765	0.733	0.874
	51-65	0.657	0.822	0.761	0.923
	66-91	0.696	0.810	0.804	0.923
North Carolina	8-12	0.601	0.791	0.773	1.112
	35-50	0.392	0.522	0.662	0.860
	51-65	0.384	0.525	0.673	0.869
	66-91	0.362	0.543	0.676	0.867
Conservative Vowel Space Boundaries (kHz ²)					
Wisconsin	8-12	0.887	1.003	0.739	0.967
	35-50	0.650	0.728	0.524	0.664
	51-65	0.657	0.780	0.567	0.672
	66-91	0.696	0.765	0.563	0.699
North Carolina	8-12	0.601	0.688	0.675	0.882
	35-50	0.392	0.480	0.591	0.708
	51-65	0.384	0.475	0.623	0.697
	66-91	0.362	0.477	0.624	0.663

Conclusions

A review of the data in Table 1 underscores the challenges and pitfalls in making generalizations and theoretical claims regarding differences in VSA between individual speakers, age groups (children versus adults) and dialects. For example, the VSA for North Carolina (NC) speakers determined on the basis of the midpoints of the corner vowels (probably the most common approach used in calculation of VSA) is much smaller than that of the Wisconsin (WI) speakers. This is primarily because the NC /u/ is much more fronted than the WI /u/. However, the positions and movements of the back vowels (especially /ɔ/) and the onglide of the /oɪ/ for NC vowels makes the difference in VSA between the NC and WI speakers much smaller (a difference which is actually reversed in the conservative approach) when VSA is calculated using all vowels and all measurement points. Our overall conclusion is that when making comparative evaluations of VSA, a rationale and motivation should be provided for the measurement selection criteria (in terms of both vowels and measurement points) in determining the boundaries of the VSA.

Acknowledgements

This work was supported by NIH/NIDCD research grant R01DC006871.

References

- Bradlow, A., Torretta, G., Pisoni, D. 1996. Intelligibility of normal speech. I. Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication* 20, 255-272.
- Fox, R.A., Jacewicz, E. 2009. Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America* 126, 2603-2618.
- Jacewicz, E., Fox, R.A., O'Neill, C., Salmons, J. 2009. Articulation rate across dialect, age, and gender. *Language Variation and Change* 18, 285-316.
- Liu, H.-M., Tsao, F.-M., Kuhl, P. 2005. The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *Journal of the Acoustical Society of America* 117, 3879-3889.
- Turner, G., Tjaden, K., Weismer, G. 1995. The influence of speaking rate on vowel working space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech, Language and Hearing Research* 38, 1001-1013.
- Vorperian, H., Kent, R. 2007. Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language and Hearing Research* 50, 1510-1545.

How to incorporate qualitative analysis into experimental research — Why Japanese students studying in the UK did not improve their fluency? —

Misa Fujio

Business Administration Department, Toyo University, Japan

Abstract

This study investigated the longitudinal improvement in fluency of three Japanese participants who studied at a graduate school in the UK. In addition to quantitative analysis of videotaped conversations focusing on core components of fluency, such as speech rates, dysfluency markers and pauses, qualitative analysis, including protocol analysis and journal analysis, were also employed to discuss why they did not improve in fluency as originally expected. The analyses revealed that these participants did not have as many chances to communicate with native speakers as originally expected, and that a sociocultural factor of being excessively afraid of making mistakes might be a reason that prevented their natural improvement.

Key words: fluency, speech rates, dysfluency markers, pauses, sociocultural elements

Introduction

In order to measure learners' linguistic ability, three constructs – fluency, accuracy, and complexity – are often considered the key constructs (i.e. Skehan 1998). Especially, fluency plays an essential role in spoken language which can be characterised by fragmentation and involvement (Chafe 1982). Although fluency is a widely-accepted term, it is difficult to define and measure it precisely, including how to treat dysfluency markers or pauses. In this study, how the measurement of fluency can be operationalised, how the participants – three Japanese learners of English – improved or did not improve in fluency during their stay in the UK, and what are the reasons if they did not improve as expected will be discussed. Especially, in order to discuss the third point, qualitative approaches such as the analysis of post-interviews and journals are employed in addition to quantitative approach. Incorporating a qualitative approach is especially suggestive for English learners in Eastern countries, who sometimes have complex cultural and personal reasons that prevent them from attaining high level of performance. Therefore, the benefit of combining the two approaches is also discussed.

Definition of Fluency

Fluency has been defined in several different ways. In the field of second language acquisition, it is associated with “the processing of language in real time” (Schmidt, 1992: 358) and is often contrasted with the term, *knowledge*, such as lexical or grammatical knowledge. In Lennon (1990), fluency is regarded as “a purely performance phenomenon” and defined as “an impression on the listener’s part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently” (1990:391). In the article, the core elements that enable us to measure fluency are also discussed: speech rates, dysfluency markers and pauses. In this study, therefore, these three elements were measured to discuss fluency.

Methodology

This longitudinal study measured and analysed the development in fluency of three Japanese learners of English who stayed in the UK for one year. The details of methodology are summarised below.

Research Questions

For this research, the following three specific questions were formulated. In this paper presentation, 1) and 3) will be focused.

- 1) How would the fluency of the three Japanese improve during their stay in the UK for one year?
- 2) How their improvement in fluency and in communication are related?
- 3) If they do not improve in fluency, what reasons are considered?

Participants

The data was collected for one year from three Japanese learners of English who studied at a graduate school in the UK. All of them had a high proficiency score (measured by a TOEIC test) even before their study in the UK; however, they had not had many chances to use English in actual communication in Japan. In this sense, they seemed to be ideal research participants who would show a lot of improvement in fluency during their stay in the UK.

Data Collection and Analytical Procedure

For the data collection, a multi-method was employed in order to make my analysis as objective and precise as possible. In addition to the main data obtained by the monthly videotaping of conversations between each Japanese participant and a native speaker of British English, retrospective comments of these conversations, comments in post-interviews, and journals about their life in the UK were collected. The main data was analysed to measure fluency, including speech rates, dysfluency markers and pauses, using voice-analysing software, Praat. As for disfluency markers, fillers, false starts and repetitions were included, but not self-corrections, which might have stemmed from grammatical or communicative purposes. The rest of the data was used to interpret the quantitative results.

Analyses and Discussion

The analyses revealed that none of the three participants showed any visible or statistically significant improvement over the one year period (as a result of ANOVA with two-way layout without repetition). This result can be explained as follows: 1) they stayed either in London or Oxford where many Japanese also resides and naturally they got together and spoke in Japanese; 2) in their programme, most of the students were from abroad and they had many fewer chances to speak with British students; and 3) since they had a lot of reading assignment, they tended to stay home, working on the assignment, rather than to communicate with others. All of these reasons reveal that studying in an English-speaking country does not necessarily guarantee the opportunities to communicate in English.

Especially, one of the participants retained extremely low speech rates, 60 wpm (including pauses), and her pauses accounted for as much as 10% of her floor-holding. In a post-interview, she commented that she needed the pauses to make better sentences. This value – making much of accuracy, even costing fluency – partially stems from her personality of being a perfectionist, and partially from the cultural trait of Japanese, who feel that making mistakes is shameful. However, she noticed gradually that taking long pauses is not advantageous because people treat her as those who cannot speak English fluently. In fact, her speech rates and her communicative attitude seem to have fixed her role as a passive listener (Shea 1994), which might be a reason that slowed down her improvement in fluency.

References

- Chafe, W.L. 1982. Integration and involvement in speaking, writing, and oral literature. In Tannen, D. (ed.) *Spoken and Written Language: Exploring Orality and Literacy*, 35-53. Norwood, NJ. Abex.
- Lennon, P. 1990. Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40(3), 384-417.
- Schmidt, R. 1992. Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14(4), 357-385.
- Shea, D.P. 1994. Perspective and production: Structuring conversational participation across cultural borders. *Pragmatics*, 4(3), 357-389.
- Skehan, P. 1998. *A cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Evaluating languages functions in a trilingual setting

Ioannis Galantomos¹, Georgia Andreou², Fotini Anastassiou²

¹Department of Special Education & Department of Language, Literature and Culture of the Black Sea Countries, University of Thessaly & Democritus University of Thrace, Greece

²Department of Special Education, University of Thessaly, Greece

Abstract

The aim of this research was to investigate the language functions of trilinguals that live in Greece. This study consisted of personal interviews conducted at a State Evening School in Volos in January 2010 and the participants were between 19 and 24 years old. They were all male and they came from immigrant families of Albanian origin. The study aimed to locate the instances that each one of the languages - L1 Albanian, L2 Greek, L3 English - is used and to record the reasons that make the students use each one of their languages.

Key words: trilinguals, language functions, attitudes, immigrants.

Introduction

According to Tucker (1998), multilinguals are far more compared to monolinguals in the world. Recently there has been rapid interest in studies on the acquisition of a third language, most of them focusing on English, as far as educational contexts, since English and multilingualism have become more and more important due to globalization (Cenoz & Genesee, 1998). Nevertheless, there has not been much published work on the ever growing number of children who are brought up as multilinguals which is mainly because of the vast mobility of populations between countries and mixed marriages (Barnes, 2005).

Trilingualism/ multilingualism is a rather recent field of research and thus there are a lot of differentiations on its definition. In fact “trying to reach to a proper definition of multilingualism could be said to be the most demanding problem of current linguistics” (Jessner, 2008). Cenoz and Genesee (1998) suggest that a person should be defined as trilingual if he can use his three languages to communicate in both oral and written speech. Furthermore, due to the fact that many scholars argue that learning an L3 differs from learning an L2, multilingualism is only used (or should be used) to refer to the learning of more than two languages (Hufeisen, 1998).

Especially during the last decade more and more immigrants have come to Greece to live and work here permanently. The most prominent immigrant community is the Albanian one in Greece. Nowadays, we host the second generation of Albanian immigrants. Therefore, we found it rather important

to investigate the languages of Albanian students in Greek schools. Their languages were L1 Albanian, L2 Greek and L3 English. Their L2 and L3 were taught at school. English is of great interest since it is a world language and the one most widely found in multilingual contexts (Barnes, 2005).

Method

Participants

Our sample was composed of thirty students of Albanian origin. All of them were male, aged between fifteen and twenty four years old. They studied at the local State Evening School in Volos and their level in English was approximately B1, according to the Common European Framework of Reference for Languages. Their mid age was 19.3 years old and their mid period of residence in Greece was 12.3 years. The research was contacted through personal interviews, which took place in Volos in January 2010, and each one of the students had to answer to two questions concerning their three languages.

Procedure

The first question aimed to determine where each one of the languages is used, that is at school, at work, within the family or friend environment.

Question:

- 1.1. Where do you use Albanian?
- 1.2. Where do you use Greek?
- 1.3. Where do you use English?

The second question aimed to record the reasons that make these students to use each one of their three languages.

Question:

- 2.1. When and why do you use Albanian?
- 2.2. When and why do you use Greek?
- 2.3. When and why do you use English?

Results

The answers that were recorded were teamed according to their common features so that they can be further analyzed:

Albanian

- 1.1 At home: 90%
 - At work: 10%
- 1.2 To communicate with their family: 65%
 - To communicate with their friends: 15%
 - During their visits to Albania: 5%
 - To preserve their national identity: 15%

Greek

- 2.1 At school/ daily life: 94%
 - At work: 4%
 - With their family: 2%
- 2.2 To communicate with their friends: 56%
 - To communicate at work: 4%
 - To help family members that can not speak Greek well: 2%
 - Frequent use contributes to better learning of Greek: 38%

English

- 1.1 At school: 98%
 - During their holidays: 2%
- 1.2 In class: 76%
 - As a sign of personal progress: 20%
 - To show off to their compatriots who can not speak English: 4%

Discussion

The participants of this research stated that they use Albanian at home and a small percentage uses Albanian at work too. This is reasonable since their family speaks their L1 at home, as it is usual in immigrant families, to preserve their inheritance and thus their national identity.

On the question where do they use Greek the majority of the students answered at school. It is likely that both students and their parents regard Greek as the lingua franca of educational advancement in Greece. Cummins (1987) emphasizes on power relations between groups within the school environment and between teachers and students. These power relations are determined to a degree by the very nature of being an L2 learner, whether the learner is adult or child. We could dare extend this to the L3 learners too.

The fact that they believe that if they use Greek more frequently they will learn the language more easily is reasonable. Also, according to Oliver & Purdie (1998) students perceive that their teachers and peers feel more positively when the environment language is used rather than their L1, in all contexts.

Regarding English the students stated that they mainly use English at school, which means in class during the English language lesson. English has a limited use due to the very nature of the specific school that this research took place.

Attitudes are determinants of the manner in which students engage in language learning at school, they influence learners' expectations for success (Cummins, 1986) and they play a significant role in students' successful maintenance of L1 (Dopke et al, 1991).

Our research tried to make a small step towards understanding attitudes and language functions of trilingual speakers. Studies on trilingualism/multilingualism in Greece have been a recent field of research, especially compared to the rest of Europe. Our first attempt was a small scale study and a larger scale study could probably enlighten us more.

References

- Cenoz, J., Genesee, F. 1998. Beyond Bilingualism. Multilingualism and Multilingual Education. Clevedon: Multilingual Matters.
- Cummins, J. 1986. Individual Differences in Second Language Learning. London, Edward Arnold.
- Cummins, J. 1987. Language development, academic achievement and empowering minority students. In Fradd, S. and Tikunoff, W. (eds.) 1987, Bilingual Education and Bilingual Special Education: A guide for Administrators, 75-97. Boston, College Hill Press.
- Dopke, S., McNamara, T.F., Quinn T.J. 1991. Psycholinguistic aspects of bilingualism. In Liddicoat, A. (ed.) 1991, Bilingualism and Bilingual Education, 21-79. Melbourne, NLLIA.
- Hufeisen, B. 1998. L3-Stand Der Forschung – Was Bleibt zu tun? In B. Hufeisen and B. Lindemann (eds) Tertiarsprachen: Theorien, Modelle, Methoden. Tübingen: Stauffenburg, 169-185.
- Jessner, U. 2008. Teaching third languages: Findings, trends and challenges. Language Teaching 41, 15–56.
- Tucker, G.R. 1998. A global perspective on multilingual and multicultural education. In J. Cenoz and F. Genesee (eds) Beyond Bilingualism: Multilingual and Multicultural Education. Clevedon: Multilingual Matters, 3-15.
- Oliver, R., Purdie, N. 1998. The attitudes of Bilingual children to their languages. Journal of Multilingual and Multicultural Development, vol.19, No. 3.

Production of inflectional morphology in a child with moderate hearing impairment

Thalia Hatzivasiliou, Marianna Hatzopoulou

Department of Logopaedics, Athens Metropolitan College, Greece

Abstract

The present study explores the production of bound morphemes in the speech of a nine-year old Greek child with moderate hearing impairment (HI). Data were obtained through the child's responses to subtests from the Diagnostic Verbal IQ Test (DVIQ, Stavrakaki et al. 2000) and samples of her speech. The analysis concerned the inflection of nouns, adjectives and verbs. The results showed that, even though the morphological features of names and verbs were marked correctly in the majority of utterances, errors were still observed in the child's speech. These errors concerned mainly the marking of the plural number in nouns and the perfective aspect in verbs.

Key words: hearing impairment, morphology, bound morphemes

Introduction

Studies on the acquisition of morphology in children with HI report that, even though grammatical morphemes are more easily perceived and processed by children with better residual hearing, there are still deficiencies observed in their speech.

Research on the production of morphemes by English deaf children using cochlear implants, which improve their residual hearing to moderate hearing loss, has shown that they exhibit difficulties in inflectional morphology and in particular that they have problems in marking the past tense of verbs and the plural number of nouns (Spencer et al. 1998; Svirsky et al. 2002). Similar findings have been reported for children with mild to moderate HI. Even though their language skills are comparable to those of their hearing peers, there are still some weaknesses observed in the production of morphemes (Elfenbein et al. 1994; Norbury et al. 2001; Hansson et al. 2007). Elfenbein et al. (1994) mention difficulties in marking verb tenses and plural number of irregular nouns, while Hansson et al. (2007) detected difficulties in marking tense of novel verbs.

Aim of the study

The aim of the present study was to assess the inflectional morphology of nouns, adjectives and verbs in a hard of hearing (HH) child and to detect any possible error patterns. It was expected that morphological errors, which

concern mainly verb tenses as well as plural forms of nouns, would be present in the child's speech.

Method

The participant of this study was a nine-year-old Greek girl (M.) with a rare congenital syndrome (Lacrimo-Auriculo-Dento-Digital syndrome - LADD), which has caused her conductive, moderate HI (hearing thresholds at 58,3dB for the right and 68,3dB for the left ear). M. has been using hearing aids since the age of seven. Data were collected by speech samples and by M.'s responses on subtests of the DVIQ (Stavrakaki et al. 2000). The analysis concerned the correct or erroneous use of inflectional morphemes in nouns, adjectives and verbs. The erroneous forms were further analyzed for markedness of the gender, number and case features in nouns and adjectives, and of the person, number, tense and aspect features in verbs.

Results

Measurements of the correct and erroneous tokens showed that there were a few errors presented in M.'s speech. M. produced a total of 32 erroneous forms, which correspond to 10.8% of her overall production of nouns, adjectives and verbs (295 uttered tokens). Even though M. produced more verbal (152) than nominal (143) tokens, she made more errors in nouns and adjectives than in verbs. The rate of errors in the production of nouns and adjectives reached 14% (20 erroneous forms out of the 143 nominal tokens), while the rate of errors in the production of verbs was 7.9% (12 erroneous forms out of the 152 verbal tokens).

Table 1. Numbers and percentages of erroneous types of uttered names and verbs.

	No of errors/No of tokens	% of errors
Nouns and adjectives	20/143	14.0%
Verbs	12/152	7.9%
Total	32/295	10.8%

The analysis of the errors regarding the production of nouns and adjectives revealed that the most frequent errors concerned the number feature, as 14 of the 20 erroneous nominal tokens were incorrect plural forms (9.8% of the total production of nouns and adjectives). M. produced an incorrect plural suffix in 8 out of the 14 erroneous plurals, while she did not mark the plural feature in 5 instances. In one instance, she used a plural instead of the obligatory singular form. The remaining 6 incorrect forms concerned errors of the gender and case features.

The most frequent error type in verbs concerned the aspect feature. M. made 5 errors of aspect out of the 12 erroneous verbal tokens (3.3% of the total production of verbs). These errors referred to production of imperfective forms in contexts where the perfective form was expected. The rest of the errors were distributed across all other features: person 1, number 1, mood 1, tense 2, voice 2.

Discussion

The aforementioned results indicate that, even though morphological features were correctly marked in the majority of the child's utterances, there were still a few erroneous forms produced, as it was expected according to the literature (Elfenbein et al. 1994; Norbury et al. 2001; Hansson et al. 2007). Interestingly, in contrast to previous studies, the child, exposed to Greek, exhibited more difficulties regarding the morphology of nouns and adjectives than that of verbs. Studies in English have shown that children with HI have more difficulties in marking tense in verbs (Spencer et al. 1998; Svirsky et al. 2002). This discrepancy could possibly be attributed to the morphological characteristics of Greek compared to English.

A further qualitative analysis of the produced nominal errors concerning the incorrect marking of the plural form showed that they were probably overgeneralizations of specific morphological rules for the formation of plural number common in many nouns and adjectives. Namely, she overgeneralized the suffixes *-es/* and *-i/* for masculine and *-es/* for feminine names. M.'s main weakness in the morphology of verbs was detected in marking the perfective aspect of verbs. This pattern could also be considered as an overgeneralization strategy, as she marks only the tense morpheme to express both perfective and imperfective types, because she does not have the perfective morpheme at her disposal.

Factors affecting the acquisition of morphology

Various factors may have affected the acquisition of morphology in M.'s speech. The perceptual prominence hypothesis, which has been proposed for children with HI (Svirsky et al. 2002), may interpret some of M.'s difficulties, such as marking aspect, which concerns internal sound changes. However, it cannot account for all the errors observed in M.'s speech, such as plural marking in imparisyllabic nouns.

In the analysis, we speculated that grammatical and lexical familiarity affected the acquisition and production of morphemes. It was observed that the child produced more errors in categories that follow particular morphological rules, which do not occur frequently in Greek or in words that she did not recognize. Lexical familiarity has been suggested to affect the morphology of children with HI. Norbury et al. (2001) maintained that

verbal morphology is affected by the verb's frequency, while Hansson et al. (2007) found that children with HI have difficulties in applying morphological rules to novel verbs.

Conclusion

Overall, the present study confirmed the previous findings on the acquisition of morphology of children with mild to moderate HI, showing that there are difficulties with the inflection of nouns, adjectives and verbs. The general assumption is that several linguistic factors, such as the perceptual salience, the grammatical and lexical familiarity of the linguistic input, may affect the morphological development of HH children. Therefore, it is important for children to be widely exposed to grammatical morphemes, especially those that are not frequently included in the linguistic input. Further research is needed on the grammatical development in Greek children with HI, as well as on the factors that affect the acquisition of morphology.

References

- Elfenbein, J.L., Hardin-Jones, M.A., Davis, J.M. 1994. Oral communication skills of children who are hard of hearing. *Journal of Speech and Hearing Research* 37, 216-226.
- Hansson, K., Sahlen, B., Maki-Torkko, E. 2007. Can a 'single hit' cause limitations in language development? A comparative study of Swedish children with hearing impairment and children with specific language impairment. *International Journal of Language and Communication Disorders* 42, 307-323.
- Norbury, C.F., Bishop, D.V.M., Briscoe, J. 2001. Production of English finite verb morphology: A comparison of SLI and mild-moderate hearing impairment. *Journal of Speech, Language and Hearing Research* 44, 165-178.
- Spencer, L.J., Tye-Murray, N., Tomblin, J.B. 1998. The production of English inflectional morphology, speech production and listening performance in children with cochlear implants. *Ear and Hearing* 19, 310-318.
- Stavrakaki S., Tsimpli, I. M. 2000. Diagnostic Verbal IQ Test for Greek preschool and school age children: Standardization, statistical analysis, psychometric properties. In *Proceedings of the 8th conference on speech therapy*: 95-106. Athens: Ellinika Grammata.
- Svirsky, M.A., Stallings, L.M., Lento, C.L., Ying, E., Leonard, L.B. 2002. Grammatical morphologic development in pediatric cochlear implant users may be affected by the perceptual prominence of the relevant markers. *The Annals of Otology, Rhinology & Laryngology* 111, 109-112.

How learning to read and write influences the dialectal traits in -sC clusters

Isabel Henriques

Clup-Centro de Linguística da Universidade do Porto, Portugal

Abstract

Words written with the initial graphic sequences “es+C” and “ex+C” raise many problematic questions to the phonology of European Portuguese since the production of these sequences is changing (Delgado Martins *et al.* 1996:302). This paper focuses on the analysis of how the speakers produce and segment these –sC clusters and on the variables that might affect these tasks. Therefore, we conducted a study to perceive what might influence the speakers’ phonological knowledge.

Key words: phonological knowledge, orthographic knowledge, dialectal traits, variation and change..

Introduction

Words with the initial –sC clusters are definitely a controversial topic in the phonology of European Portuguese, as in other Romance languages, especially when the syllabic status of the cluster has to be defined since the production of these sequences is being modified (Delgado Martins *et al.* 1996:302) by the loss of the initial vowel. The difficulty arises when the speakers have to identify the syllabic constituents, since the word initial -sC clusters cannot be accepted as tautosyllabic (Veloso 2002), because they violate the Sonority Principle (Selkirk 1982:343; 1984:110).

The problem only arises in initial clusters of words with the graphic sequences “es+C” and “ex+C”, as the initial vowel tends to disappear and they are also produced differently due to the dialectal origin of the speakers, especially among speakers from the septentrional (Northern area-Oporto) and central meridional dialect (Southern area-Lisbon). There is a clearer tendency to vowel deletion in the central meridional dialect, while in the septentrional dialect vowel insertion is more frequent, according to the dictionaries and the characteristics of the two dialects (Barros Ferreira *et al.* 1996: 495, based on Cintra 1970).

Some of the core concerns in the study of these sequences, as well of this paper, are:

- i) If these sequences are considered tautosyllabic, we are facing a violation of the Sonority Principle (Selkirk 1984:110) and of the Dissimilarity Condition.

- ii) It may be a difficult task for speakers to clearly define the syllable boundaries.
- iii) There may be some variation due to the dialectal origin of the speakers.
- iv) There may be some influence of the orthographic knowledge in the way the speakers produce and segment these words, since there is an initial graphic vowel, <e>.
- v) The orthographic knowledge may be changing the dialectal traits.

Experimental study

Elaborating on the analyses available in the literature for European Portuguese (Freitas 1997, 2000; Andrade and Rodrigues 1999; Rodrigues 2000; Rodrigues 2003; Freitas and Rodrigues 2003), we considered to be fundamental to conduct an experimental study which was orientated with following aims:

- i) to find evidence that the orthographic knowledge interferes with the phonological knowledge (Velooso 2003) and with the changing trend in the production of these sequences;
- ii) to determine if we can talk about variation in case the sequence fricative plus plosive may be considered tautosyllabic by the speakers;
- iii) to test the influence of the dialect in the two orthographic sequences (“es+C” and “ex+C”).

Having these purposes in mind, we selected some speakers from different age groups and from different dialectal origin. We chose 10 speakers who attended the 5th grade and the 9th grade from Lisbon (central meridional dialect) and Oporto (septentrional dialect). The speakers were asked to produce the words with –sC clusters (e.g. *espada/sword*; *escola/school*; *exposição/exhibition*; *experiência/ experience*) and segment the syllables of those words, after some images were shown.

All the interviews were recorded; then, the recordings were carefully heard several times and finally the data were analysed with SPSS16, in order to see what were the statistically significant differences. We have chosen the Mann-Whitney test because we were testing two independent variables: the orthographic knowledge and the dialect.

Results

The results of the production of the –sC clusters without initial vowel were not statistically significant among the speakers of the 5th grade from the two dialects ($z=-0,506$; n.s.). While the results of the production without initial vowel of the –sC clusters among the speakers of the 9th grade of the two

dialectal groups were statistically significant ($z=-2,834$; $p<0,005$). The results of the task of segmentation were not statistically significant in both groups, regardless of the dialectal origin.

Conclusions

Some of the data showed that the dialectal traits may be disappearing due to the influence of learning to read and write. Therefore, one must conclude that some of the dialectal traits may be changing, when the speakers produce and segment these sequences. While there was a clear tendency to vowel deletion in the production (e.g. [ʃkɔlə]-*escola/school*), the segmentation led to vowel insertion (e.g. [iʃ-kɔlə]). Nonetheless, some dialectal traits still persist, such as: the speakers from Oporto attending the ninth grade showed a clear tendency to the insertion of vowels ([i],[i] and diphthongs [ej], [ej]) both in the production and segmentation of these clusters.

After the experimental study we can conclude that:

- i) The speakers from the Northern dialect tend to produce these sequences with an initial vowel but there is no statistically significant difference, with the exception of the speakers attending the 9th grade.
- ii) Words with initial -sC clusters are segmented as heterosyllabic.
- iii) Speakers introduce a vowel in the segmentation.
- iv) The results were similar among the different groups regardless of the school year or dialect. So, the dialect differences are fading as a consequence of the process of learning to read and write.
- v) There was a certain influence of the orthographic knowledge, especially in the words with the graphic sequence “ex+C”.

On the other hand, some of the early predictions proved to be wrong or at least inconclusive, such as:

- i) The possibility of tautosyllabic segmentation, based on Delgado-Martins (1996:302). This hypothesis did not occur in this study.
- ii) The possibility of change and variation. The only change is the fact that the results of both dialectal groups are similar, which was unexpected since there are some clear differences in the production of these sequences. This is possibly the result of orthographic knowledge.

To sum up, the similar results among the different groups from different dialects may be due to the influence of the orthographic knowledge acquired at school.

Nonetheless, there are still questions related to the ‘grey area’ of the position of the fricative [ʃ] in the syllable. These can be some of the topics to

be considered in a future study with more data from language acquisition, tests with larger number of speakers and also with different tasks, such as reading and spontaneous speech.

References

- Barros Ferreira, M. *et al.* 1996. Variação Linguística: perspectiva dialectológica. In *Introdução à Fonologia Geral e Portuguesa*. Lisboa. Caminho, 479-502.
- D'Andrade, E., Rodrigues, C. 1999. Das Escolas e das Culturas: História de uma sequência consonântica. *Actas do XIV Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: APL/Colibri, 117- 133.
- Delgado-Martins, R. *et al.*. 1996. Fenómenos de Reestruturação Silábica em Português de Lisboa. In: M. R. Delgado-Martins. 2002. *Fonética do Português. Trinta anos de Investigação*. Lisboa, Caminho, 301-318.
- Freitas, M.J. 1997. *Aquisição da Estrutura Silábica do Português Europeu*. Dissertação de Doutoramento. Universidade de Lisboa.
- Freitas, M.J. 2000. O grupo consonântico s+C em início de palavra na aquisição do português europeu. *Actas do XV Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: APL/Colibri, 499-512.
- Freitas, M.J., Rodrigues, C. 2003. O grupo consonântico s+C em início de palavra na aquisição do português europeu. *Actas do XV Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: APL/Colibri, 499-512.
- Rodrigues, C. 2000. Novos Dados acerca de /# S\$C/. *Actas do XV Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: APL/Colibri, II, 287-299.
- Selkirk, E.. 1982. The Syllable. In Van der Hulst, H., Smith, Norval. *The Structural of Phonological Representations. Part II*. Dordrecht, Foris Publications. 337-361.
- Selkirk, E.. 1984. On the Major Class Features and Syllable Theory. In Aronoff, M. *Language, Sound and Structure. Studies Presented to Morris Halle by his teacher and student*, 107-136.
- Veloso, J. 2002. Do Fricative+ Plosive Onsets exist word-initially in European Portuguese?. In: A. Eftimova, K. Petrova (Eds.). *Litora Psycholinguistica*. Sofia: Sema RSH, 49- 56.
- Veloso, J. 2003. *Da Influência do Conhecimento Ortográfico sobre o Conhecimento Fonológico*. Dissertação de Doutoramento apresentada à Faculdade de Letras da Universidade do Porto.

Tarascan dialectal and generational differences around aspirated stops

Francisco Javier Bucio García

Department of linguistics, El Colegio de México-UAM, Mexico

Abstract

The Tarascan or phurépecha language contrasts aspirated stops from non-aspirated word initially, but this contrast was hardly heard in Hiwatzio, a lake dialect (Friedrich 1971:177). We measured the length of the aspirated feature following stops and vowels /i a u/'s formants F1 and F2 so as to identify differences among subjects. The results show that adult men speakers from a mountain dialect make an aspiration shorter than men from a lake dialect. The opposite pattern applies for women belong to the same dialects. Regarding generational differences, a young man from a lake dialect does a shorter aspiration than the older men do and a young girl's F1 and F2 of /i, a, u/ tend to be higher than older women.

Key words: tarascan, aspirated, generational, dialectal differences.

Previous analysis

The Tarascan language is spoken in Michoacán Mexico and is divided into the lake and the mountain dialects. The first dialectal work reported that the contrast between aspirated and non-aspirated stops was hardly perceived for Hiwatzio speakers, a lake dialect (Friedrich 1971:177). A second dialectal work (Chavez 2004) points some phonological differences but does not report any lost in the stop contrast. A subsequent dialectal work focuses on the morpho-syntactical level (Chamoreau 2005). Neither Friedrich nor Chavez mention age nor gender of their speakers and did not use any computer programme for analyzing their data so as to support their claims. We address the same questions as theirs: what are some dialectal differences? And add a new one: Is there any generational difference among speakers of the same gender?

We applied a questionnaire to subjects whose first language is Tarascan and use it in everyday communication with family and friends, but they also speak Spanish for commercial purposes and with foreigners. The subjects were as follows: i) two men from Janitzio and one man from Santa fe (lake dialects); one man from Capacuaro, (mountain dialect). All men are in between 50 and 60 years old. We also applied it to ii) one woman from Janitzio and one woman from Capacuaro, 45 and 47 years old respectively; iii) one young man, 25 and a girl, 23 years old both from Janitzio.

The questionnaire is made up with 80 isolated words based on a grammar and a dictionary (Foster 1969, Velásques 1987). We asked speakers repeat each word twice and recorded them with a Sony digital recorder ICD-UX71. These words had i) a non-aspirated stop initially, i.e. *tareri* 'snake', *tumina* 'money' etc.; ii) an aspirated stop initially i.e. *phiran* 'take off', *thiren* 'eat'; iii) the five vowels /i, e, a, o, u/ word initially and iv) each vowel following both types of stops.

Methodology

We analyzed all data in PRAAT programme and i) measure the length of the stop's aspirated feature; ii) find the middle part of the vowel (/i, a, u/) that followed the aspirated stop so as get its F1 and F2; iii) got some mean values by gender, dialect, speaker and age.

Results

Men from the lake dialect do an aspiration longer than that of men from the mountain dialect (Table 1). The opposite pattern applies for the mountain woman who does a slightly longer aspiration than the woman from the lake (Table 2).

Table 1. Men: different aspiration length in dialects.

Stop	Lake dialect	N	mountain dialect	N
ph	0.134	11	0.065	4
th	0.108	7	0.068	4
kh	0.155	15	0.084	8

Table 2. Women: different aspiration length in dialects.

Stop	Lake dialect	N	mountain dialect	N
ph	0.072	4	0.090	3
th	0.096	3	0.101	5
kh	0.106	7	0.105	4

There are also dialectal differences based on men's F1 and F2 (table 3). The man's F2 from the mountain dialect is lower than men's F2 from lake, except for vowel /i/ following the three aspirated stops /ph, th, kh/ and one instance of the vowel /u/ following an aspirated velar /kh/. There is a much transparent dialectal difference between women. F2 values from the mountain dialect are lower than F2 from the lake dialect, except for vowel /u/ following a velar /kh/ (table 4).

Table 3. >en: difference in F1 and F2 values in dialects.

Stop	Vowel	Lake dialect		N	Mountain dialect		N
	/a/	F1	F2		F1	F2	
ph		917	1457	3	745	1298	2
th		697	1611	2	752	1599	1
kh		786	1844	4	760	1466	1
	/i/						
ph		375	2084	2	376	2175	1
th		359	2010	4	372	2144	2
kh		426	1956	3	370	2142	3
	/u/						
ph		597	1746	6	443	1182	1
th		838	1869	1	711	1161	2
kh		560	1748	8	618	1857	3

Table 4. Women: difference in F1 and F2 values in dialects.

Stop	vowel	Lake dialect		N	Mountain dialect		N
	/a/	F1	F2		F1	F2	
ph		897	1762	2	953	1586	1
th		935	1891	2	835	1523	1
kh		807	1847	2	801	1779	1
	/i/						
ph		376	2611	1	390	2152	1
th		442	2675	2	419	2338	2
kh		485	2531	2	494	2196	2
	/u/						
ph		496	949	1	600	869	1
th		813	1423	1	714	1187	3
kh		383	936	2	470	1136	1

Regarding generational differences, a young man does a shorter aspiration than older men (table 5), while there is not a clear pattern for women of different age (table 6).

Table 5. Men generational differences in aspiration length.

Stop	young man 25 y.ld	N	men 60-70 y.ld	N
ph	0.082	5	0.102	16
th	0.091	6	0.099	12
kh	0.087	6	0.135	19

Table 6. Women generational differences in aspiration length.

Stop	girl 23 y.ld	N	women 45-47 y.ld	N
ph	0.079	8	0.068	6
th	0.069	1	0.098	8
kh	0.102	6	0.089	10

So far we can partially answer our questions: what are some dialectal differences? There are some dialectal differences given by the measurement of the aspirated feature in stops. The man from the mountain dialect does a shorter aspiration, while the women from the same dialect does a slightly longer aspiration than their respective gender speakers from the lake dialect. Another difference is given by the F2 value. Both, the woman and the man from the mountain tend to have an F2 lower in all vowels following aspirated stops, except for the man's F2 for vowel /i/ which is higher than that of the lake dialect.

Is there any generational difference? The young man's F1 and F2 values tend to be higher than the old men's, while it is the opposite pattern for the girl whose F1 and F2 values are lower than older women.

The length of the aspirated feature has some consequences for the hearer since the longer the aspiration the more perceptible it is. Probably Friedrich (1971) had not perceived this difference and made him state that a lake dialect was losing its aspiration contrast with stops. The vowel's F1 and F2 also show generational differences that have not been considered before in this language. Although all these results are based on few speakers, future data is needed to confirm or debunk the above claims.

References

- Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5:9/10, 341-345.
- Chamoreau, C. 2005. Dialectología y dinámica: reflexiones a partir del p'urépecha. *Trace*, 47, 61-81.
- Chávez, R.D.A. 2004. Aproximación a la dialectología de la lengua purépecha. MA thesis, México, CIESAS.
- Foster, M.L. 1969. *The Tarascan Language*. Berkley-Los angeles, University of California press.
- Friedrich, P. 1971. Dialectal variation in tarascan phonology. *IJAL*, 3, 32, 164-187.
- Velásques, G.P. 1987. *Diccionario de la lengua phorépecha*. México, Fondo de cultura económica.

Boundary-related durations in Modern Greek

Evia Kainada

Department of Linguistics and English Language, University of Edinburgh, UK

Department of Linguistics and Phonetics, University of Leeds, UK

Abstract

This paper investigates the influence of prosodic boundary strength on the temporal organisation of speech. It tests whether Modern Greek (MG) exhibits pre- and post-boundary lengthening, and whether the lengthening effect decays gradually the further away from the boundary, or is attracted to stressed syllables, even if those are not in the proximity of the boundary. It is shown that MG exhibits pre-boundary lengthening, which is at its strongest on the pre-boundary syllable, followed in strength by the stressed syllable, and is least pronounced on the intervening unstressed syllables. No clear post-boundary lengthening effects were identified.

Key words: prosodic structure, pre- and post-boundary lengthening, Modern Greek

Introduction

It is a well-attested phenomenon that the duration of segments at the vicinity of prosodic boundaries is influenced by the strength of the boundary, i.e. longer durations are generally found in pre- and often post-boundary positions close to stronger prosodic boundaries (Wightman et al. 1992, Cho & Keating 2001). This phenomenon is known as *pre-* and *post-boundary lengthening*.

The first question addressed in this study is whether pre- and post-boundary lengthening is attested in MG, as indirectly suggested by previous research (e.g. Baltazani 2006). In post-boundary positions, Botinis (1989) and Arvaniti (2000) have shown that low order prosodic constituents do not exhibit post-boundary lengthening, but there has been no research testing the effect in higher order domains.

The theory of π -gestures (Byrd & Salzman 2003) proposes that the boundary strength slows down the movements in the vicinity of the boundary, therefore the effect is at its strongest closer to the boundary and cannot skip syllables. Turk & Shattuck-Hufnagel (2007), on the other hand, reported that stressed syllables could attract a boundary-related lengthening, irrespectively of whether unstressed syllables intervene between the prominent syllable and the pre-boundary one. The second question of this paper is whether the stressed syllable of the pre-boundary word attracts a lengthening effect or not in MG when further away from the boundary.

Method**Materials**

The boundary strength effect was tested on the duration of pre- and post-boundary syllables. The materials consisted of ADJECTIVE # NOUN structures (#=prosodic boundary of varying strength). The pre-boundary word was tetrasyllabic stressed on the antepenultimate syllable, and the post-boundary word was at least trisyllabic stressed on the second syllable. Sentences were designed to manipulate the boundary strength by changing the syntactic affiliation of the words, and the length of the phrases (8 items*5 boundary conditions*2 repetitions*5 speakers = 400 utterances). Following is an example of the boundary conditions (BC1-5 in increasing size):

- BC1 [a'po a'ftin tin **ci'notopi** # **ko'pela**, to pe'rimenta]
 from this the common girl it expected_{PASTIMP-1S}
 'From this common girl, I expected it'.
- BC2 [me em'fanisi **ci'notopi**, # **ko'pela** ðen Θa vri]
 with appearance common girl not will_{AUX}find_{FUT-3S}
 'With a common appearance, he will not find a girl'.
- BC3 [a'fu 'eçis em'fanisi **ci'notopi**, # **ko'pela** ðen Θa vris]
 since have_{PRES-2S}appearance common girl not will_{AUX}find_{FUT-2S}
 'Since you have a common appearance, you will not find a girl'.
- BC4 [a'fu 'eçis ka'ta jeni'ki omolo'jia em'fanisi **ci'notopi**, # **ko'pela** ðen Θa vris
 'xoris rizi'ki ala'ji]
 since have_{PRES-2S} by general consent appearance common girl not will_{AUX}find_{FUT-2S}
 without radical change
 'Since by general consent you have a common appearance, you will not find a girl without a radical change'.
- BC5 ['eçis ka'ta jeni'ki omolo'jia em'fanisi **ci'notopi**, # **ko'pela** ðen Θa vris
 'xoris rizi'ki ala'ji]
 have_{PRES-2S} by general consent appearance common girl not will_{AUX}find_{FUT-2S}
 without radical change
 'By general consent you have a common appearance. You will not find a girl without a radical change'.

Participants and Recordings

Five native speakers of MG, two male and three female, were recorded. They were all in their twenties at the time of the recordings and were students of the University of Edinburgh, having lived in the UK for a period of time no longer than a year. None reported a speech, hearing, or reading difficulty. Recordings took place at the recording booth of the University of Edinburgh.

Results

Results on the duration of the ultimate, penultimate and antepenultimate syllable of the pre-boundary word are reported, and of the first post-boundary syllable (measurements for the onset, rime and whole syllable). The effect on their duration was tested using repeated measures ANOVAs. Boundary strength was found to be significant on the onset, rime and whole syllable of the antepenultimate, penultimate and ultimate syllables of the pre-boundary word [ULTIMATE syllable: onset: $F(4,244)=37.4$, $p<.001$, rime: $F(4,244)=43.4$, $p<.001$, whole syllable: $F(4,228)=53.6$, $p<.001$, PENULTIMATE syllable: onset: $F(4,160)=10.7$, $p<.001$, rime: $F(4,132)=4.7$, $p<.001$, whole syllable: $F(4,132)=14.3$, $p<.001$, ANTEPENULTIMATE syllable: onset: $F(4,128)=6.38$, $p<.001$, rime: $F(4,128)=19.55$, $p<.001$, whole syllable: $F(4,128)=22.2$, $p<.001$]. In post-boundary position, the boundary strength effect was found to be significant only for the onset of the post-boundary syllable ($F(3,180)=3.76$, $p=.012$), however no clear cumulative pattern was found (duration of onset: BC1=70ms, BC2=73ms, BC3=77ms, BC4=75ms, BC5=68ms, but see also Katsika 2009).

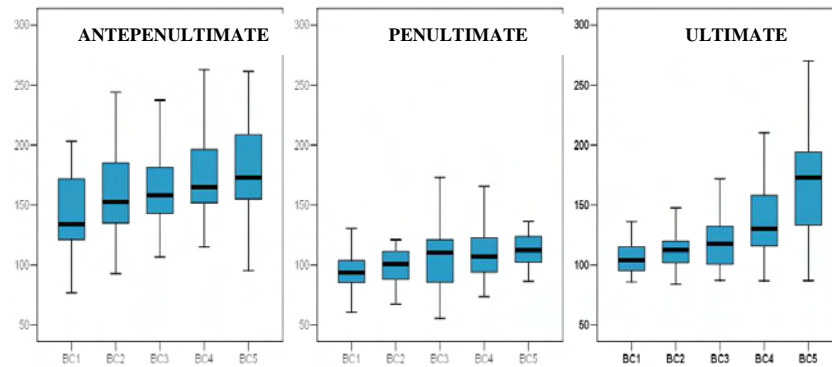


Figure 1. Duration (ms) of antepenultimate (stressed), penultimate and ultimate syllables of pre-boundary word for all boundary conditions (BC1-5).

The magnitude of the boundary strength effect was then examined on each syllable; the percent lengthening of BC2-BC5 with respect to BC1 was calculated ($BC_{2345}lengthening = [(BC_{2345}duration - BC1duration) * 100] / BC1duration$) and was found to be significant for all three syllables [ULTIMATE $F(3,171)=43.191$, $p<.001$, PENULTIMATE $F(3,99)=5.97$, $p=.001$, ANTEPENULTIMATE $F(3,96)=9.5$, $p<.001$]. It was most pronounced for the pre-boundary syllable: the percent lengthening starts from about 10% (BC1 to BC2) and reaches up to 60% (BC1 to BC5). The antepenultimate

(stressed) syllable followed in terms of magnitude of lengthening (15% from BC1 to BC5 reaching up to 31% for BC1 to BC5), and it was the least pronounced for the intervening unstressed syllable (8% for BC1 to BC2 reaching up to 22% for BC1 to BC5).

Discussion

This paper showed that MG exhibits a clear pre-boundary lengthening pattern. In post-boundary position the result is less clear showcasing a possible asymmetry between the two positions. Importantly, the effect of boundary strength in MG seems to support Turk & Hufnagel's (2007) finding that boundary-induced lengthening seems to be attracted to prominent syllables, and is not necessarily in line with the prediction by the π -gestures theory (Byrd & Saltzman 2003). However, the present paper did find a pre-boundary lengthening effect in the intervening unstressed syllable, which was not as pronounced as the one on the stressed syllable.

Acknowledgements

I would like to thank Prof. D. R. Ladd and Prof. A. Turk for their constant support and comments throughout my research. This study was funded by the Economic and Social Research Council, UK.

References

- Arvaniti, A. 2000. The phonetics of stress in Greek, *Journal of Greek Linguistics* 1, 9-39.
- Baltazani, M. 2006. Focusing, prosodic phrasing, and hiatus resolution in Greek, in L. Goldstein, D. Walen & C. Best, eds, *Proceedings of Laboratory Phonology VIII*, Mouton de Gruyter, Berlin, pp. 473-494.
- Botinis, A. 1989. *Stress and Prosodic Structure in Greek*, Lund University Press.
- Byrd, D., Saltzman, E. 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening, *Journal of Phonetics* 31(2), 149-180.
- Cho, T., Keating, P. 2001. Articulatory and acoustic studies on domain initial strengthening in Korean, *Journal of Phonetics* 29, 155-190.
- Katsika, A. 2009. The interaction of boundaries and prominence in Greek, poster presentation at the 157th Acoustical Society of America, May 2009.
- Nespor, M., Vogel, I. 1986. *Prosodic Phonology*, Foris Publications.
- Turk, A., Shattuck-Hufnagel, S. 2007. Multiple targets of phrase-final lengthening in American English words, *Journal of Phonetics* 35, 445-372.
- Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M., Price, M.J. 1992. Segmental durations in the vicinity of prosodic phrase boundaries, *Journal of the Acoustical Society of America* 91, 1707-1717.

Action and object naming in mono- and bilingual children with language impairment

Maria Kambanaros¹, Kleanthes K. Grohmann², Eleni Theodorou²

¹Speech and Language Therapy Program, European University Cyprus, Cyprus

²Department of English Studies, University of Cyprus, Cyprus

Abstract

Specific Language Impairment (SLI) is a severe limitation in language ability in the absence of other factors that typically accompany language problems (e.g., hearing impairment, low non-verbal IQ, neurological damage). SLI is the most common and most studied type of developmental language disorder, yet research comparing bi- and monolingual development is surprisingly lacking, leaving potential implications of bilingualism for children with language disabilities an under-explored area. It is known that monolingual children with SLI are less accurate at naming pictures of objects than age-matched peers with typical language development (TLD), strongly suggesting that difficulties with lexical access are related to a breakdown at the level of the phonological word form, which the present study clearly support.

Key words: Cypriot Greek, mono- and bilingual first language acquisition

The Study

Four groups of children participated in this study:

- (i) 30 monolingual children with TLD (15 girls, 15 boys), aged 6;0–6;11 years (mean: 6;3) — all children were recruited randomly from three public primary schools in the Nicosia district after approval from the Ministry of Education and upon written parental consent, and no child was receiving speech and language therapy service.
- (ii) 12 monolingual children (3 girls, 9 boys), aged 5;5–9;9 years (mean: 6;9) — they were all diagnosed with SLI and recruited from speech-language therapists in private practices (MoSLI group).
- (iii) 4 bilingual children (2 girls, 2 boys), aged 6;6–8;10 years (mean: 7;4) — they were all diagnosed with SLI and recruited from speech-language therapists in private practices (BiSLI group).
- (iv) 10 younger children with TLD (2 girls, 8 boys) aged 3;5–5;2 years (mean: 4;4) — all children were recruited from a kindergarten in the Nicosia district after approval from the Ministry of Education and upon written parental consent, and no child was receiving speech and language therapy service.

Subject selection criteria included:

- a Greek Cypriot, monolingual Cypriot Greek-speaking background for the monolingual children
- a bilingual background for the bilingual children where Cypriot Greek was one of the languages spoken (either L1 or L2)
- no history of neurological, emotional, or behavioural problems
- no obvious learning difficulties (teacher report)
- no gross motor difficulties
- hearing and vision adequate for test purposes
- normal articulation
- normal performance on screening measures of non-verbal intelligence (or as reported by school psychologist)
- medium-high socio-economic status

The Cypriot Object and Action Test (COAT) of Kambanaros et al. (in submission), adapted to Cypriot Greek from the GOAT for Standard Modern Greek (Kambanaros 2003), was administered to assess retrieval of object and action names. For the present study, 74 coloured photographs measuring 10x14cm in size were used, 39 depicting actions (verbs) and 35 objects (nouns). Object names are single, concrete inanimate nouns and include manipulated instruments such as garage tools, garden equipment, kitchen utensils, household items, and office and personal implements, used for activities of daily living. All verbs were monotransitive with either simple internal word structures of [root + affix] or more complex ones of [root + affix + affix]. Actions were restricted to past stereotypical roles, that is, a woman is shown performing household activities (e.g. sweeping), for example, and a man is performing more “manly” duties (e.g. hammering). All action names corresponded to either an instrumental verb (where an instrument is part of the action, e.g. cutting) or to a non-instrumental verb (e.g. climbing). All target nouns in object naming were also items in the noun comprehension task. All target verbs in action naming were also targets in the verb comprehension task.

Table 1 summarizes the characteristics of items in each word class.

Table 1: Characteristics of items in each word class.

	Lemma frequency	Syllable length	AoA	Image-ability	Picture complexity
Objects	40.91	2.88 (0.803 SD)	2.98 (0.76 SD)	6.49 (0.49 SD)	6.49 (0.28 SD)
Actions	40.11	2.95 (0.731 SD)	2.82 (0.58 SD)	6.42 (0.16 SD)	6.16 (0.67 SD)

Lemma frequencies for object and action names were calculated based on the printed word frequency count for Standard Modern Greek (Hatzigeorgiou et al. 2000); at this time, there are no word frequency data available for Cypriot Greek. A Mann-Whitney test revealed no significant difference between object and action lemmas ($z = -0.154$, $p = 0.878$). In addition, there was no significant difference in syllable length between object and action names either ($z = -0.610$, $p = 0.542$). Furthermore, object and action names were measured for key psycholinguistic variables, including age of acquisition (AoA), imageability, and picture complexity.

The object and action tasks were presented in one or two sessions. Testing was conducted in a quiet room at the school. Each child was tested individually.

Children were asked to name the object or action represented in the photograph in a single word. Action names were required in the third person singular. Two examples were provided before testing. The stimulus question(s) was repeated once for children who did not respond. If no response was given, the item was scored as incorrect. No time limits were placed and self-correction was allowed. Responses were recorded and transcribed verbatim by the first-named author and checked by the second.

Results

The results of two subtests of the COAT are reported in the present study: object/noun naming and action/verb naming. The percentages of correct responses were calculated for object and action names as provided by all children. A summary of the results is given in Table 2 according to picture type.

Table 2: Correct production percentages for object and action names.

	MoSLI	BiSLI	yTD	oTLD
Object names	72%	65%	63%	85%
Actions names	65%	55%	64%	77%
Picture naming (mean)	68%	60%	63.5%	

Although the MoSLI group have higher percentages correct than the BiSLI group, this difference failed to reach significance using the Mann-Whitney U-test.

A one-way analysis of variance (ANOVA), carried out on the results from verb and noun naming performances between the four groups, revealed a statistically significant difference between the children with TLD and those with SLI — with the latter showing significantly more difficulties retrieving object and action names compared to typically developing peers.

Discussion

The present study investigated object and action picture naming accuracy in four groups of Cypriot Greek-speaking children: four- and six-year-olds with TLD, monolingual and bilingual children with SLI, in a highly inflected language (Cypriot Greek, patterning morphosyntactically for all items tested just as Standard Modern Greek). Modern Greek (whether Cypriot Greek or Standard Modern Greek) is a highly inflected language where nouns and verbs are clearly differentiated on the basis of inflectional suffixes (Holton et al. 1997).

Generally, children with SLI are less accurate in (object and action) naming than both the age-matched and the younger children with TLD, but interestingly, error type cannot differentiate the two groups. This suggests strongly that children with SLI are delayed — but *not atypical*. The fact that both groups of children with SLI also showed no grammatical-class effect fully supports our initial hypothesis that Cypriot Greek children with SLI would show an undifferentiated grammatical class dissociation.

The bilingual children with SLI (albeit 4 only) did not show a significant difference in naming accuracies for action and object names compared to their monolingual counterparts with SLI. This finding is in line with research indicating that bilingualism does not impact negatively on children affected with SLI (see Paradis et al. 2003). In other words, the outcome of SLI children learning two languages for verb and noun retrieval at the single word level revealed no significant differences between the bilingual and monolingual SLI groups. Our results are more consistent with the predictions of a representational than processing account of SLI suggesting that difficulties retrieving verbs and nouns in SLI is internal to the linguistic system.

References

- Hatzigeorgiou, N., Gavrilidou, M., Piperidis, S., Carayannis, G., Papakostopoulou, A., Spiliotopoulou, A., et al. 2000. Design and implementation of the online ILSP corpus. *Proceedings of the Second International Conference of Language Resources and Evaluation (LREC)*, vol. 3, 1737–1740.
- Holton, D., Mackridge P., Philippaki-Warbuton, I. 1997. *Greek: A Comprehensive Grammar of the Modern Language*. London: Routledge.
- Kambanaros, M. 2003. *Verb and noun processing in late bilingual individuals with anomic aphasia*. Flinders University, Adelaide: PhD dissertation.
- Kambanaros, M., Grohmann, K.K., Michaelides, M. in submission. Picture action and object naming in Cypriot Greek first graders.
- Paradis, J., Crago, M., Genesee, F., Rice, M. 2003. Bilingual children with specific language impairment: How do they compare with their monolingual peers? *Journal of Speech, Language and Hearing Research* 46, 1–15.

Is that a *bnik* I see? Testing phonotactics using word-picture matching

Vsevolod Kapatsinski, Lamia Haddad Johnston
Department of Linguistics, University of Oregon, USA

Abstract

Phonotactic knowledge is typically tested using metalinguistic wordlikeness judgment tasks. We introduce a new method for testing phonotactics, where subjects are asked to match a set of pictures of novel objects with a larger set of pseudowords, only some of which are phonotactically legal. The subjects tend to pick the pseudowords that are phonotactically legal in their language and leave the illegal ones unassigned. We compare the results of the word-picture matching to a traditional rating task and show that the two tasks produce somewhat different results despite sharing some (25%) of the variance. We argue for methodological pluralism in the study of phonotactics.

Key words: phonotactics, word-picture matching, rating.

Introduction

Speakers' knowledge of the phonotactic constraints of their language affects the development of the lexicon in the domains of borrowing, coining, and word retention over time. This can happen through listeners misperceiving phonotactically illegal words as similar-sounding legal words (Berent et al. 2007, Ohala 1981) or speakers mispronouncing phonotactically illegal words. In addition, even if a word *can* be correctly perceived and produced, speakers may minimize effort by choosing to produce or adopt phonotactically legal words over phonotactically illegal competitors (Berg 1998, Martin 2007, Schwartz & Leonard 1982).

The main method for testing phonotactic knowledge, at least in adults, is a rating task, where subjects are asked to judge the acceptability or 'wordlikeness' of various pseudowords (e.g., Bailey & Hahn 2001, Frisch et al. 2000, Shademan 2005, Treiman et al. 2000). Although recent studies have developed methods for examining the effect of phonotactics on misperception (Berent et al. 2007), the effect of phonotactics on lexical choice remains unexplored experimentally. In this paper, we propose one method (word-picture matching) for testing an influence of phonotactics on lexical choice. We ask subjects to select from a set of words and assign them to a smaller set of referents. While requiring much more time than a ratings task, the method avoids explicit solicitation of phonotactic judgments and provides a way to examine other potential factors that influence lexical choice, including sound symbolism (Sapir 1929) and lexical analogy (Bailey & Hahn 2001, Shademan

2005). Using the same set of stimuli, we will compare the results of a ratings task with our novel picture-word matching paradigm.

Methods

40 native English speakers were presented with a series of Powerpoint slides. The first slide contained a short background story and instructions. The story asked subjects to imagine arriving on another planet in the distant future and discovering the remains of an Earth colony established by speakers of English and Wilkipaengo (an invented language name). They were told that they would hear lists of both English and Wilkipaengo names for creatures along with creatures' pictures and they were instructed to assign English words to the creatures.

The following six slides each contained 6 creatures and 12 sound files of pseudowords (6 legal and 6 illegal, equally balanced between /i/ and /a/ arranged in a column on the left side of the slide) that the subjects could listen to as many times as necessary by double-clicking on them. The subjects assigned a name to a creature by dragging a sound file icon on top of the chosen creature. Thus, we were able to avoid presenting spellings of the words.

For each phonotactically illegal word, there was a corresponding phonotactically legal word. The two words in a pair differed from each other only by the onset, which did or did not contain an illegal cluster (e.g., blick/bnick, fnek/frek, sren/fren). The two words in a pair were presented to different subjects but on the same slide and in the same location within the slide.

An additional group of 23 native English speakers were asked to rate the same stimuli on a scale from 1-5 for how much each word sounded like a typical English word. If the word sounded completely normal, the subjects were to pick 5, if it sounded "totally weird", they were to pick 1.

Results

Phonotactically legal words were significantly more likely to be assigned to creatures than the corresponding phonotactically illegal words: $t(71)=8.05$, $p<.0001$ (in Figure 1, there are more words above the horizontal axis than below it), and were rated as sounding more like typical English words: $t(71)=5.09$, $p<.0001$ (in Figure 1, there are more words to the right of the vertical axis than to the left of it).

Figure 1 shows that there is a highly significant correlation between popularity in being assigned to creatures and being rated as sounding typically English ($r=.5$, $t(70)=4.78$, $p<.0001$). However, the correlation explains only 25% of the variance, leaving 75% of the variance between tasks unaccounted for. We tested whether such a difference would be observed between two

different groups of subjects exposed to the same task using resampling. We drew a random sample of 10 subjects from a population of 20 subjects with the same distribution of scores as in each of our actual samples of subjects exposed to a certain set of stimuli in a certain task. We then correlated the scores for words across subject samples within a task and across tasks. The correlation coefficients between samples that came from within the same task were reliably higher than the coefficients for samples coming from different tasks ($p < .001$), indicating that the two tasks reliably differ in which words they favor. Based on the confidence interval on the difference between within-task and across-task correlation coefficients, at least 4% of the variance in the data is due to the difference between tasks rather than random between-subject differences (compared to 25% of shared variance across tasks explainable by phonotactics).

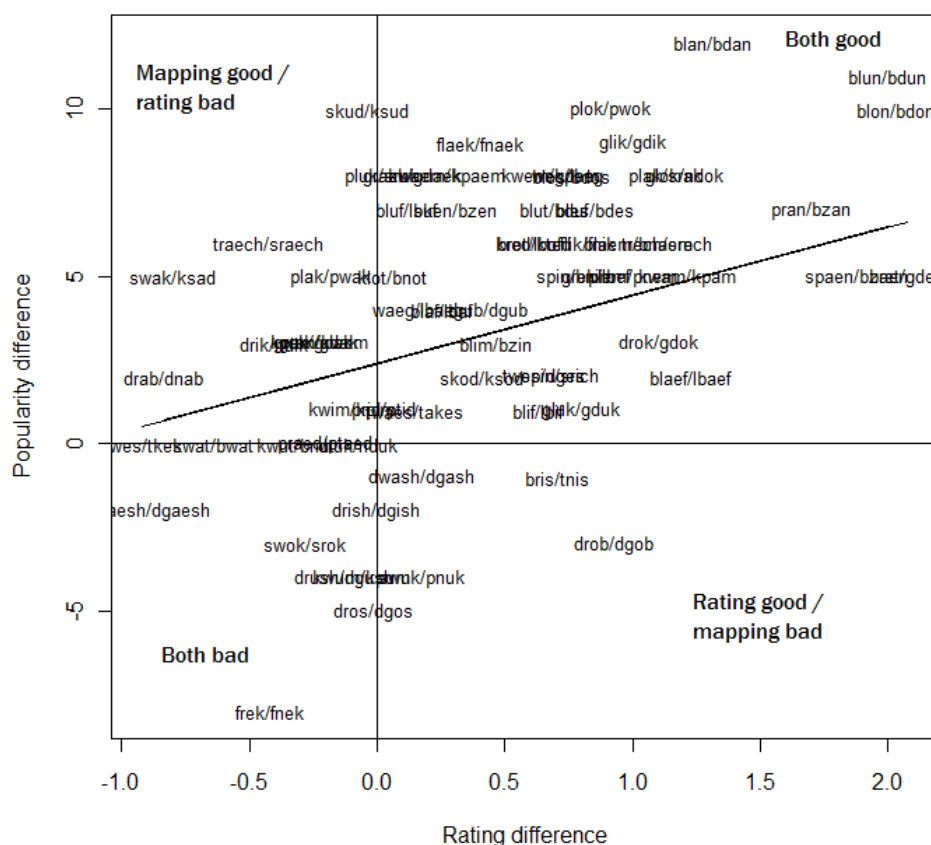


Figure 1. Standardized ratings vs. popularity in word-picture mapping.

Overall, the two tasks are comparable in the number of word pairs in which the illegal word is unexpectedly preferred to the legal one. In both tasks, such

exceptional words tend to involve legal #Cw clusters, which are very low in type frequency in English, and #dr clusters, which are strongly affricated in the stimuli and may be perceived as the illegal #dʒr by the participants who do not affricate as much. In addition, the exceptional preference for /fnek/ (the most popular and highly-rated illegal word) over /frek/ appears to be due to lexical analogy. While this would just be speculation based on rating data, the mapping task allows us to test for non-random picture-word co-occurrences. In this experiment, /fnek/ was mapped onto the only horizontally elongated legless creature on the slide, thus its high popularity and ratings are likely due to the word 'snake'.

Conclusion

We confirm English speakers' sensitivity to the phonotactics / frequency of occurrence of onset clusters in a novel task, auditory word-picture matching, which allows experimental testing for the influence of the subjects' phonotactic knowledge on lexical choice and provides possible explanations for unexpected results due to lexical analogy. A word that is phonologically similar to an existing word in English, may be rated as acceptable, even if a portion of the word violates English phonotactics; analogies that are unclear in a ratings class can surface when we examine the assigned referents in a word-picture matching task.

References

- Bailey, T., Hahn, U. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory & Language* 44, 568–591.
- Berent, I., Steriade, D., Lennertz, T., Vaknin V. 2007. What we know about what we have never heard. *Cognition* 104, 591–630.
- Berg, T. 1998. *Linguistic Structure and Change*. Oxford University Press.
- Frisch, S. A., Large, N. R., Pisoni, D. B. 2000. Perception of wordlikeness. *Journal of Memory & Language* 42, 481–496.
- Martin, A. 2007. *The Evolving Lexicon*. PhD Dissertation, UCLA.
- Ohala, J.J. 1981. The listener as a source of sound change. *Chicago Linguistic Society* 17-2, 178–203.
- Sapir, E. 1929. A study in phonetic symbolism. *J. of Exper. Psychology* 12, 225–239.
- Schwartz, R. G., Leonard, L. B. 1982. Do children pick and choose? *Journal of Child Language* 9, 319–336.
- Shademan, S. 2005. Is phonotactic knowledge grammatical knowledge? *Proceedings of the West Coast Conference on Formal Linguistics* 25, 371–379.
- Treiman, R., Kessler, B., Knewasser, S., Tincoff, R., Bowman, M. 2000. English speakers' sensitivity to phonotactic patterns. *Laboratory Phonology* 5, 269–282.

Embedded aspect in L2 acquisition: Evidence from L1-Russian learners of Greek

Sviatlana Karpava, Kleanthes K. Grohmann

Department of English Studies, University of Cyprus, Cyprus

Abstract

This work investigates first-language (L1) influence on second-language (L2) acquisition of embedded aspect, comparing participants with homogeneous L1 background (Russian) in Greece (L2 Standard Modern Greek) and Cyprus (L2 Cypriot Greek), where verb complementation takes a finite form instead of an infinitival as in Russian. The focus of the experimental study is on those embedded sentential environments which allow only perfective aspect in Greek but either perfective or imperfective in Russian. The findings are taken to support the Full Transfer/Full Access Hypothesis, according to which aspect is part of Universal Grammar; L2 learners can reach native-like attainment due to access to UG, while at the initial stage of L2 acquisition transfer from L1 into L2 takes place.

Key words: acquisition, cross-linguistic interference, embedded aspect, imperfective, perfective.

Introduction

According to the Interpretability Hypothesis (Tsimpli 2003), aspect is a grammaticalized, interpretable feature in Greek relevant to the syntax–semantics and the syntax–discourse interfaces. This study investigates L1 influence on L2 acquisition of aspect, comparing participants with homogeneous L1 background (Russian) in Mainland Greece (L2 Standard Modern Greek) and Cyprus (L2 Cypriot Greek).

In both varieties of Modern Greek, Standard Modern Greek (SMG) and Cypriot Greek (CG), verb complementation takes a finite form instead of an infinitival form as in Russian. The particle *na*, traditionally analyzed as a subjunctive marker (Veloudis and Philippaki-Warbuton 1983, Roussou 2009), introduces subjunctive clauses; the *na*-clause is a complement clause controlled by the main verb. Aspect in the subjunctive subordinate *na*-clauses depends on the kind of verb in the main clause (Malagardi 1993). According to Moser (1993), there is an interaction between lexical and grammatical aspect that influences the aspect of the embedded verb in *na*-clauses. Four lexical aspects (states, activities, achievements, and accomplishments) interact with two grammatical aspects (perfective and imperfective).

In Russian, if the complementation is infinitival, the subject of the main clause and the embedded clause should be the same. If the complementation

is with a finite verb, the complementizer *čtoby* ‘in order to’ is used; there is a restriction on tense of the embedded finite verb: The verb should be only in the past, and the subjects of the main and embedded clauses should be different.

The Study

276 participants took part in the study: 75 bilingual Russian–CG speakers (25 adults, 50 children), 63 bilingual Russian–SMG speakers (13 adults, 50 children), and two control groups of 75 monolingual SMG speakers (25 adults, 50 children) and 63 monolingual CG speakers (25 adults, 38 children).

The participants were offered a sentence picture-matching task and an elicited production task. They were expected to choose (i) perfective in embedded environments which allow only perfective aspect in Greek but either perfective or imperfective in Russian: with an accomplishment main verb that has a continuous interpretation (*prospathusa* ‘try’), a perfective non-ingressive state main verb (*fovithika* ‘be afraid’), a perfective ingressive state main verb (*boresa* ‘be able’), a perfective volitional verb (*ithela* ‘want’), and an activity verb expressing a purpose or a goal with *na* being a short form of *gia na* ‘in order to’; and (ii) imperfective aspect of the complement verb in sentences where both Russian and Greek allow only imperfective: when the main verb is inchoative (*arhise* ‘start’).

Results

Not surprisingly, native speakers performed better than the bilingual Russian–SMG and Russian–CG speakers in both Greece and Cyprus. In addition, and perhaps more surprising, the performance of both monolingual and bilingual groups was better for SMG-speakers in Greece than the corresponding groups of CG-speakers in Cyprus.

According to an ANOVA (means) statistical analysis, there is a statistically significant difference between all groups except of the monolingual SMG and CG: *t*-value -0.363 , *df* = 136, *prob.* = .7174. According to the two-sample *t*-test between percentages, there is a statistically significant difference only between the SMG monolingual and the Russian–CG groups: *t*-value 1.964 , *df* = 136, *prob.* = .05.

In both countries, the test production of monolingual adults was better than bilingual adults. Bilingual Russian–CG adults performed better than bilingual Russian–SMG speakers, while monolingual SMG adults performed better than CG adults.

Table 1. Non-target test production of the groups.

	GREECE		CYPRUS	
Non-target Production	63 Russian–SMG adults + children	63 SMG adults + children	75 Russian–CG adults + children	75 CG adults + children
prospathuse	1.21%	0.33%	1.94%	0.38%
fovithike	3.30%	1.10%	3.66%	2.46%
mborese	0.97%	0.28%	1.35%	0.16%
gia na	1.52%	1.21%	2.48%	1.20%
arhise	1.52%	0.33%	2.09%	0.40%
ithele	0.90%	0.11%	1.16%	0.31%
TOTAL	9.43%	3.37%	12.70%	4.94%

The perfective condition was more problematic than imperfective for non-native groups, presumably since in L1 Russian both perfective and imperfective aspect are possible, so either positive or negative transfer might take place.

The most problematic types of the main verbs for all groups were *fovithike* ‘afraid’, activity verbs with *gia na* ‘in order to’ (perfective condition), and *arhise* ‘start’ (imperfective condition).

In both countries, monolingual children performed better than bilingual children and the test production of SMG and Russian–SMG children was better than that of CG and Russian–CG children.

Table 2. Non-target test production: adults vs. children

	GREECE		CYPRUS	
non-target production	13 Russian–SMG adults	25 SMG adults	25 Russian–CG adults	25 CG adults
TOTAL	28.41%	2.33%	15.7%	4.38%
non-target production	50 Russian–SMG children	38 SMG children	50 Russian–CG children	50 CG children
TOTAL	4.5%	4.05%	11.16%	5.22%

Bilingual children, both in Greece and in Cyprus, performed better than adults, while monolingual adults performed better than children.

L2 children (both Russian–SMG and Russian–CG) get systematic language instruction before or within the critical period; L2 adults on the other hand may not have a systematic language instruction and the language input is far beyond the critical period.

The findings of the study support the Full Transfer/Full Access Hypothesis (Schwartz and Sprouse 1994, Slabakova 2005): Aspect is part of Universal Grammar and L2 learners can reach native-like attainment due to the access to UG, while transfer from L1 into L2 takes place at the initial stage of L2 acquisition. This study provides evidence in support of the Full Transfer/Full

Access Hypothesis by investigating how learners with a homogeneous L1 background acquire L2 aspect in bi-dialectal settings (SMG and CG).

The non-target test production by the Russian–CG group might be explained by the fact that Russian–SMG participants have more years of exposure to L2 in comparison to Russian–CG, or it might be due to the diglossic or bilingual-like situation in Cyprus that influences language acquisition and learning in very interesting ways and puts subjects at a disadvantage similar to what is observed in delayed language development.

References

- Malagardi, I. 1993. Problems of Greek aspect morphology and the identification of projection for Tense and Aspect. In Philippaki-Warbuton, I., Nicolaidis, K. and Sifianou, M. (eds.), *Themes in Greek Linguistics*, 161-169. Amsterdam, John Benjamins.
- Moser, A. 1993. The interaction of lexical and grammatical aspect in Modern Greek. In Philippaki-Warbuton, I., Nicolaidis, K. and Sifianou, M. (eds.), *Themes in Greek Linguistics*, 137-145. Amsterdam, John Benjamins.
- Roussou, A. 2009. Selecting complementizers. *Lingua* 119, 1811-1836.
- Schwartz, B.D., Sprouse, R. 1994. Word order and nominative case in nonnative language acquisition: A longitudinal study of (L1 Turkish) German interlanguage. In Hoekstra, T. and Schwartz, B.D. (eds.), *Language Acquisition Studies in Generative Grammar*, 317-368. Amsterdam, John Benjamins.
- Slabakova, R. 2005. What is so difficult about telicity marking in L2 Russian? *Bilingualism: Language and Cognition*, 8, 63-77.
- Tsimpili, I.M. 2003. Clitics and determiners in L2 Greek. *Proc. of Generative Approaches to Second Language Acquisition*. Somerville, Mass., Cascadia Press.
- Veloudis, I., Philippaki-Warbuton, I. 1983. Η Υποτακτική στα Νέα Ελληνικά [The subjunctive mood in Modern Greek]. *Studies in Greek Linguistics*, 151-169.

Evidence for depression and schizophrenia in speech prosody

Roi Kliper¹, Yonatan Vaizman¹, Daphna Weinshall¹, Shirley Portuguese²

¹The Interdisciplinary Center for Neural Computation and Computer Science
Department, The Hebrew University of Jerusalem, Israel

²Maclean Psychiatric Hospital, Boston, USA

Abstract

We developed automatic computational tools for the monitoring of pathological mental states – including characterization, detection, and classification. We show that simple temporal domain features of speech may be used to correctly classify up to 80% of the speakers in a two-way classification task. We further show that some features strongly correlate with certain diagnostic evaluation scales, suggesting the contribution of such acoustic speech properties to the perception of an apparent mental condition.

Key words: Speech analysis, Schizophrenia, Depression.

Introduction

Changes in the acoustic characteristics of speech prosody in the course of different mental disorders, notably depression and schizophrenia, are a well-documented phenomenon (Alpert, Pouget, & Silva, 2001; Cohen, Alpert, Nienow, Dinzeo, & Docherty, 2008; Michaelis, Frhlich, & Strube, 1998). The evaluation of speech constitutes a standard part of the repertoire of a mental status examination where standard scales include variables such as Alogia, Affective Flattening and Apparent sadness. Psychiatrists who use these scales, report prosody as an important diagnostic tool in their daily work. However, prosody is currently measured by subjective rating scales acquired by highly trained staff, and as a result these clinical ratings tend to be skewed and restricted in range in comparison to computerized acoustic measures, which may lead to rating bias in the global clinical impression.

In an effort to develop automatic computational tools for the monitoring of these pathological mental states, we followed two parallel and tightly connected research tracks: (1) Study the physical signal properties that highlight the examined mental states, (2) Develop an automatic, real-time, reliable and objective assessment tool for diagnostic screening, illness progression monitoring, and response to treatment assessment.

Illness characterization: Data, Feature Extraction

Our data consisted of speech recordings from schizophrenic, depressed and control subjects performing three different tasks: reading a standardized list of words, reading an emotionally neutral passage, and participating in a semi structured clinical interview performed by a clinician. These recordings were used separately to extract various acoustic features that characterize the speech signal. We focused here on features extracted in the temporal domain. Though spectral features have been proven beneficial in various speech analysis tasks, we leave the examination of these features for future research.

Alterations of the speech signal in abnormal conditions can occur at different time-scale levels, including the meso-scale (25 ms to 1 second) and the micro-scale (10 ms or less) levels. By focusing on these scales, we follow reports showing the relevance of meso-scale and micro-scale level to the task of emotion detection. (Cowie & Cornelius, 2003; Rong, Li, & Chen, 2009). Specifically, for each speaking task we used an energy based Utterance-Silence segmentation from which we extracted utterance and gap statistics (e.g. mean utterance duration). We then used the YIN pitch detection algorithm (de Cheveign & Kawahara, 2002) to detect pitched segments (among uttered segments) and to further segment the Uttered segments into Voiced-Unvoiced segments, followed by the extraction of power and pitch related statistics (e.g. Inflection, Emphasis). We then used pitched segments to refer to the finer resolution of ‘cycle to cycle’ variation and extract micro-scale statistics (e.g. Jitter, Shimmer).

Some researches have claimed that speech measurements have arguable validity and poor reliability, pointing in part to the large within subject variability (Rong et al., 2009). This issue is indeed crucial in all applications of speech analysis and poses a major problem for the development of reliable classification tools. Ideally, for the purpose of classification, we would have liked to employ measures which are (speech) task-independent on the one hand and (mental) condition-dependent on the other hand.

We observed that both large-scale and meso-scale measures seem to incorporate a larger variability component due to the specific task (between-task-variability: S_b) as compared to the within speaking-task variability (S_w): $S_b/S_w > 2.8$. This task dependency is significantly reduced for micro-scale measures where one sees the ratio $S_b/S_w < 1.8$ and generally around 1. This observation does not disqualify the larger scale measures from being useful in a classification task (as the task is known in advance); however, it highlights micro-scale features as candidates to be used by a robust general-purpose classifier, in order to achieve task invariance.

To check the reliability of these measures for classification of the different conditions over general speaking tasks, we examined the variance ratio of the between condition vs. within condition. The higher this ratio is, the more discriminative a measure is for classification. While, in general, we did not get high ratios (due to the general signal to noise ratio), the highest ratios were obtained for micro-scale features, notably shimmer, and mean-waveform-correlation. Features that display such robustness (both task invariance and condition dependence) may serve as efficient indicators of a mental illness.

Results: Classification and Correlations

We used the extracted acoustic features and a basic linear classifier (Chang & Lin, 2001) to classify the different conditions: Normal (NL), Schizophrenia (SZ) and Depression (DP) in a two-way classification task. For each classification scenario (e.g. NL vs. SZ) and all relevant recordings of a given task (e.g. List of words), one subject was left out for testing and the rest were used to train a classifier. Table 1 contains the average success rate over all leave-one-out test realizations. Due to the small training sample size, PCA was used to reduce dimensionality of the data-set. As can be seen in the table SZ patients are discriminated nicely from NL with classification rates ranging from 73.81-80.95%. Similar success is achieved for a two-way classification of DP vs. NL. The task of discrimination DP from SZ is reported to be harder but succeeds in two out of three speaking tasks.

Table 1: Classification Results:

List of Words (n(NL)=20, n(SZ)=22, n(DP)=20), Passage (n(NL)=11, n(SZ)=10, n(DP)=15), Interview (n(NL)=33, n(SZ)=31, n(DP)=29)

	NL vs. SZ	NL vs. DP	SZ vs. DP
List of words	0.7381	0.70	0.7143
Passage	0.8095	0.69	0.5200
Interview	0.75	0.8710	0.7667

In addition to the strict labels of mental condition, every subject was clinically evaluated by a psychiatrist using standard psychiatric evaluation scales, for instance: Scale for the Assessment of Negative Symptoms (SANS) for SZ and Hamilton Rating Scale for Depression (HAM-D) for DP. We calculated the correlation coefficients (cc) between the acoustic features extracted from the interview recordings, and the psychiatric evaluations, for the SZ and DP subject groups. Selected results are shown in Table 2.

For SZ, high ratings of negative symptoms correlated with longer gaps and shorter (but denser) utterances and with lower emphasis measures. For

DP, higher depression severity correlated with longer utterances and with lower pitch variability, both in the meso-scale (inflection) and in the micro-scale (jitter). These correlations are a good indication that the extracted acoustic features capture elements of the speech signal which are perceived by the listener as deviating from normality, and thus may contribute to a better identification and characterization of the underlying mental condition.

Table 2: Correlation Results

Schizophrenia Group	
Acoustic feature	SANS Total (cc, p)
Utterance duration	(-0.5953, 0.0004)
Gap duration	(0.4223, 0.0180)
Spoken ratio	(-0.5134, 0.0031)
Fragmented speech	(-0.4831, 0.0059)
Emphasis	(-0.4782, 0.0065)
Depression Group	
Acoustic feature	HAM-D
Utterance duration	(0.4661, 0.0164)
Inflection	(-0.5007, 0.0092)
Jitter	(-0.4555, 0.0194)

Notes

This study was supported by the European Union under DIRAC integrated project IST-027787.

References

- Alpert, M., Pouget, E., Silva, R. 2001. Reflections of depression in acoustic measures of the patient's speech. *Journal of affective disorders* 66, 59-69.
- Chang, C., Lin, C. 2001. LIBSVM: library for support vector machines: Citeseer.
- Cohen, A., Alpert, M., Nienow, T., Dinzeo, T., Docherty, N. 2008. Computerized measurement of negative symptoms in schizophrenia. *Journal of psychiatric research* 42, 827-836.
- Cowie, R., Cornelius, R. 2003. Describing the emotional states that are expressed in speech. *Speech Communication* 40, 5-32.
- de Cheveign , A., Kawahara, H. 2002. YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America* 111, 1917.
- Michaelis, D., Frohlich, M., Strube, H. 1998. Selection and combination of acoustic features for the description of pathologic voices. *The Journal of the Acoustical Society of America* 103, 1628.
- Rong, J., Li, G., Chen, Y. 2009. Acoustic feature selection for automatic emotion recognition from speech. *Information Processing & Management* 45, 315-328.

The effect of cochlear implants on phonological acquisition

Evaggelia Lazarou, Marianna Hatzopoulou

Department of Logopaedics, Athens Metropolitan College, Greece

Abstract

The aim of the present study was to compare the phonological errors that occur in the speech of two 12-year-old Greek children with profound hearing loss. The first child used a cochlear implant (CI) and the second one used hearing aids (HA). The children's phonological development has been assessed using the Phonetic and Phonological Development Test (PPDT), (Levandi et al. 1995)¹. In alignment with previous studies, the child with HA made more phonological errors than the child with CI, even though some unexpected findings were observed regarding the phonological processes used by both children.

Key words: Phonological errors, cochlear implants.

Introduction

Several researches have examined whether deaf children benefit from a CI in speech production, focusing either on the accuracy of articulating specific sounds (consonants, vowels, clusters), or on the phonological processes used. Regarding the accuracy of articulating consonants, Tobey et al. (1991) and Bouchard et al. (2007) reported that the increase in consonant accuracy resulted in greater speech intelligibility of children with CI, whereas Ertmer et al. (1997) and Tye-Murray and Kirk (1993) found improved and more accurate production of back vowels. However, Chin and Pisoni (2000) stated that back vowels were difficult to produce by a child with CI and Chin and Finnegan (2002) observed distortions of the first segment in clusters production. With regard to phonological processes Chin and Finnegan reported that substitutions of the first segment in clusters production and simplification of either the first or the second segment of consonant clusters were the most common patterns in CI children's speech. Chin and Pisoni (2000) found that final consonant omissions were the most frequent speech errors and Bouchard et al. (2007) recorded substitutions of fricatives by stop consonants (stopping). Focusing on consonant clusters, Chin (2007) observed frequent use of voiceless features instead of voiced sounds (devoicing).

The benefit of HA in the phonological development of profoundly deaf children is investigated in a restricted number of studies, due to the fact that the speech production of deaf children with profound hearing loss is usually very limited. Okalidou (2002) as well as Elfenbeim and Hardin-Jones (1994)

reported frequent distortions of sounds, omissions of consonants in initial, medial or final position of words, cluster simplifications and substitutions, such as devoicing, frontalization and voicing in the speech of deaf children with HA. Finally, no studies were found concerning the comparison of speech in children with CI and HA.

Aim

The aim of the present study is to investigate:

- (a) the types of phonological errors made by profoundly deaf children with CI and HA,
- (b) the effect of CI on phonology acquisition of profoundly deaf children in comparison with the use of HA.

Method

The subjects of the present study were two 12-year-old prelingually deaf children exposed to Greek: a boy having a CI at his left ear and a girl using two HA; both children have been attending speech and language therapy sessions for 10 years. Children's performance in naming simple pictures and describing complex ones have been assessed with the PPDT (Levandi et al. 1995). Children's elicited answers were audio-taped, transcribed with the IPA symbols and their phonological errors were measured and analyzed.

Results

The child with CI gave more correct answers than the child with HA in both tasks.

Table 1. Phonological errors made in the tasks of simple and complex pictures.

Participants	Ph. errors in naming of simple pictures		Ph. errors in description of complex pictures	
	No	%	No	%
Child with HA	27/70	38.6%	8/31	25.8%
Child with CI	9/70	12.9%	4/31	12.9%

In particular, in the task of naming simple pictures, the CI child made less errors (12.9%) compared to the errors made by the HA child (38.6%). Similarly, in the task of describing complex pictures the CI child's errors were less (12.9%) compared to those made by the HA child (25.8%). From Table 1 it becomes obvious that the CI child's performance remained stable in both tasks (12.9%), whereas in contrast to the expected results the HA child made less errors in the task of describing complex pictures.

In particular in the task of naming simple pictures, both participants followed similar phonological processes: cluster simplification (HA:10%, CI:2.9%), stopping (HA:7.1%, CI:4.3%), devoicing (HA:7.1%, CI:2.9%) and cluster substitution (HA:1.4%, CI:1.4%). However, it was unexpected that omission of final consonant, was found only in the HA child's production (7.1%) and not in that of the CI child. Furthermore, the CI child made frontalization (1.4%), whereas the HA child did not make any errors of that type.

In the task of describing complex pictures it was observed that both participants followed similar phonological processes: devoicing (HA:6.5%, CI:3.2%) and cluster substitution (HA:3.2%, CI:3.2%). Stopping (9.7%), omission of final consonant (3.2%) and consonant substitution with vowel (3.2%) occurred only in the HA child's answers, whereas cluster simplification (3.2%) was made solely by the CI child. However, palatalization (3.2%) that was observed in the CI child's speech was unexpected.

Discussion

Overall the profoundly deaf child with CI gave more correct answers and produced less phonological errors than the child with HA. Answering the research question, and in alignment with previous studies (Tobey et al. 1991; Chin and Pisoni 2000), the CI seemed to have a positive effect on the child's speech, since the boy made less phonological errors and demonstrated accuracy in the production of vowels and the majority of consonants. With regard to the type of phonological errors, results showed that both participants used devoicing and cluster substitutions in both tasks, which is in accordance with previous studies (Chin and Pisoni 2000; Bouchard et al. 2007; Chin 2007). Interestingly, no omissions of final consonant occurred in the CI child's speech.

Although the CI child has an advantage over the HA child, he still made mistakes. More specifically, the percentage of errors he made remained stable in both tasks (12.9%). The results showed that the disadvantage of the HA child over the CI child was obvious. However, what is challenging is that the HA child had improved her performance (25.8%) in the second task compared to the first task (38.6%). Concluding, due to the limited number of participants, our outcomes are indicative. For generalizing our findings a larger number of profoundly deaf children with CI and HA is needed. Furthermore, the fact that the error categories of frontalization and palatalization did not occur in the speech of children with CI of the present study is a stimulus for further research.

Notes

1. Translation by Lazarou Evaggelia of the Dokimasia Fonitikis kai Fonologikis Exelixis.

References

- Bouchard, M., Le Normand, M., Cohen, H. 2007. Production of consonants by prelinguistically deaf children with cochlear implants. *Clinical Linguistics and Phonetics* 21, 875-884.
- Chin, S.B. 2007. Variation in Consonant Cluster Production by Pediatric Cochlear Implant Users. *Ear and Hearing* 28, 7S-10S.
- Chin, S.B., Finnegan, K.R. 2002. Consonant Cluster Production by Pediatric Users of Cochlear Implants. *The Volta Review* 102, 157-174.
- Chin, S.B., Pisoni, D.B. 2000. A phonological system at 2 years after cochlear implantation. *Clinical Linguistics and Phonetics* 14, 53-73.
- Elfenbeim, J.L., Hardin-Jones, M.A. 1994. Oral communication skills of children who are hard of hearing. *Journal of Speech and Hearing Research* 37, 216-226.
- Ertmer, D.J., Kirk, K.I., Sehgal, S.T., Riley, A.I., Osberger, M.J. 1997. A comparison of Vowel Production by Children with Multichannel Cochlear Implants or Tactile Aids: Perceptual Evidence. *Ear and Hearing* 18, 307-315.
- Levandi, E., Kirpotin, L., Kardamitsi, E., Kabouroglou, M. 1995. Dokimasia Fonitikis kai Fonologikis Exelixis. Omada Erevnas Panelliniou Syllogou Logopedikon.
- Okalidou, A. 2002. Barikoia-Kofosi: Meleti tis paragogis tou logou kai therapeutiki paremvasi. Athina: Ellinika Grammata.
- Tobey, E. A., Pancamo, S., Staller, S.J., Brimacombe, J.A., Beiter, A.L. 1991. Consonant production of children receiving a multichannel cochlear implant. *Ear and Hearing* 12, 23-31.
- Tye-Murray, N., Kirk, K.I. 1993. Vowel and diphthong production by young users of cochlear implants and the relationship between the phonetic level evaluation and spontaneous speech. *Journal of Speech and Hearing Research* 36, 488-502.

Observations on the prosodic characteristics of Finnish-speaking youngsters with Asperger Syndrome

Mari Lehtinen

Department of Modern Languages, University of Helsinki, Finland

Abstract

This paper presents some preliminary results of a post doctoral research project directed towards the prosodic and nonverbal deficiencies of French- and Finnish-speaking youngsters with Asperger Syndrome. More precisely, the paper is focussed on the salient prosodic features that characterise the speech of the Finnish-speaking informants of the project. The term 'prosody' is conceived here in a large sense, encompassing also such paralinguistic features as breathiness, creak, nasalization and whisper (Couper-Kuhlen 2000). The data consists of audiovisual recordings of two neuropsychiatric group therapy sessions where two different groups of 11–13-year-old boys discuss with two therapists and with each other. Methodologically, the study is based on Conversation Analysis and, more precisely, on the interactional approach to the study of prosody.

Key words: Asperger Syndrome, prosody, interaction, conversation analysis

Introduction

Asperger Syndrome (AS) is a form of high-functioning autism which is characterized e.g. by qualitative impairment in social interaction and by stereotyped and restricted patterns of activities and interests (APA 2000). Though persons afflicted with AS generally have no remarkable delay in language acquisition, and their speech typically lacks significant abnormalities, the language use and the prosody of persons with AS is often atypical (Partland & Klin 2006). People with AS often have a limited range of intonation, their speech may be overly fast, jerky or loud, and they also typically have abnormal nonverbal behaviours (such as eye contact, facial expressions, postures and gestures). It has also been shown that people with AS have problems to recognise affective prosody (e.g. Korpilahti et al. 2007) and that they have difficulties to produce affective prosodic patterns (Scott 1985). According to Tantam et al. (Tantam, Holmes & Cordess 1993) the impression of abnormality in AS subjects' interaction results mainly from a lack of integration of expression, speech and gaze.

This study presents some preliminary results of an ongoing post doctoral research project directed towards the prosodic and nonverbal deficiencies of French- and Finnish-speaking youngster afflicted with AS. More precisely, the objective of the current paper consists of providing a brief overview of the salient prosodic features that characterise the speech of the Finnish-

speaking informants of the project. The term ‘prosody’ is conceived here in a large sense, encompassing also such paralinguistic features as breathiness, creak, nasalization and whisper (Couper-Kuhlen 2000), in addition to dimensions of pitch, intensity, speech rate and pauses that are more traditionally qualified as prosodic features.

AS being five times more common for males than for females, all informants of this study are male. The data consists of audiovisual recordings of two neuropsychiatric group therapy sessions where two different groups of 11–13-year-old boys discuss with their therapists and with each other at the Hospital for Children and Adolescents in Helsinki (Finland) in the winter 2009–2010. The duration of each session is two hours. One of the groups consists of three participants (informants) and two therapists, and the other one consists of four participants and two therapists. Most participants of these groups have been diagnosed with AS, but both groups include also one member who has not been officially diagnosed with AS, although according to the therapists these persons have same kind of symptoms as the other members of the group. Each participant had an own microphone behind his ear, and two cameras were used to film the sessions. The filming and the recording were executed by professionals.

Methodologically, the study is based on Conversation Analysis and, more precisely, on the interactional approach to the study of prosody (Couper-Kuhlen & Selting 1996; Couper-Kuhlen 2000).

Prosodic characteristics of the informants

This section will provide a brief overview of some preliminary observations concerning the prosodic characteristics of the Finnish informants of the project. Due to the lack of space, no examples can be given here.

Loudness

As previous studies have already shown (Partland & Klin 2006), AS subjects’ speech can be overly loud. This feature is particularly salient in the case of two informants: the intensity of their speech can be over 80 dB, and even the mean intensity level is about 75 dB. On the other hand, some other informants speak very quietly, at the level of 50–55 dB. (The distance between the microphone and the mouth of the speaker is constant, and it is approximately the same for all participants.)

Prominent word stress

In Finnish, the word stress (primary stress) is fixed: it always falls on the first syllable of the word. The word stress is indicated with a combination of increased intensity, raised F0, longer duration and/or clearer articulation of the syllable (e.g. Iivonen 1998). Generally speaking, the Finnish stress is however relatively weakly expressed (Iivonen 1998). Indeed, one salient

prosodic feature occurring in this data consists of an exceptionally strong word stress. This is particularly clear in the case of the same informants who also speak remarkably loud (cf. sub-section 'Loudness' above). The prominent word stress makes their intonation bounce up and down, which, in turn, creates an effect of peculiarity in Finnish.

On the other hand, the word stress of some informants of the data is very weak. This feature seems to be typical of participants who also talk in a very quiet voice. The combination of these two features – the lack of word stress and a quiet voice – makes the speech of these subjects a bit unclear.

Accelerated speech rate

Partland and Klin (2006) remark that AS persons' speech can be overly fast. In this data, this concerns at least one informant: where the mean speech rate of Finnish is approximately five syllables per second, this informant's speech rate is frequently more than seven syllables per second. His speech is also remarkably loud and characterised by a prominent word stress (cf. sub-sections 'Loudness' and 'Prominent word stress' above).

Monotonous pitch vs. large pitch movements

It has been shown that AS persons often have a limited range of intonation (Partland & Klin 2006). This also seems to be the case of some of the informants of this data: on the whole, participants who have a quiet voice and whose speech carries a weak word stress also use a narrow pitch range.

However, Finnish intonation being on the whole characterized by a certain monotony due to its relatively small pitch intervals and a narrow variation range of pitch movements (e.g. Iivonen 1998), the lack or rarity of large pitch movements does not really constitute a salient feature. The opposite case, in turn, is highly prominent: indeed, one of the informants regularly use a pitch pattern in which the F0 level wanders from the lowest level (H1) of his pitch range (150 Hz – 350 Hz) to the highest level (H4) of the range several times within a few seconds. This makes his speech sound almost like singing. The phenomenon does not appear the same way in other informants' speech, but sudden pitch rises in unexpected places can also be found in other participants' talk.

Creaky voice and other salient features

Creaky voice is a very frequent phenomenon in this data. It occurs repeatedly in the speech of most informants, and there can be surprisingly long sequences produced with a creaky voice. According to Iivonen (1998), creaky voice very often occurs in Finnish in utterance-final positions. In this data, this paralinguistic phenomenon however seems to occur in a large variety of positions and not only at the end of utterances.

Another frequent phenomenon consists of an atypical pausal structure of utterances. Several pauses can occur within one utterance, and there can be

remarkably long pauses inside a turn. Recurrent pauses – especially when associated with accelerations – may contribute to the creation of the impression of “jerky speech” (Partland & Klin 2006). Stretched vowels as well as repeatedly occurring hesitation words and other vocalisations also constitute typical phenomena in the data.

References

- American Psychiatric Association (APA). 2000. Diagnostic and Statistical Manual of Mental Disorders. 4th edition, Text Revision, DSM-IV-TR. Arlington, American Psychiatric Association.
- Couper-Kuhlen, E. 2000. Prosody. In Verschueren, J., Östman, J.-O., Blommaert, J., Bulcaen, C. (eds.) 2000, *Handbook of Pragmatics*, 1-19. Amsterdam/Philadelphia, John Benjamins Publishing Company.
- Couper-Kuhlen, E., Selting, M. 1996. Towards an interactional perspective. In Couper-Kuhlen, E., Selting, M. (eds.) 1996, *Prosody in conversation. Interactional studies*. 11-56. Cambridge, Cambridge University Press.
- Iivonen, A. 1998. Intonation in Finnish. In Hirst, D., Di Cristo, A. (eds.) 1998, *Intonation Systems. A Survey of Twenty Languages*, 311-327. Cambridge, Cambridge University Press.
- Korpilahti, P., Jansson-Verkasalo, E., Mattila, M.-L., Kuusikko, S., Suominen, K., Rytty, S., Pauls, D.L., Moilanen, I. 2007. Processing of affective speech prosody is impaired in Asperger syndrome. *Journal of Autism and Developmental Disorders* 37(8), 1539-1549.
- Partland, J., Klin, A. 2006. Asperger's syndrome. *Adolescent Medicine Clinics* 17(3), 771-88.
- Scott, D.-W. 1985. Asperger's syndrome and non-verbal communication: a pilot study. *Psychological Medicine* 15(3), 683-687.
- Tantam, D., Holmes, D., Cordess, C. 1993. Nonverbal expression in autism of Asperger type. *Journal of Autism and Developmental Disorders* 23(1), 111-133.

Metalanguage or bidialectism? Acquisition of clitic placement by Hellenic Greeks, Greek Cypriots and binationals in the diglossic context of Cyprus

Evelina Leivada, Paraskevi Mavroudi, Anna Epistithiou

Department of English Studies, University of Cyprus, Cyprus

Abstract

Acquisition of object clitics is one of the more investigated aspects of the largely understudied variety of Modern Greek spoken in the Republic of Cyprus. Previous studies on the acquisition of clitics in Cypriot Greek usually acknowledge that the linguistic reality in Cyprus involves a state of diglossia, where the sociolinguistically ‘high’ Standard Modern Greek co-exists with the ‘low’ Cypriot Greek. Acquisition of clitic placement in simple declaratives is here approached through a picture-based elicitation task by testing both varieties in three populations residing in Cyprus — Greek Cypriots, Hellenic Greeks, and binationals — so as to examine implications for bidialectism and its connections with enhanced metalinguistic abilities.

Key words: acquisition, clitic placement, Cypriot Greek, diglossia, metalanguage.

Acquisition of clitics in Cypriot Greek

The sociolinguistic status of the Republic of Cyprus has been linked to trilingualism, bilingualism, or bidialectism (very much depending on the study and the perspective behind it). What is unequivocally admitted is the status of diglossia (Ferguson 1959), and numerous studies have discussed the diglossic status of Cyprus and sociolinguistic aspects of the ‘high’ variety Standard Modern Greek (SMG) and the ‘low’ variety Cypriot Greek (CG) as well as the relationship between them (e.g., Papapavlou and Pavlou 1998).

From formal and developmental perspectives, clitics in CG constitute one of the domains that has enjoyed comparatively rich research. Petinou and Terzi (2002) published findings of a longitudinal study to argue that children at age 3 have mastered general clitic production and placement. A more recent pilot study (Grohmann, to appear) reports on a picture-based elicitation task that investigated the acquisition of object clitics in monolingual Greek Cypriot children, between 3 and 6 years of age, which supports the claim that acquisition of object clitics occurs by age 3 in simple declaratives (i.e. indicative environment). In the same study, adults’ performance shows 100% enclitic production; this would reasonably lead to an assumption of adult CG being exclusively enclitic in indicative environments.

Initial results reported by Grohmann et al. (2010), which lay the basis for the presentation in Grohmann (to appear), seem to confirm this assessment.

They extended the groups of CG-speaking children from 3–4- and 5-year-old children to differentiated groups of equal size for 3-, 4-, 5-, and 6-year-olds respectively (around 20 children per age group, but over twice as many 5-year-olds), as well as a language-impaired group. We will extend the tested groups further and thereby contribute to the study of clitic placement in young speakers by adding another dimension.

The Study

The present experiment readdresses the issue of the acquisition of object clitic placement, yet it does so by approaching it from the other aspect of diglossia in Cyprus, that is, by examining the placement of object clitics in children native in the ‘high’ variety (SMG). These are subsequently referred to as Hellenic Greek children residing in Cyprus (but born and, for a certain period, schooled in Greece, with both parents from Greece) and what we call ‘binational’ children (born in Greece or Cyprus, with one parent from each country). We employed a picture-based elicitation task (from COST Action A33; Varlokosta et al., to appear) in which children had to complete 12 target sentences inside a *because*-clause by producing a verb and a clitic.

The experiment consisted of two parts: (i) an SMG-version, slightly differing from the version of COST Action A33 carried out in Greece, and (ii) the CG-version used by Grohmann et al. (2010). The correct answer was provided, if necessary, only in the 2 warm-up stories. Three groups participated in the present study: (i) 40 monolingual Hellenic Greek children, divided into 4 sub-groups (10 in each of the following age groups: 3;0–4;5, 4;6–5;11, 6;0–7;5, and 7;6–8;11), (ii) 30 binational children (one age group, 3;5–9;1, mean age 6;5), and (iii) 40 Greek Cypriot children, divided into 4 sub-groups (10 in each age group: 3;0–4;5, 4;6–5;11, 6;0–7;5, and 7;6–8;11).

All children were tested twice, once in each (but first in their native) variety, by two different investigators, one SMG-, the other CG-native, and a 7-day period was allowed between the two sessions. The control groups were 6 Hellenic Greek adults and 6 Greek Cypriot adults, also tested twice. Children were recruited randomly from 36 primary and pre-primary schools in the Nicosia district and tested in the school environment.

Findings

Children aged 3;0–4;5 produced at least 7 out of 12 clitics in our experiment, a finding that agrees with Petinou and Terzi (2002) and Grohmann et al. (2010) regarding age of acquisition. In the CG-version, Hellenic Greek children ignored the enclisis feedback that the investigator provided them with in the two warm-up stories and did not code-switch. The Hellenic Greek adults had lived in Cyprus for 2 to 6 years, yet they code-switched, if at all, for only the first few test structures, just to show their ability to produce enclisis (as admitted verbally to the experimenter)), but then

returned to SMG and their proclitic productions. In the SMG-version, all Hellenic Greek adults produced 12 out of 12 pre-verbal clitics. Hellenic Greek children produced very high percentages of proclisis in both versions of the tests. Greek Cypriot adults were different from Hellenic Greeks in significantly producing proclisis (i.e. non-target placement) even when taking the CG-version of the test. Moreover, they performed proclisis (i.e. target placement) when addressing the Hellenic Greek experimenter in the SMG-version of the test to such a great extent that their productions gave rise to a paradox: Their target placement was higher in SMG (98.6%) than in their native CG (76.6%). The binational children behaved like the monolingual Hellenic Greek adults, especially in terms of placement, although the vast majority (25/30) was born in Cyprus and certainly schooled for some time, if not entirely, in Cyprus.

Based on the reports of previous acquisition studies, it was initially expected that target placement is proclisis in SMG and enclisis in CG indicative declaratives. However, adult target placement is not 100% enclitic according to the productions that our task elicited: 76.6% in the CG version of the test. It should be clarified that all adult Greek Cypriots were tested in the house of one of the experimenters and they also had a certain degree of familiarity with the experimenter, so there is nothing that should make us expect the use of SMG. A look at their production in the fillers that involved productions outside the V-CL/CL-V pattern that our task aimed to elicit shows both the use of inflectional morphology and lexical items that are CG-specific. Based on these data, we suggest that there is no evidence that our adult Greek Cypriots code-switched in SMG, hence proclisis cannot be considered the result of code-switching.

The fact that some Greek Cypriot children who performed 100% non-target placement in the CG version commented on their performance or on the experiment's pictures in CG suggests that in bilingual/bidialectal populations especially, children are metalinguistically aware. If Crain and Fodor (1987) are right in suggesting that metalanguage is innate as a medium of representation used to encode observations about language, the link between enhanced metalinguistic abilities and multilingualism established by Bialystock (1991) becomes relevant also for diglossic environments like the one in Cyprus (see Ibrahim et al. 2007 for Arabic). The question raised here is whether the performance of Greek Cypriot children is an instance of code-mixing, as a result of bidialectism, or a(n) (un)conscious demonstration of metalinguistic awareness driven by linguistic anxiety to (show that they are able to) speak 'properly'. Some Greek Cypriot children, even at age 5;10, while trying to align themselves with the 'high' variety, employed non-target placement combined with verbs that do not exist in either variety. By misapplying SMG morphological suffixes to the CG verbal stem, they produced forms that *prima facie*

resemble SMG, but are not SMG. Such forms were totally absent from the production of our true bidialectals, that is, the group of binational children, the overwhelming majority of whom used correct SMG, not due to absence of metalinguistic awareness but due to native(-like) SMG performance that disallowed such errors.

Acknowledgements

At all stages of this experiment, we have received valuable comments by all members of the Cyprus Acquisition Team (CAT), to whom we are grateful. We especially wish to thank Kleanthes Grohmann for co-ordinating, supervising, and (endlessly) supporting our work. All faults remain our own.

References

- Bialystock, E. 1991. Metalinguistic dimensions of bilingual language proficiency. In Bialystock, E. (ed.), *Language Processing in Bilingual Children*. CUP.
- Crain, S., Fodor, J.D. 1987. Simplicity and generality of rules in language acquisition. In MacWhinney, B. (ed.), *Mechanisms of Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum.
- Grohmann, K.K. To appear. Some directions for the systematic investigation of the acquisition of Cypriot Greek. In Rinke E., Kupisch T. (eds.), *The Development of Grammar: Language Acquisition and Diachronic Change — Volume in Honor of Jürgen M. Meisel*. Amsterdam: John Benjamins.
- Grohmann, K.K., Theodorou, E., Pavlou, N., Leivada, E. 2010. Object clitics in typically-developing and language-impaired Cypriot Greek children. Poster presentation at the COST Action A33 final conference *Let the Children Speak*. Wellcome Collection, London, January 22-24.
- Ibrahim, R., Eviatar, Z., Aharon Peretz, J. 2007. Metalinguistic awareness and reading performance. *Journal of Psycholinguist Research* 36(4), 297–317.
- Papapavlou, A., Pavlou P. 1998. A review of the sociolinguistic aspects of the Greek Cypriot dialect. *Journal of Multilingual and Multicultural Development* 19, 212-220.
- Petinou, K., Terzi A. 2002. Clitic misplacement among normally developing children and children with specific language impairment and the status of Infl heads. *Language Acquisition* 10, 1-28.
- Varlokosta, S. et al. [= a large number of participants from COST Action A33: WG1]. To appear. A cross-linguistic study of the acquisition of pronoun and clitic production. *Language Acquisition*.

Contextual information in (non) native speech perception in noise

Bin Li¹, Ratrete Wayland², Caicai Zhang¹

¹Department of Chinese, Translation & Linguistics, City University of Hong Kong, Hong Kong S.A.R.

²Linguistics Program, University of Florida, Gainesville, Florida, U.S.A.

Abstract

Non-native listeners' perception of speech has been shown to be more adversely affected by interference from background noise than native speakers. This study focused on effects of different levels of linguistic information on speech perception in noise. We examined roles of phonetic and contextual cues in sentence perception in babble noise by Chinese learners and native speakers of English. Our sentences varied in degrees of syntactic and semantic acceptability. Results revealed similar patterns of perception in the performance of listener groups. Secondly, listeners benefited more from certain sentence types under different listening conditions.

Key words: speech perception in noise, ESL/EFL learners, sentence perception

Introduction

Everyday speech communication is often characterized by noise, which adversely affects the understanding of speech. Studies showed that non-native listeners are less tolerant with the degradation of speech than native listeners (Roussohatzaki, Florentine 1990). To alleviate the communication burden on the listener, researchers suggested that the talker use a clear speaking style (Picheny et. al. 1985, Hazan, Simpson 1998, Bradlow, Bent 2002, Wayland et. al. 2006). Few previous studies, however, explored roles of higher-level linguistic information in perceiving speech in noise.

The current study

We aimed at examining effects of linguistic information at various levels in perceiving sentences with varying degrees of acceptability under different listening conditions.

Stimuli

74 English sentences with varying syntactic and semantic properties were created. All were evaluated by 6 adult native speakers of English (NE) for their acceptability on a scale of 5. Sentences were then grouped in to 5 types: context dependent (CD) sentences with semantically related and predictable keywords, e.g. The boat sank in the storm; sentences containing context neutral (CN) keywords, e.g. A friend came for lunch; sentences with

syntactically marked (SY) structures, e.g. Up the hill, Mary ran; and semantically marked (SE) sentences with unpredictable collocation, e.g. They are drinking oil; and lastly, syntactically or semantically unacceptable (U) sentences, e.g. They liked in cold water. Production of all sentences was recorded from 2 adult NE speakers, which was digitized and normalized for intensity and then mixed with babble noise at two Signal-to-Noise (SNR) ratios: -4dB and -8dB. Altogether 148 (74x2) test sentences were generated.

Listeners

70 adult Chinese EFL/ESL learners were recruited with normal hearing ability from China's mainland (MN) and Hong Kong (HK) participated. All were full-time university students, and had taken TOEFL or IELTS. 10 adult NE speakers, none of whom participated in sentence preparation, were recruited as our control group. Listeners were aged between 18-35 years.

Procedure

Listeners were instructed to listen to each sentence as many times as they wished before rewriting. They also completed acceptability evaluation on all sentences.

Results and Discussion

Mean percentages correct for keywords rewriting were calculated and compared across listener groups, sentence types and listening conditions.

English proficiency

Self-reported scores of IELTS and TOEFL (TOEFL scores were converted to IELTS bands for comparison) helped group Chinese listeners into 4 groups: Band 5, Band 6, Band 7, Band 8. Mainland listeners had an average score of 7.13 whereas Hong Kong listeners' average score was significantly lower at 6.48 ($p < 0.001$). Listeners' performance was generally positively correlated with their English proficiency (Table 1). A closer look into each SNR revealed that the correlation was stronger under the -4dB SNR condition than under the -8dB one, suggesting a less robust advantage of language proficiency in more adverse listening conditions for non-native listeners.

Listening conditions

All listeners did significantly better under -4 dB SNR condition than under the -8dB one ($p < 0.01$ for all) (Table 1, Fig. 1). NE speakers scored highest in both conditions and MN listeners outperformed HK listeners ($p < .05$ for -8dB SNR and $p < .01$ for all other conditions). As reported earlier, MN listeners were more proficiency in English than HK listeners. This suggested again that the adverse effects of noise on speech perception are more significant

for less proficient learners. When noise reached a level where even NE speakers found perception difficult (41.05% correct rewriting), the difference between learners at various proficiency levels started to diminish.

Sentence types

Our acceptable sentences differed in keywords predictabilities and thus were divided into CD (context dependent), CN (context neutral), SY (marked syntactic structures), and SE (marked semantic information). Both Chinese and NE listeners evaluated all sentences on their acceptability, results of which were highly correlated (Table 2). Rewriting results showed that SE, CD, CN types as a group were significantly better than SY and U for all listeners (all $p < 0.02$) under -4dB SNR condition (Fig. 1). Under -8dB, NE listeners did numerically better with the CD and CN types, but poorer with SE. On the other hand, for Chinese listeners, CN was better than the group of SE, SY and U, which was better than CD (all $p < 0.04$). So when noise level was relatively low, e.g. -4dB SNR, all listeners found SE sentences easier despite its lower degree of acceptability. However, under -8dB SNR condition, such facilitative effect diminished for even NE speakers (though not significantly worse than others, $p > 0.05$). Another interesting finding was the greater deteriorating performance of Chinese listeners with the CD type than SE under -8dB SNR condition. This may be due to their greater dependence on semantic information than on syntactic structures under -4dB SNR condition. Under -8dB SNR condition, however, Chinese listeners may rely more on phonetic cues. The effects of predictable keywords in CD sentences were only consistently beneficial to NE listeners.

Table 1. Mean percent correct (Std. error of mean) of keywords rewriting.

Chinese listeners	SNR -4dB		SNR -8dB	
Band 8	49.95	(1.72)	23.34	(1.50)
Band 7	48.19	(1.82)	24.38	(1.42)
Band 6	45.28	(1.75)	19.58	(1.28)
Band 5	41.77	(1.48)	18.40	(1.12)
NE speakers	75.65	(1.59)	41.05	(1.96)

Table 2. Mean (Std. error of mean) sentence acceptability, 1 meaning totally acceptable.

	Chinese listeners	NE listeners	Correlation
CN	2.054 (.010)	1.173 (.014)	$r = 0.85, p < 0.001$
CD	2.216 (.012)	1.334 (.027)	
SY	2.498 (.011)	2.176 (.023)	
SE	2.614 (.014)	2.835 (.026)	
U	3.093 (.009)	3.880 (.012)	

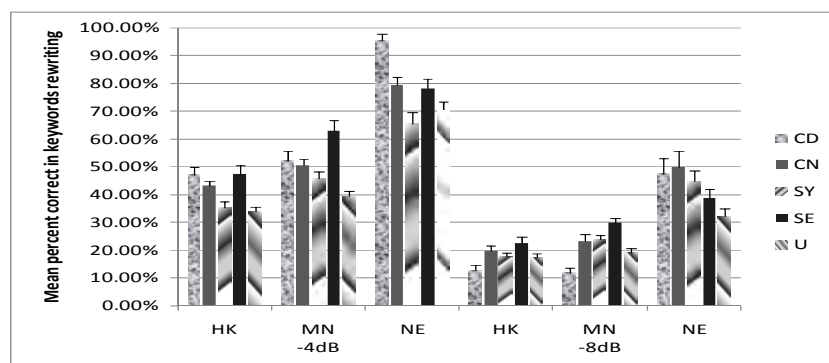


Figure 1. Mean percent correct in keywords rewriting by listener groups and sentence types.

Conclusion

Our study aimed to investigate whether perception of impaired speech by listeners is influenced by their limited usage of the contextual cues or their inability to process speech acoustic signals. We found that listeners' abilities in using semantic or syntactic information in speech perception were commensurate with their proficiency in the target language. Second, semantic information had a stronger facilitating effect on speech perception in noise than syntactic information for both native and non-native listeners. Thirdly, the facilitating effects may be reduced by noise at higher SNRs, whereas acoustic-phonetic information may become more useful in such situations.

Acknowledgements

This study was supported by SRG Grant #7002191 of City University of Hong Kong. We thank Mr. Peng Yukang for his help in data collection.

References

- Bradlow, A.R., Bent, T. 2002. The clear speech effect for non-native listeners. *J. Acoust. Soc. Am.* 112 (1), 272-284.
- Hazan, V., Simpson, A. 1998. The effect of cue enhancement on the intelligibility of nonsense word and sentence materials presented in noise. *Sp. Comm.* 24, 211-226.
- Picheny, M.A., Durlach, N.I., Braida L.D. 1985. Speaking clearly for the hard of hearing I. *J Speech Hear Res.* 28 (1), 96-103.
- Roussohatzaki, M., Florentine, M. 1990. Perception of American-English in noise by Greek listeners, *J. Acoust. Soc. Am.* 87, S72.
- Wayland, R, Li, B., Csonka, L., Harnsberger, J. 2006. Effects of experience on speech perception in noise. *J. Acoust. Soc. Am.* 119, 3422.

Exploring sound symbolism in the investigation of speech expressivity

Sandra Madureira¹, Zuleica Antonia de Camargo^{1,2}

¹Integrated Acoustic Analysis and Cognition Laboratory-LIAAC- Catholic University of Sao Paulo, São Palo, Brazil

² Voice and Speech Laboratory- Cefac Institute, Sao Paulo, São Paulo, Brazil

Abstract

The objective of this paper is to examine specific uses of sound symbolism concerning segmental and prosodic properties. The typology developed by Hilton et al (2004) is taken as reference. The corpus of this work is a poem recorded by a professional actor. A research methodology comprising prosodic perceptual analysis, prosodic acoustic analysis and affective states evaluation tests is proposed. The results indicate the speaking strategies used by the actor have been found to make use of three types of sound symbolism (synesthetic, imitative and metalinguistic). Correlations among acoustic properties, perceived affective states and text meaning production demonstrate productive use of sound symbolism and corroborate the discussion on the direct links between sound and meaning.

Key words: speech expressivity, sound symbolism, acoustic and perceptual cues

Introduction

Our current understanding of the processes involved in speech expressivity is rather limited. One of the reasons for that, being the lack of understanding of how sound symbolism, the direct link between sound and sense (Chuenwattanapranithi, 2008).

The objective of this paper is to examine specific uses of sound symbolism concerning segmental and prosodic properties in the recording of a poem by a professional actor. The actor interprets the poem and it is his production of meaning that is taken into account. His acting task is not restricted to saying a sentence with various kinds of feelings as explored in some expressive speech studies, but it is rather meaning production oriented.

Methods

The corpus of this work is a poem¹, written in the nineteenth century by the Brazilian poet Casimiro de Abreu (1837-1860). The poem was recorded by a professional actor. The recording is available in a commercial CD entitled “Quatro Séculos de Poesia Brasileira” which was released in 2002.

The poetic narrative takes into account the narrator's feelings towards his beloved one and his love rival while watching them dance. The poem has twenty three-syllable verses structured in eleven-line stanzas. One of the

stanzas is repeated five times throughout the poem and although the syntactic and lexical items are the same, the affective states reported in the poem change throughout the text and affects their interpretation.

A research methodology comprising text analysis, prosodic perceptual analysis, prosodic acoustic analysis and affective states evaluation tests is proposed. For the perceptual evaluation of prosodic aspects (voice quality settings and voice dynamics), the VPAS, the scheme adapted by Camargo, Madureira (2008) has been used. For the acoustic analysis, based on PRAAT, manual and automatic measures (SG detector and SG Expressive Evaluator developed by Barbosa to analyse speech expressivity (Barbosa, 2009) have been taken into account. Statistically measures concerning f0 in Hz, VV duration in ms, Long Term Average Spectra, relative intensity and spectral tilt have been considered. For the perceptual evaluation of the expressive uses of prosodic aspects, a group of 30 college students, aged from 20 to 30 years old, have answered a semantic differential scale questionnaire with the following descriptors (activation: calm/activated; valence: pleasant/unpleasant). They have also been asked to identify affective states (joy, sadness, anger and exasperation) and speech acts (advice, admonition, order and plea). Repeated verses throughout the text were the object of the evaluation.

Results

The results indicate the speaking strategies used by the actor have been found to make use of three types of sound symbolism (synesthetic, imitative and metalinguistic).

The poet makes use of sound symbolism to indicate both the dynamics of the dance as well as the dynamics of the observer's conflicting affective states. The use of imitative kind of symbolism has been found in the production of the very first verse of the poem. The rising intonation pattern, the lengthening of the vowel (753 ms) of the first word and the gliding transition to the following vowel segment in the production of the first verse (1411 ms) of the poem *Tu ontem* (You yesterday) are thought to be effective in representing the compass of the dance and the dancer's movements. It sounds like music. It contrasts with the *Tu* (70 ms) in the verse *Tu davas* (795 ms) which occurs in a stanza in which the poet expresses his anger. Figure 1 shows these two productions of *Tu* followed by words which exhibit the same number of syllables and stress pattern.

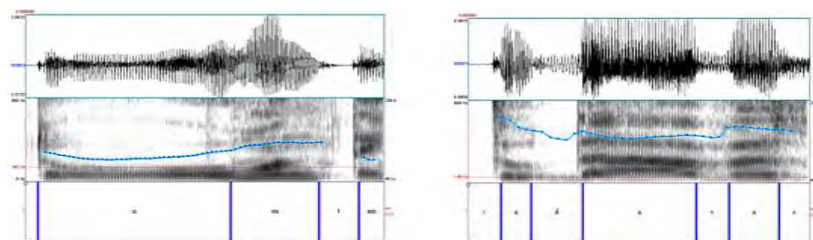


Figure 1. Waveform, wideband spectrogram, f0 contour and orthographic transcription of the segments in the verses Tu ontem (You yesterday) and Tu davas (You give).

The synesthetic kind of symbolism has been used to express the dynamics of the dance which starts fast and gets slower towards the end. Comparison among the stanza repetitions shows that changes in speech rate, rhythm and intonation are introduced by the actor to represent changes related to the dynamics of the dance. Lengthening of fricative sounds and nasalized vowels have also been used in that context. Measurement of the VV units in ms indicated the fifth repetition has longer VV units than the others. Oneway Anova showed that it differs statistically from the first, the second and the third repetitions ($p=0.000$) as well as from the fourth ($p=0.001$). Figure 1 contrasts the duration of VV units of the utterances in the five repetitions of the stanza.

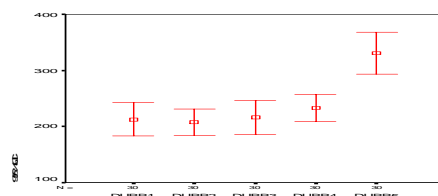


Figure 2. Graphic showing the median values and the standard deviation of the duration in ms of the VV units in the five repetitions of a stanza.

The fifth repetition, which was found to differ from the others in relation to f0 (median, 99,5 quantil, skewness and its first derivate mean, standard-deviation and skewness), occurs after the act of dancing ceases and follows the stanza in which the poet's beloved is described as *batida, caída, sem vida, no chão* (beaten, fallen lifelessly onto the ground). The ceasing of the dance is expressed by changes in speech rate, rhythm and intonation. The actor creates a sound metaphor, speaking at a lower speech rate, producing a lowered larynx voice quality setting and introducing silent pauses.

Instances of metalinguistic symbolism comprised the use of intonation patterns, pitch range and voice quality to signal affective states and

communicative functions. The verse *Não mintas* (Do not lie) was produced with varied intonation patterns, pitch ranges and voice qualities throughout the text. There were also differences in F0 alignment and duration. The judges reported higher degrees of activation for the second, third and fourth repetitions which were produced with hyperfunction and close jaw voice qualities and higher degrees of pleasantness for the fifth repetition which was produced with expanded pharynx. The second, the third and the fourth repetitions were correlated with the expression of admonition, anger and exasperation, the first to request and the fifth to advice and plea (Figure 3).

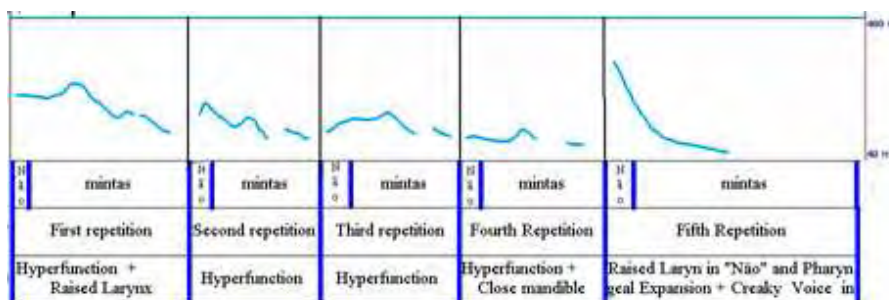


Figure 3. The waveform, the f0 contour, and tiers of annotation for the five repetitions of the utterance *Não mintas* (Do not lie). The first indicates the limits between the two words, the second refers to the number of the repetition and the third identifies the kind of voice quality setting.

Conclusions

Correlations between production characteristics, acoustic properties, perceived affective states and text meaning production demonstrate productive use of sound symbolism and corroborate the discussion on the direct links between sound and meaning.

References

- Barbosa, P.A. 2009. Detecting changes in speech expressiveness in participants of a radio program. In Proc. of Interspeech. v.1, 2155-2158. Brighton, Reino Unido.
- Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., Maneewongvatana, S. 2008. Encoding emotions in speech with the size code: a perceptual investigation. *Phonetica* 65, 210-230.
- Hinton, L. Nichols, J., Ohala, J.J. (eds.). 1994. *Sound Symbolism*. Cambridge: Cambridge University Press, 1994.
- Camargo, Z., Madureira, S. 2008. Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BP-VPAS). In Proc. of the 4th Conf on Speech Prosody v.1, 57–60. Campinas, Brazil.

A dynamic view of the prosodic structure: The example of French

Philippe Martin

EA 3967 CLILLAC-ARP, UFR Linguistique, Université Paris Diderot, France

Abstract

The sentence prosodic structure is traditionally viewed globally in intonation phonology, taking into account all the prosodic events at once from the beginning to the end, without explicitly integrating the fact that these events occur one after the other in a time sequence. Indeed, whereas the speaker can achieve some planning ahead while reading a written text, it is barely the case for the listener, who has to process the sequence of syllabic units perceived one by one along the time scale. It is suggested here that prosodic events are used as acoustic signals triggering partial processing and storage of the already perceived syllables, followed by a concatenation of linguistic units (normally stress groups) already stored. This paper looks in some details into the mechanism of this decoding process, focusing on the role of prosodic events in the specific case of French.

Key words: Prosodic structure, French, Fo contour, cognitive process

Introduction

The formal representation of the prosodic structure in intonation phonology normally extends from the beginning to the end of the sentence. Whether hypothetical-deductive or inductive, prosodic theories using the concept of prosodic structure consider prosodic events (PE) statically as part of the sentence as a whole. For instance, for French, the autosegmental-metrical approach assigns specific tones to intonation phrases normally aligned on syntactic boundaries (Post, 2000; Jun and Fougeron, 2002), whereas other theories (Martin, 1975; Mertens, 1993) assign rising or falling melodic contours to stressed syllables according to some mechanism such as the contrast of melodic slope depending on the declarative or interrogative terminal contour.

Applied to the encoding and decoding processes performed by the speaker and the listener, most of these formal accounts of sentence intonation assume that the entirety of the prosodic structure of each sentence is known in advance. Whereas this might be true to some extent for read speech, where punctuation marks and visual word detection may give some indications on the prosodic structure to be associated with the text, it is less the case for spontaneous speech, except perhaps for short sentences containing only a few number of stress groups.

A dynamic cognitive view

From the point of view of the listener, prosodic events are perceived one after the other in function of time. Therefore their categorization (i.e. the identification of the class they belong, assuming of course that all PE are not identical) depends only on past and not on future events.

The prosodic structure reconsidered dynamically results from a process by which strings of syllables are hierarchically assembled thanks to the identification of each PE as belonging to classes known by the listener. This process involves 2 mechanisms: a) processing of the string of syllables into a linguistic unit such as a stress group followed by the storage of their processed unit and 2) concatenation of this unit with all units belonging to the same level (i.e. whose storage was triggered by same class of PEs).

Among all PEs occurring in French, only those located on the last syllable of stress groups take part in this mechanism, at the exclusion of events located on the first syllable of lexical words, which are treated as secondary stress (a stress group contains either a Noun, a Verb, an Adjective or an Adverb). This means that the identification of PEs implied in the dynamic prosodic structure elaborated by the listener involves the identification of a syllable as stressed in order to qualify the associated PE as a triggering signal to process to storage-concatenation mechanism.

The Storage-Concatenation process

The dynamic view leads to consider that each PE, instantiated by various contrasts of variation and melody height, duration, of intensity, or vocalic quality, appears as a signal triggering the processing and storage of syllables perceived since the appearance of the last prosodic signal belonging to the same class. This syllabic storage is accompanied by concatenation of the elements already present at the same prosodic “address”.

We can look into this mechanism in some details with the simple example schematized on Fig. 1 which gives a diagram of the storage-concatenation mechanism relative to a syllabic sequence $\sigma_1 \sigma_2 \dots \sigma_n$. In this sequence, some syllables are stressed and are therefore associated with a PE in the prosodic structure. When one of these stressed syllables is perceived, the PE attached to that syllable is identified by a grid of perception acquired by the listener.

In our examples, this grid has 4 classes and pertains to the knowledge of the linguistic system of French, but can easily be adapted to other systems as well.

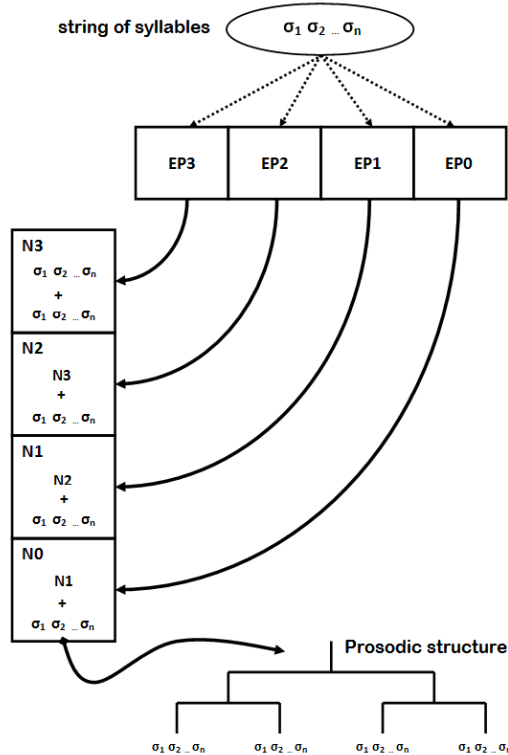


Figure 1. Schema of the storage-concatenation process

Each one of these identifications by the grid starts the following operation:

- 1) Storage of the syllabic sequence appeared since the last PE;
- 2) Conversion of the sequence into a linguistic unit (typically corresponding to a stress group, containing a verb, noun, adverb or adjective and some grammatical words depending on them;
- 3) Concatenation with the converted linguistic units stored at the lower levels, if they exist, and deletion of the existing strings of syllables already stored. This procedure implies a hierarchy between the PEs: $EP0 > EP1 > EP2 > \dots > EPn$.

In the example, EP1 assembles all the sequences stored in the lower memories N2, N3 taken in the order with the sequence of syllables appeared since the last PE. This operation can be generalized on all levels. We will call this general procedure Storage-Concatenation Process (SCP).

It also results from this dynamic cognitive view that prosodic marks enter a network of relative contrasts, and do not need invariant acoustic characteristics, as long as they are identified by the listener as belonging to the same class. Another argument going in this direction comes from the large variety of the styles of production of word. An emotionally depressed speaker will realize little or no melody variations. In whispered voice, contrasts between prosodic marks will have to be realized by other means than laryngeal frequency variations, with segment durations, about 50 to 70% higher than the equivalent achievements of duration in not whispered voice. There cannot thus be invariant characteristics for the prosodic marks, as opposed to what many of researchers in this field seem to admit.

Some cognitive consequences

The storage-concatenation dynamic view leads to a better understanding of some constraints governing the prosodic structure (Martin, 2009). First, the production of a string of syllables is not only constrained by the respiratory capacity of the speaker (Gilbert and al., 2007), but as well by the human immediate memory possibilities, which prevent stress groups to exceed some 7 or 8 syllables (Gilbert and al., 2010). Furthermore, the syntactic collision rule, which does not allow the stress groups whose syntactic units are dominated by distinct nodes in the syntactic structure to be concatenated by the prosodic structure (Martin, 2009) prevents the listener to build a misleading sentence structure at the first step of the concatenation process.

References

- Gilbert, A.C. Boucher, V.J. 2007. What do listeners attend to in hearing prosodic structures? investigating the human speech-parser using short-term recall", In INTERSPEECH-2007, 430-433.
- Gilbert, A.C., Boucher, V.J., Jemel, B., Lalonde, B. 2010. Segmentation de la parole : des groupes rythmiques et des énoncés, pas des mots et des phrases. Présenté au 78e Congrès de l'ACFAS, Montréal, Québec.
- Jun, S.-A., Fougeron, C. 2002. The Realizations of the Accentual Phrase in French Intonation, *Probus* 14, 147-172.
- Martin, Ph. 1975. Analyse phonologique de la phrase française, *Linguistics*, (146) Fév. 1975, 35-68.
- Martin, Ph. 2009. *Intonation du français*, Paris: Armand Colin, 256 p.
- Mertens, P. 1993. Intonational grouping, boundaries, and syntactic structure in French, in House, D., Touati, P. (eds.), *Proc. ESCA Workshop on Prosody*, September 27-29, 1993, Lund (S) Working Papers 41 (Lund Univ., Dept. of Linguistics), 156-159.
- Post, B. 2000. Tonal and phrasal structures in French intonation. The Hague: Holland Academic Graphics, 289 p.

Acquisition of lightverbhood: Evidence from Cypriot Greek

Paraskevi Mavroudi, Evelina Leivada, Anna Epistithiou
Department of English Studies, University of Cyprus, Cyprus

Abstract

This experiment investigates the acquisition of (true) light verb constructions (in the sense of Kearns 1988/2002) as well as other constructions that involve a light verb in Cypriot Greek. Focusing on the acquisition of the light verb *kamno* ‘do/make’, this study aims to identify (i) whether light verb constructions are more frequent in Cypriot Greek than in Standard Modern Greek, (ii) what the age of acquisition is for true light verb constructions, and (iii) how sensitive to (light verb) priming children of different ages are in a picture-based elicitation task; in addition, this research tests whether adults can be sensitive enough to priming so as to reproduce light verb constructions that are inexistent in Cypriot Greek.

Key words: acquisition, Cypriot Greek, lightverbhood, priming.

Lightverbhood: Theory and acquisition

Kamno ‘do/make’ is a contentful verb (CV) that functions as ‘light’ in certain environments. In such light verb construction (LVC) environments, light verbs (LVs) combine with their complement (such as VP, NP, or AP) to form primary or complex predication. Primary predication arises from a phrase structure such as [_{VP} DP [_{v'} v XP]], with DP the external argument (Hornstein et al. 2005: 108). The meaning of the LVC is the same as the equivalent CV, and the main semantic content of the predication is provided not by the LV but by its complement (Kearns 1988/2002). With respect to its semantic status, we argue that *kamno* is a lexical verb that functions as light in such constructions by making no semantic contribution to the LVC, yet making an aspectual contribution, similar to what van Hout (1998) reports for ‘aspectual’ LVs in Dutch and Alexaki (2003) for *kano*, the Standard Modern Greek (SMG) variant of the Cypriot Greek (CG) *kamno*. Since we are unable to detect any semantic contribution of *kamno* to the overall meaning of the LVC, we suggest that it is semantically empty in the attested LVCs. The best way to account for the inability to detect its individual meaning in the meaning of the LVC is to suggest idiomaticity inside LVCs (Marantz 1998).

In CG, the only reference to lightverbhood in acquisition studies is an observation made by Grohmann et al. (2010): Some of their test structures in the clitics-in-islands test they carried out were completed with LVCs, despite the fact that the experimenter primed children with a CV. In SMG, LVs were

discussed in Alexaki (2003): In her longitudinal study, Christos until 2;0 used LVs to a limited extent (15%–28% of his total verb production). At the age of 2;4 this became 60% and after 2;5 LV-production decreased to 20%.

The present experiment

We employed a picture-based elicitation task with 17 test sentences and 7 fillers. Test sentences were completed inside a *because*-declarative island, by producing an LVC and its internal argument, if one was available. Fillers required a simple picture description. All test sentences appear *prima facie* alike with respect to the LV and to the non-verbal part of the construction, which is always an NP; however, they can be divided into five types: (i) LVCs pass the Greek adaptation of Kearns' (1988/2002) tests for determining a true LVC (Type 1), (ii) LVCs that do not pass more than one of Kearns' tests (pronominalization and *wh*-movement, since passivization was not considered a reliable test in Cypriot Greek) (Type 2), (iii) the (true) LVC can be used in contexts in which the CV is not felicitous, whereas in Types 1 and 2 the two could be used interchangeably (Type 3), (iv) the CV can be used in contexts where the (true) LVC is not felicitous (Type 4), and (v) constructed LVCs that do not exist in adult speech, where the CV is stem-identical with the non-verbal part in 2 of the 3 structures (Type 5).

The subjects are divided into six age-groups ranging in age from 4;0 to 9;11, 20 in each age group: 10 receive as input the CV and 10 the corresponding LVC, to evaluate the significance of priming in such elicitation tasks. All children were monolingual in CG and were recruited randomly from 12 primary and pre-primary schools in the Nicosia district. The control group consisted of 12 adults (aged 20–54): 6 adults (mean 39;8) received LVC input, 6 adults (mean 31;5) received CV input. A typical example is the following (Type 1 item):

Experimenter: The girl is happy because the boy hugs her (= does/makes her hug). Why is the girl happy? The girl is happy because the boy...

(To koritsi en xarumeno jati to ayori kamni tu agaja. jati to koritsi en xarumeno? To koritsi en xarumeno jati to ayori...)

Child: ... *does her hug*. (... kamni tu agaja.)

Findings

First findings show that children extensively produce true LVCs with *kamno* already by the age of 4; even when primed with CVs, they employ LVCs in particular test sentences. All children are affected by LVC-priming to the extent that they reproduce Type 5-LVCs or construct their own inexistent ones or even produce, to a limited extent though, serial verb constructions of the form [LV–CV] which are inexistent in adult speech.

In the LVC-version, across age groups, Type 1-items (true LVCs) are usually the ones less reproduced, whereas there is no observed pattern across

age groups for a type that is more reproduced. Two children (aged 5;2 and 8;0) produced four serial verb constructions (inexistent in adult speech) of the form [LV–CV], with *kamno* the LV and the corresponding contentful verb the CV. The production of our inexistent LVCs (Type 5 items) was surprisingly high across age groups:

Age group 4;0–4;11:	50.0% production
Age group 5;0–5;11:	73.3% production
Age group 6;0–6;11:	56.6% production
Age group 7;0–7;11:	96.6% production
Age group 8;0–8;11:	76.6% production
Adult control group:	27.7% production

These percentages show the effects of priming in this kind of elicitation task, since all participants were primed enough so as to produce structures that do not exist in their production otherwise. One of the primary aims of this study was to see whether LVCs are more frequently used in CG than in SMG. Calculating the LV use in findings from another elicitation task, the clitics-in-islands test (Leivada *et al.* 2010), this hypothesis seems to be correct. Different populations showed different LV use without LV-priming in this case: Both Hellenic Greek children and Greek Cypriot children received as input CVs. However, the latter group produced more LVs: 3.54% for Hellenic Greek and 6.45% for Greek Cypriot children, each percentage calculated for productions in a given participant's native variety.

Finally, the term employed so far to refer to verbs like *kamno* is 'light verb'. However the term 'General All-Purpose' (GAP) verbs has also been suggested in literature to refer to that type of verbs: "GAP verbs are defined as verbs that do not have a semantically specific meaning" (Goldberg *et al.* 2008: 43), referred to also as 'semantically flexible'. Stavrakaki (1998), in her discussion of the SMG *kano*, draws a distinction between the light *kano* and the idiomatic one, which she argues to be stored in the lexicon in the form of the idioms in which it appears. However, if we do not want to overload the conceptualization of our lexicon, there is no reason to assume two or three different *ka(m)no* stored there — one contentful, one idiomatic, one semantically flexible with light or general purpose, etc. Since none of the productions that our task elicited provides evidence for semantic flexibility in the use of *kamno*, we consider our test structures as involving aspectual light verbs: These are contentful verbs functioning as light in these environments by not making any semantic contribution to the meaning of the LVC.

Acknowledgements

This experiment was an idea that emerged from meetings of the Cyprus Acquisition Team (CAT), therefore we are grateful to all CAT members that

participated for their insightful comments. Within the team, many thanks go to Maria Kambanaros who was the person both to answer all our queries about the connections of light verbs with Specific Language Impairment, but also to provide useful references and comments. Finally, this experiment would not have been carried out, had it not been for Kleanthes Grohmann and his constant help and support at every stage of this study.

References

- Alexaki, C. 2003. Γενικές παρατηρήσεις πάνω στην κατάκτηση του ρήματος στην Νέα Ελληνική ως μητρική γλώσσα. [General observations on the acquisition of verbs in L1 Modern Greek]. In Catsimali, G., Kalokairinos, A., Anagnostopoulou, E. and Kappa, I. (eds.), *Proceedings of the 6th International Conference on Greek Linguistics*, Crete, Greece.
- Grohmann, K.K., Theodorou, E., Pavlou, N., Leivada, E. 2010. Object clitics in typically-developing and language-impaired Cypriot Greek children. Poster presentation at the COST Action A33 final conference *Let the Children Speak: Learning of Critical Language Skills across 25 Languages*. Wellcome Collection, London, January 22-24.
- Goldberg, H., Paradis, J., Crago, M. 2008. Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics* 29, 41-65.
- Hornstein, N., Nunes J., Grohmann, K.K. 2005. *Understanding Minimalism*. Cambridge: Cambridge University Press.
- van Hout, A. 1998. *Event Semantics of Verb Frame Alternations: A Case Study of Dutch and its Acquisition*. New York/London: Garland.
- Kearns, K. 1988/2002. *Light verbs in English*. Ms., Massachusetts Institute of Technology, Cambridge/University of Canterbury, Christchurch.
- Leivada, E., Mavroudi, P., Epistithiou, A. 2010. (Meta)language in diglossic environments: Clitic placement by Hellenic Greek, Greek Cypriot and binational children. Poster presentation in *Language Disorders in Greek 3*, European University Cyprus, Nicosia (June 12-13).
- Marantz, A. 1998. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In Dimitriadis, A. (ed.), *Proceedings of the 27th Annual Penn Linguistics Colloquium*. Philadelphia, PA: Working Papers in Linguistics 4.
- Stavrakaki, S. 1998. *Kano*: The case of a light verb in MG. In Lambropoulou, S. (ed.), *Proceedings of the 12th International Symposium of Theoretical and Applied Linguistics*, vol. 1, 171-185. Thessaloniki, Greece.

Phonetic factors influencing /l/-rhoticisation in Greek

Daniela Müller

CLLE-ERSS, Universitat de Tolosa 2 – Lo Miralh, France & Romanisches Seminar,
Ruprecht-Karls-Universität Heidelberg, Germany

Abstract

This study investigates phonetic factors – vocalic/consonantal context, duration, stress - favouring rhotic perception of the lateral approximant in Greek. Nonsense words containing the lateral in different lengths in both intervocalic and internal coda position, alongside similar words containing a rhotic, were presented to Greek listeners for categorisation as /l/ or /r/. Results showed that laterals were perceived as rhotics if they matched the rhotic in duration. Rhotics, on the other hand, were only perceived as laterals if they were articulated as an approximant rather than as a tap.

Key words: laterals, rhotics, perception, sound change

Rhoticisation of laterals in Greek

The alternation of rhotics and laterals is a widespread cross-linguistic phenomenon (Maddieson 1980, Proctor 2009), and the confusion of the two sounds may be enhanced in environments where reduction is expected, such as in unstressed syllables. In all Greek dialects, a lateral in internal coda position rhoticised into an alveolar tap, regardless of place of articulation of the following obstruent, even though the standard language has since reintroduced forms with /l/ (Newton 1972). Dialectological evidence from other languages, e.g., Italian dialects, shows that following velar or labial sounds may favour /l/-rhoticisation in coda position (Rohlf's 1966).

Lambdacisation, i.e., /r/ > /l/, such as occurs e.g. in Andalusian Spanish (Quilis-Sanz 1998), on the other hand, is a much rarer phenomenon than rhoticisation. It has been reported for Greek only in cases of liquid dissimilation.

In order to assess which phonetic factors (duration, stress, vocalic and consonantal context) favour /l/-rhoticisation in Greek, a perception experiment has been conducted.

Experiment

Stimuli

The stimuli consisted of disyllabic nonsense words containing either /l/ or /r/ in intervocalic position or in word-internal coda, as well as either one of the vowels /a, i, u/, either one of the consonants /p, t, k/ (where applicable), and

stress either on the first or the last syllable. A male speaker of Standard Thessaloniki Greek read the stimuli three times from a list at a self-chosen speaking rate. All tokens obtained were used for the acoustic analysis. For the perception test, the best one among the three repetitions was selected (or the two best ones in the case of rhotics, provided that both taps and approximants were available (see below)). In the stimuli containing /l/, the duration of the lateral was manipulated to obtain two further stimuli of the same word: one with a 75-ms-long lateral (matching the average length of the approximant in the speaker's speech) and another in which the lateral had a duration of 30 ms (corresponding to the length of the tap closure).

Laterals and rhotics in the Greek speaker

The laterals of the Thessaloniki Greek speaker were greatly influenced by the (symmetrical) vocalic context. They were somewhat dark in the environment of /a, u/ (Δ F2-F1 1041 Hz and 1146 Hz, respectively), and extremely clear in the /i/-context (Δ F2-F1 2147 Hz).

The rhotics of the speaker could take either one of two realisations: a tap (with possible undershoot) accompanied by a short svarabhakti vowel when in coda position, or an approximant. Taps tended to occur mainly with back and low vowels, while approximant articulation was more readily found in /i/-contexts. Intervocalic rhotics were always taps.¹

Subjects

All the stimuli (three repetitions of each) were presented in random order over headphones to 21 native listeners of Greek (age range 18 – 37) from diverse regions of Greece and Cyprus. The listeners were asked to identify the nonsense words as containing either a rhotic or a lateral by clicking on the orthographic transcription of the word on a computer screen.

Results

Rhotic perception of laterals

The rhoticisation rate of the 30 ms-long laterals in intervocalic position was 20% in the /a/-context, 28% with /i/, and 60% when the context vowel was /u/. Laterals of greater durations were not perceived as rhotics. The significant effect of duration ($F=126,35$, $p<0,0001$) was confirmed by statistical analysis (ANOVA). A post-hoc Tukey test of multiple comparisons of means revealed that any significant effect other than that of duration (30 ms vs. 75 ms and original duration) hinged on the fact that the token /úlu/ yielded an extremely high number of rhotic responses (83%).

In internal coda position, laterals, again those with 30 ms-duration, were perceived as rhotics 15% of the time in /a/-context, 33% with /i/, and 53% in /u/-context. Statistical analysis (ANOVA) yielded a highly significant effect

for vowel quality ($F=56,372$, $p<0,0001$) (/u/ leading to rhoticisation more than half of the time with 30 ms-tokens) and lateral duration ($F=239,6598$, $p<0,0001$). Lateral duration thus has about the same effect as with the intervocalic tokens. Contrary to expectations and predictions from the dialectological literature, place of articulation of the following stop consonant didn't yield any significant effect ($F=0,7354$, $p=0,4796$). Stress was also not significant in either intervocalic or internal coda tokens.

Lateral perception of rhotics

Due to the presence of two types of rhotic in the speech of the Thessaloniki Greek speaker, both were presented to listeners for categorisation wherever possible. Both rhotic types were not, however, available in each case, which leads to the results being somewhat unbalanced. Notwithstanding these drawbacks, it is useful to separate the two rhotic types, for they proved to be prone to lambdacisation to different degrees. On average, only 2% of the taps were perceived as /l/, whereas 28% of the approximant rhotics led to misidentification as a lateral. This difference was highly significant (Student's t-test: $p=0,003$). Within the approximant rhotics, lateral perception occurred 35% of the time in /a/-context, 10% with /i/, and 54% with /u/. Due to the unbalanced occurrences of approximant rhotics across vowel contexts, these results were not further analysed statistically, but the high rate of confusion in the context of the back rounded vowel /u/ is striking. Among the taps, the tokens /artá/, /urtú/, and /urkú/ also yielded high rates of lambdacisation (16%, 11%, and 6%, respectively).

Discussion

It has been shown that rhotic perception of laterals is due primarily to variability in duration. The original length of the laterals used in this study, around 120 ms, has been obtained under laboratory conditions; in spontaneous speech, however, much shorter durations are likely to occur for lateral sounds, in which case rhoticisation would be an expected sound change. This is in line with the fact that laterals rhoticise more readily in unstressed environments where segmental shortening can usually be encountered.

The reverse case of lambdacisation, on the other hand, seems to hinge on more than simple duration issues. A tap, even a reduced one, being a fast ballistic movement of the tongue tip (Barry 1997) designed to create a short interruption in the acoustic carrier signal, cannot readily be lengthened. If an approximant rhotic is considered as the reinterpretation of undershot taps and thus becomes a target articulation for the rhotic in free allophonic variation with the tap in contemporary Thessaloniki Greek, it may be hypothesised that a gestural reorganisation is a necessary first step in the

perception of the rhotic as a lateral. Given these more complex prerequisites, lambdacisation is expected to be a far less common phenomenon than rhoticisation, which cross-linguistic dialectological facts seem to confirm.

Notes

1. Baltazani 2005 finds a somewhat different distribution of taps and approximants in her study of Greek rhotics. Specifically, her speakers present more approximants intervocally than in internal coda position.

Acknowledgements

I'd like to thank Andrew Isaak for technical assistance, as well as Vangelis Karatsiolis and Nontas Apostolopoulos for help with finding the subjects of this study. Special thanks to Giorgos Frangos without whom the study would not have been possible.

References

- Baltazani, M. 2005. Phonetic variability of the Greek rhotic sound. Poster presented at Phonetics and Phonology in Iberia, Barcelona, 20-21.06.2005.
- Barry, W. 1997. Another R-tickle. In *Journal of the International Phonetic Association* 27, 35-45.
- Maddieson, I. 1980. A survey of liquids. *UCLA Working Papers in Phonetics* 50, 93-112.
- Newton, B. 1972. *The Generative Interpretation of Dialect. A Study of Modern Greek Phonology*. Cambridge, CUP.
- Proctor, M.I. 2009. *Gestural Characterization of a Phonological Class: the Liquids*. Doctoral Dissertation, Yale University.
- Quilis-Sanz, M.J. 1998. Las consonantes [-r] y [-l] implosivas en Andalucía. *Revista de Filología Española* 78, 125-156.
- Rohlf, G. 1966. *Grammatica storica della lingua italiana e dei suoi dialetti*. Fonetica. Torino, Einaudi.

Perception of “tonal focus” in Greek

Olga Nikolaenkova

Phonetics Laboratory, University of Athens, Greece

Abstract

The present paper reports on the way tonal correlates of focus impact its identification in Greek simple declaratives. The material was based on 4 utterances with different focus placement. Manipulation of the F0 contour and duration of the original utterances resulted in a list of 18 utterances, repeated 10 times each, and presented randomly to each of 10 informants. The informants were asked to indicate what part of each utterance bore focus. The results showed that: (1) F0 is the most reliable parameter for focus perception, especially the rate of tonal rise/fall, (2) duration expansion is not sufficient for focus perception, and (3) combining F0 and duration affects successful focus identification.

Key words: Greek, focus, tonal structure, tonal movement, perception

Introduction

Traditionally, tonal rise, longer duration and increase of intensity have been considered as correlates of focus. The so called focal accent, meaning tonal rise, was suggested as the main correlate of focus by Bruce (Bruce 1977) based on the assumption that the element in focus is given prosodic prominence by adding a focal accent rise after the word accent fall, performing the local tonal range expansion in order to be more contrastive as global tonal structure compression is being performed (Botinis et al. 2000, Baumann et al. 2006).

According to the results of our previous studies of focus production, local tonal expansion and global tonal compression are not observed in most cases so clearly (Nikolaenkova 2009). F0 rise does not seem to be an obligatory cue for focus production, which agrees with the observations made by Heldner for the case of focus perception in Swedish (Heldner, 1998). A parameter that was suggested to be more reliable for focus identification is the slope, the rate at which the tone falls.

The present research was aimed at testing F0 as well as duration expansion, which are widely accepted as tonal correlates of focus perception.

Methodology

The speech material for the present study consists of a Greek declarative sentence “i me-LI-na MA-lo-ne ti-MA-na-mu” (Melina was arguing with my mother). For the scope of the experiment, different focus placement was elicited by instructing the speaker to answer questions designed to elicit

focus placement on the desired sentence element. Four focus conditions were elicited: neutral, S-focus, VP-focus and O-focus. A female native speaker of Greek, in her late twenties, was recorded producing the utterances seated in a sound treated room at the University of Athens Phonetics Laboratory. The speech material was recorded to computer disk using the PRAAT software package.

For the needs of the present experiment three series of manipulations on the originally recorded material have been made: pitch contour manipulation, duration manipulation and combination of both.

The first set of pitch manipulated stimuli was based on flattened natural pitch contour manipulation, either neutral or with naturally produced focus (see Figures 1, 2).

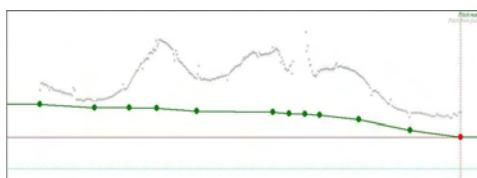


Figure 1. Pitch flattening trace with neutral focus placement.

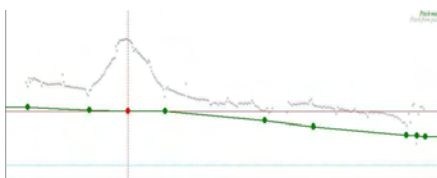


Figure 2. Pitch flattening trace with S focus placement.

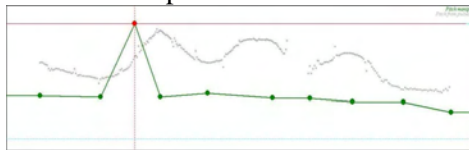


Figure 3. Pitch manipulation trace with S focus placement on basis of flattened O focus contour.

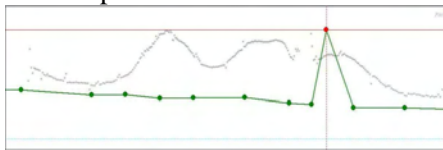


Figure 4. Pitch manipulation trace with O focus placement on basis of flattened V focus contour.

On the basis of the flattened pitch contour, a pitch rise-fall on stressed syllable nucleus was simulated for Subject, Verb and Object focus elicitation (see Figures 3, 4).

In each manipulation pitch range has been expanded in accordance with the pitch peak range of naturally produced focus. As a result of manipulating pitch, 6 utterances (3 based on flattened neutral contour and 3 based on flattened focused production) have been used.

The set of manipulations with duration included stimuli where the element in focus got stressed syllable duration expanded by 2 periods equally for consonant (liquid or nasal) and vowel (approx. 30 ms in total). The

expansion was made by duplicating two naturally produced periods from the middle of the respective sound.

The third set of stimuli consisted of utterances involving both pitch range expansion and stressed syllable duration expansion.

The experiment was based on an exclusively designed computer application. It was calling the 18 WAV files organizing them in 10 sets with different order of stimuli every time. Each listener was instructed to mark the most prominent element if any, while there was also an option “none”. All the options were visualized in four interactive buttons. The time interval between the stimuli was 1,5 sec giving no return or break option – missed stimuli were ignored. The listeners could only take breaks between the sets.

The perception test yielded 1800 responses (18 utterances x 10 repetitions x 10 listeners).

Results

The results showed that F0 expansion had the best total identification rate amounting to 92,67% (see Table 1) in case of manipulating the flattened neutral pitch contour. This rate was higher than the one obtained with similar manipulation on naturally focused contour (79%). Such difference in perception rates may be indicating other tonal parameters present in natural stimuli.

Table 1. Successful focus identification in relation to the parameters used for tonal manipulation.

Total Results – Parameters						
Parameter	F0	Duration	F0 on Focus	Duration on focus	F0 and Duration	Natural
Matching	278	36	237	15	273	139
% Success	92,67%	12,00%	79,00%	5,00%	91,00%	46,33%

Table 2. Successful focus identification in relation to focus placement.

Total results - Focus placement			
Focus type	Subject	VP	Object
Matching focus type	369	326	283
% Success	61,50%	54,33%	47,17%

It is worth mentioning that duration expansion did not contribute to successful focus identification having low identification rates (see Table 1). However in combination with F0 local expansion it had impressively high perception rate being 100% identified by most informants. Some of our informants even showed preference to this combination having even lower rates of 80% in cases of F0 expansion alone.

Furthermore the perception experiment confirmed that focus placement on Subject is much better perceived than placement on Object and Verbal Phrase (see Table 2).

Conclusions

According to the results of the present study, local tonal expansion combined with global compression especially in the post-focus area being widely acknowledged as focus-indicator by several investigators (e.g. Botinis et al. 2000, Botinis 2003, Xu 2005) has the highest impact on successful focus identification for Greek.

Close ties between F0 change and duration revealed also by our previous production studies (Nikolaenkova 2009) have been reconfirmed also by the perception data which showed not only high identification rating but also preference by some speakers.

Acknowledgements

I would like to thank Pandelis Karamolegos for implementing the conceptual design of the experiment having undertaken programming the relative application. I am very grateful to Antonis Botinis for his kind guidance and advisory, to Marios Fourakis for his constructive comments as well as to all the volunteers for the present experiment.

References

- Baumann, S., Grice, M., Steindamm, S. 2006. Prosodic marking of focus domains – categorical or gradient? *Proc. Speech Prosody 2006, Dresden*. 301 – 304.
- Botinis, A. (Ed.) 2000. *Intonation: Analysis, modelling and technology*. Dordrecht: Kluwer Academic Publishers.
- Botinis A. 2003. Focus and gender tonal interactions. *Proc. XVth International Congress of Phonetic Sciences, 1826-1829. Barcelona, Spain*.
- Bruce G. 1977. *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.
- Heldner, M. 1998. Is an F0-rise a necessary or a sufficient cue to perceived focus in Swedish? In Werner, S. (Ed.), *Nordic Prosody: Proceedings of the VIIth Conference*, 109-125, Joensuu 1996. Frankfurt am Main: Peter Lang.
- Nikolaenkova O. 2009. *Rhythmic and tonal correlates of focus in Greek*. Master Dissertation. University of Skovde, Sweden.
- Xu Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46, 220-251.

Acoustic structure of fricative consonants in Greek

Elina Nirgianaki¹, Anthi Chaida¹, Marios Fourakis²

¹Phonetics Laboratory, University of Athens, Greece

²Department of Communicative Disorders, University of Wisconsin-Madison, USA

Abstract

The present study examines the temporal and spectral characteristics of Greek fricatives (duration and spectral moments, i.e. mean, variance, skewness, kurtosis) as distinctive cues for their place of articulation. The effects of voicing, speaker's gender and post-fricative vowel on both duration and spectral moments are also investigated. The results indicate that noise duration does not distinguish fricatives in terms of place of articulation. However, voiceless fricatives have longer durations than voiced ones. Spectral moments distinguish fricatives in terms of place of articulation, except for the labiodental from dental place.

Key words: Greek, fricatives, duration, spectral moments

Introduction

Very little is known about the Greek fricative consonants. Most studies examine [s] (e.g. Nicolaidis, 2002); while there is almost no information about the acoustics of the other Greek fricatives except for the duration data of Fourakis (1986). The present study examines duration and spectral moments of Greek fricatives in word initial position as a function of voicing, speaker's gender, post-fricative vowel and place of articulation.

Noise duration has been shown to distinguish English voiced from voiceless fricatives, with voiceless being longer than voiced (Baum & Blumstein, 1987; Behrens & Blumstein, 1988). Spectral moments' analysis is a statistical procedure that differentiates English fricatives' place of articulation (Jongman et al., 2000). According to it, an FFT derived spectrum is treated as a random probability distribution from which the first four moments (center of gravity, standard deviation, skewness and kurtosis) are computed. Center of gravity corresponds to the mean of the distribution. Standard deviation captures the amount of spectral energy's dispersion around the mean. Skewness indicates whether the distribution is tilted to the left or the right. Kurtosis captures whether the shape is more peaked or flat.

Methodology

Four speakers, two females and two males, recorded the experimental material. All are native speakers of Greek and none of them has any history of speech or hearing disorders. The ten Greek fricative consonants [f], [v], [θ], [ð], [s], [z], [ç], [j], [x] and [ɣ] were recorded in real, two-syllable words

(CVCV) stressed on the first syllable. Each fricative was in initial position and the following vowel varied over the five Greek vowels [a], [e], [o], [i], [u]. Words beginning with [x] and [ɣ] were followed by [a], [o], [u], since /x/'s allophone, [ç], appears before [e] and [i]. The carrier phrase was "I said ...again" ('ipa...ksa'na). Each token was repeated five times, yielding a total of 230 tokens per speaker. Fricative segmentation involved the simultaneous consultation of waveform and wideband spectrogram. Spectral moments were computed from FFTs calculated using a 40-ms full Hamming window at three different locations in the fricative: onset, middle and end.

Results

Frication duration

Table 1 shows the duration of each fricative. A four-way ANOVA (placeXvoicingXvowelXgender) revealed a main effect of voicing [$F(1,720)=780.588$, $p<0.0001$], i.e. voiceless fricatives were significantly longer (122.50) than voiced ones (93.38). A main effect of gender [$F(1,720)=1483.418$, $p<0.0001$] showed that fricatives produced by males (89.20) were significantly shorter than those produced by females (126.68). Place of articulation was also significant [$F(1,720)=9.929$, $p<0.0001$], although not all places, in accordance with post hoc tests. Differences between almost all vowels of different height were significant.

Table 1. Mean duration (ms) for each fricative and each place of articulation.

Fricative		Voiced	Voiceless	Place of articulation
[v]	[f]	99.05	120.84	109.95
[ð]	[θ]	95.61	119.07	107.34
[z]	[s]	95.68	127.43	111.55
[j]	[ç]	87.92	121.02	104.47
[ɣ]	[x]	85.48	125.21	105.34

Spectral moments

One-way ANOVAs for all factors across window locations, with four moments as dependent variables, revealed a main effect of spectral mean on place of articulation [$F(4,915)=85.623$, $p<0.0001$]. Labiodentals, dentals and alveolars were not differentiated by first moment, although differences were significant for all other paired comparisons (Table 2). Variance had a main effect on place of articulation [$F(4,915)=492.092$, $p<0.0001$], although it was not significant neither for labiodentals-dentals, nor for alveolars-palatals. Skewness was also significant [$F(4,915)=114.051$, $p<0.0001$], although post hoc tests showed that it did not differentiate alveolars and dentals from labiodentals; palatals from velars. Kurtosis [$F(4,915)=52.246$, $p<0.0001$] failed to differentiate labiodentals, dentals and alveolars.

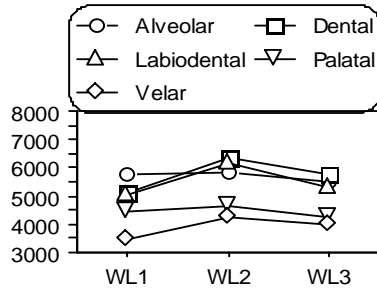


Figure 1. Spectral mean in Hz (averaged across vowels, voiced-voiceless tokens, gender) for each window location, as a function of place of articulation.

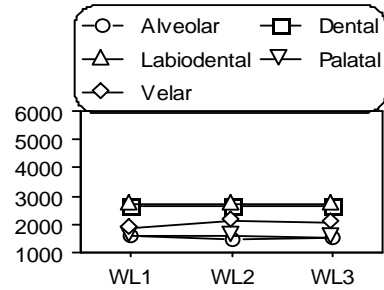


Figure 2. Spectral variance in Hz (averaged across vowels, voiced-voiceless tokens, gender) for each window location, as a function of place of articulation.

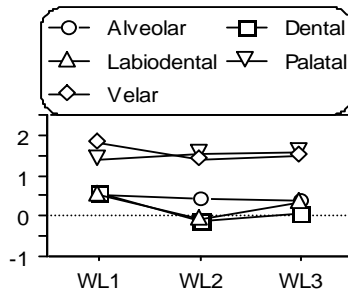


Figure 3. Spectral skewness (averaged across vowels, voiced-voiceless tokens, gender) for each window location, as a function of place of articulation.

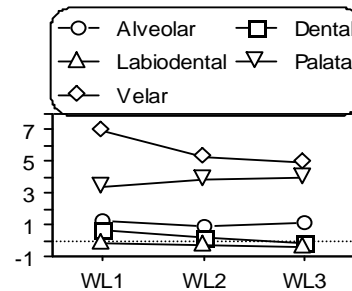


Figure 4. Spectral kurtosis (averaged across vowels, voiced-voiceless tokens, gender) for each window location, as a function of place of articulation.

Voicing had a significant effect on fourth spectral moment [$F(1,918)=6.400$, $p=0.0116$]. Voiced fricatives were characterized by higher values for kurtosis (2.178) than voiceless fricatives (1.365), revealing that the first had better defined peaks. Gender was proved to be significant for spectral mean [$F(1,918)=23.535$, $p<0.0001$], indicating higher values for female than male speakers (5367 vs. 4953, respectively). Post-fricative vowel was significant for first [$F(4,915)=7.596$, $p<0.0001$], third [$F(4,915)=7.362$, $p<0.0001$] and fourth moment [$F(4,915)=4.990$, $p=0.0006$], since post hoc tests revealed significant differences mainly between fricatives before [a] - [e] and [a] - [i].

Table 2. Mean spectral moment values for each place of articulation, averaged across window location, speaker's gender, voiced-voiceless tokens and vowel context.

Place of articulation	Mean (Hz)	Variance(Hz)	Skewness	Kurtosis
Labiodental	5484	2721	0.270	-0.278
Dental	5738	2649	0.161	0.181
Alveolar	5700	1526	0.453	1.079
Palatal	4471	1573	1.508	3.731
Velar	3905	1995	1.496	9.647

Conclusions

The results indicate that both duration and spectral moments provide important information for the acoustic structure of Greek fricatives.

First, voiced fricatives of all places of articulation are considerably shorter than voiceless ones. Also, fricatives produced by females are significantly longer than those produced by males. Post-fricative vowel examination revealed that the higher the vowel, the longer the preceding fricative. It was also found that dental and labiodental place of articulation do not exhibit distinctive values for all spectral moments. Velars are distinguished by all four moments and palatals by three (mean, skewness, kurtosis), though alveolars are highly related to labiodentals and dentals. They can be distinguished from labiodentals by variance and from dentals by variance and skewness. Across moments, window location 1 (onset) contains the most distinctive in-formation. Finally, voiced fricatives can be distinguished from voiceless ones, apart from duration, by the fourth moment (kurtosis).

References

- Baum S.R., Blumstein S.E. 1987. Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English. *Journal of the Acoustical Society of America* 82, 1073-1077.
- Behrens, S.J., Blumstein, S.E. 1988. Acoustic characteristics of English voiceless fricatives: A descriptive analysis. *Journal of Phonetics* 16, 295-298.
- Fourakis, M.. 1986. A Timing model for word-initial CV syllables in Modern Greek. *Journal of the Acoustical Society of America* 79, 1982-1986.
- Jongman, A., Wayland, R., Wong, S. 2000. Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America* 108, 1252-1263.
- Nicolaidis, K. 2002. Durational variability in vowel-consonant-vowel sequences in Greek: The influence of phonetic identity, context and speaker. *Selected Papers on Theoretical and Applied Linguistics from the 14th International Symposium*, 280-294, April 20-22, 2000, Thessaloniki.

The IPA training on the pronunciation of difficult English words for Cantonese speakers

Sunyoung Oh

Department of Chinese, Translation and Linguistics, City University of Hong Kong,
Hong Kong

Abstract

This study examined English pronunciation by Cantonese speakers in Hong Kong with the International Phonetic Alphabet (IPA) to see how the application of IPA can be used in language education. Twenty university students, divided into two groups, one with the IPA training and the other without the training, were asked to read 30 English words (5 controls) that contain challenging sounds (e.g., *liaison*, *knead*) with and without the IPA transcriptions. Words were divided into three categories - difficult vowels, silent sounds, and ambiguous sounds. Results showed that the pronunciations by the IPA trained group improved greatly for all categories when IPA was provided; however, the response time was longer compared to the non-trained group. No significant improvement was found in the non-trained group though the pronunciation of the words with silent sounds improved with the IPA. The study suggests that the IPA training can be an alternative way to assist ESL learners to have access to proper English pronunciations when resources are limited.

Key words: Phonetic alphabet, pronunciation, second language acquisition, English, Cantonese

Introduction

English has been used in Hong Kong for many decades. While Cantonese, the first language largely spoken in Hong Kong, Macau, and Guangdong regions of the Southern China, has been a dominant language in daily use, English has played a role as an instructional language in education. With a recent increase in the use of Putonghua (Mandarin Chinese) after the 1997 handover, the use of English has been rather reduced to classroom language, and, contrary to what many people think, most students in Hong Kong, except for those who go to international schools, do not have enough exposure to native English during their primary and secondary education. Lack of native English speakers and/or trained ESL teachers in the language learning environment in addition to the interference with Cantonese phonology has been a barrier to a certain degree to mastering correct sounds, and, as a result, incorrect perception of English sounds and pronunciation errors are very common among Cantonese speakers (Chan & Li 2000; Yip & Oh, *forthcoming*). The International Phonetic Alphabet (IPA) can be used as an alternative way to provide an optimal learning output in order to minimize incorrect input of English sounds under the circumstances of a shortage of trained

teachers and native English speakers in the language learning environment in Hong Kong. IPA has been widely used in linguistics such as phonetics, phonology, language documentation, and child language transcription and suggested that the IPA could help second language learners to master the correct forms of words in a target language (IPA 1999, Pullum & Ladusaw 1996). However, not much study has been done on its actual application to language education or the effects of the IPA training in language learning in Hong Kong. Thus, the current study investigates how IPA training can be used to assist language learners to improve their pronunciations of difficult English words.

Methods

Participants

Twenty speakers of Hong Kong Cantonese participated in the study. The participants were divided into two groups, based on their knowledge of IPA: One group with previous IPA training (nine female and one male) and the other group with no IPA training (three female and seven male). The mean age of the IPA trained group was 23 years ranging from 21 to 28; and the mean age of the non-trained group was 24 years ranging from 22 to 29. All participants were born and raised in Hong Kong, went to local schools taking English courses since Form 3 (equivalent of grade 9), and had no foreign friends at the time of the study. None of them, except one speaker with a short-term stay, had studied or travelled overseas at the time of the study. All participants were undergraduate students.

Stimuli and Procedures

Twenty five English words that are difficult to read were selected along with five control words from Kwok (1993). Test words consisted of three categories: (1) words with a silent sound (e.g., *debris*), (2) words contains an ambiguous sound 'g' or 'ch' (e.g., *analogous*), and (3) words with double vowels (e.g., *quay*). The target words were presented twice to the speakers. In the first trial, the test words were shown without the IPA transcriptions; and in the second trial, the same words were presented with the IPA transcriptions (e.g., *debt* [det]). The speakers were recorded individually, producing all thirty words twice (20 speakers x 30 words x 2 repetitions). The recordings were made using a MP4 recorder with a built-in microphone in a quiet room. Data of total 1200 tokens were digitized and sampled at 44010 HZ for acoustic analysis.

Results

Findings were presented for each trial for the two speaker groups (non-IPA, IPA) in three word categories (double vowels, silent sounds, ambiguous sounds). Comparisons were made for overall performance by category, improvement rate and response time for the speakers. The ANOVA showed that the means for Speaker Group, IPA transcriptions (before, after) were all significant ($p < .0001$, $p < .0009$).

Word Category

Results showed improvements in pronunciation for all three categories for the IPA-trained group after IPA transcriptions were provided. For the non-trained group, no effect of the IPA on the pronunciations was found; however, the pronunciations of the words with a silent sound (e.g., *receipt*) improved with the IPA transcriptions. The performance by the groups were significant for both trials with ANOVA and *t*-test ($p < .0001$).

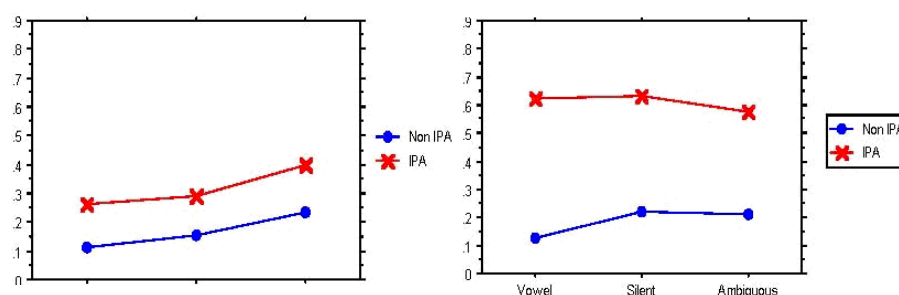


Figure 1. The performance of the correctly pronounced words by the two groups of speakers are shown in three word categories without (left)/with (right) IPA transcriptions.

Response Time

While the pronunciations by the trained group improved for all categories with the IPA transcriptions, the increase of response time to the target words with the IPA was greater for the trained group than the non-trained group.

Table 1. Response Time to the test words (in second).

	Non-trained Group	IPA-trained Group
Trial 1	8.39	9.75
Trial 2	6.85	9.82

Discussion and Conclusion

The effect of the IPA on the pronunciation was significant for English words with challenging sounds. The IPA-trained group benefited from the IPA for all categories with which showed above 60% of improvement, but the words with double vowels received the most improved pronunciations with the IPA, as shown in Figure 1. On the contrary, the non-trained group did not show any improvement in pronunciation with the IPA, and the words with double vowels were the most difficult ones to pronounce correctly. This suggests that the IPA training is effective to learn different vowel sounds as the IPA symbols can specify various vowels that cannot be learned from the English alphabet. Moreover, the IPA transcription was useful even to the non-IPA trained group: absent consonant symbols in the transcription had helped the speakers realize that those were silent sounds.

The first language learners are naturally exposed to a target language whereas the second language learners are not given such an environment and face interference with their first languages. For effective learning of pronunciation, enough exposure to a target language is necessary for words to be stored in one's lexicon for the correct use; however, when such resources are limited or not available, IPA could assist ESL learners to acquire correct pronunciations of English. This study provides a foundation for the application of the IPA in language education.

Acknowledgements

This project was supported by Grant 7002192 from City University of Hong Kong. The author thanks to Amy Yip and Carmen Suen for data collection.

References

- Chan, Y.W, Alice, Li, C.S. David. 2000. English and Cantonese Phonology in Contrast: Explaining Cantonese ESL Learners' English Pronunciation Problems. *Language, Culture and Curriculum*, Vol. 13, No 1.
- International Phonetic Association. 1999. *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge; New York: Cambridge University Press.
- Kwok, D. 1993. *Practical English Phonetics and Phonology*. Hong Kong: International Phonetic (H.K.) Association.
- Pullum, G.K, Ladusaw, W.A. 1996. *Phonetic Symbol Guide*, 2nd edition. University of Chicago Press.
- Yip, L., Oh, S.Y. forthcoming. English pronunciation teaching with IPA in Hong Kong.

Towards an articulatory characterization of European Portuguese /l/

Catarina Oliveira¹, António Teixeira², Paula Martins³

¹Health School (ESSUA)/ IEETA, University of Aveiro, Portugal

²DETI/ IEETA, University of Aveiro, Portugal

³Health School (ESSUA)/ IEETA, University of Aveiro, Portugal

Abstract

Based on EMA data, this study investigates timing and magnitude properties of lingual gestures involved in the production of European Portuguese /l/ in onset and coda positions. Syllable position effects were found: coordination pattern for syllable-initial /l/ is distinct from those observed for syllable-final /l/; the tongue dorsum is more retracted in the syllable-final /l/; /l/s in coda showed, for one of the speakers, a reduction in magnitude of the tongue tip gesture.

Key words: (European) Portuguese /l/, EMA, clear/dark allophony

Introduction

For the last decades, it has been argued that European Portuguese (EP) exhibit allophonic variation in /l/. The “clear” variant occurs syllable-initially and the “dark” variant occurs in coda position (Mateus & d’Andrade 2000). However, this point of view is not completely consensual (Andrade 1999). Unfortunately, the majority of the descriptions are based on impressionistic observations and the only extensive empirical descriptions of Portuguese /l/ come from the acoustic data of Andrade (1999).

For other languages, the most commonly cited articulatory distinction between the two allophones is that there is a greater tongue dorsum retraction in the dark /l/ than in light variety (Giles & Moll 1975; Sproat & Fujimura 1993; Narayanan 1997). Also, the consonantal gesture of alveolar contact is considered to be articulatorily stronger in onset position, and weaker in coda position (Giles & Moll 1975; Browman & Goldstein 1995; Gick 2003). The tongue dorsum gesture does not follow the same pattern and seems to be unaffected by syllable position (Sproat & Fujimura 1993). Studies about the temporal organization of /l/ gestures (Browman & Goldstein 1995, Sproat & Fujimura 1993, Gick 2003) have also revealed different timing patterns that distinguish initial and final allophones. In syllable-final /l/s, tongue dorsum movement precedes tongue tip raising, while in initial allophones, tongue tip constriction tend to occur earlier than, or simultaneous with, tongue dorsum gesture.

The purpose of this paper is to improve the knowledge of the articulatory properties of EP /l/ in order to bring some new light into the controversial

question whether the Portuguese /l/ is categorically associated with two positional allophones or not. Based in Electromagnetic Articulography (EMA) data, tongue gestures amplitudes and coordination were investigated for onset and coda /l/.

Experimental method

Data collection

Two native speakers of European Portuguese (AT and PM) were recruited for the experiment.

The speech material includes short meaningful expressions (e.g. “laca de cabelo” “hair spray”), where /l/ occupies the word-initial position, intervocalic position and coda position. The lateral was flanked by corner vowels ([a], [i], [u]). Four to six repetitions of each item were collected for each speaker.

Recording sessions for both subjects took place at the Gipsa-lab in Grenoble, using the Carstens AG100 Articulograph. Kinematic data were sampled at 500Hz and acoustic speech signal was sampled at 20 kHz. EMA coils were placed on each subject's tongue (near the tongue tip and at the dorsum), upper lip, lower lip and jaw. In this study we focus on data from the tongue tip (TT) and tongue dorsum (TD) coils. Standard calibration and post-processing procedures were applied.

Data analysis

The EMA data were processed in *Matlab*, using routines written by the second author. Following Nakamura (2009), the time and position for a constriction were defined as the minimum of the tangential velocity of TT and TD. Gestural coordination was quantified as the time between gestural landmarks (*tip delay*), algorithmically derived from local velocity profiles of the relevant receivers in the vocal tract.

Results

A first topic of investigation was the **magnitude** properties of the two gestures (TT and TD) in the production of syllable initial and final /l/.

The analysis of the TT vertical position (TTy) revealed that for AT coda /l/ exhibits a significantly lower position than onset /l/ while for the other speaker (PM) the situation is the opposite (Fig1). The TD horizontal position (TDx) is also affected by syllable position in both speakers: the coda /l/ shows a more posterior position than the onset /l/.

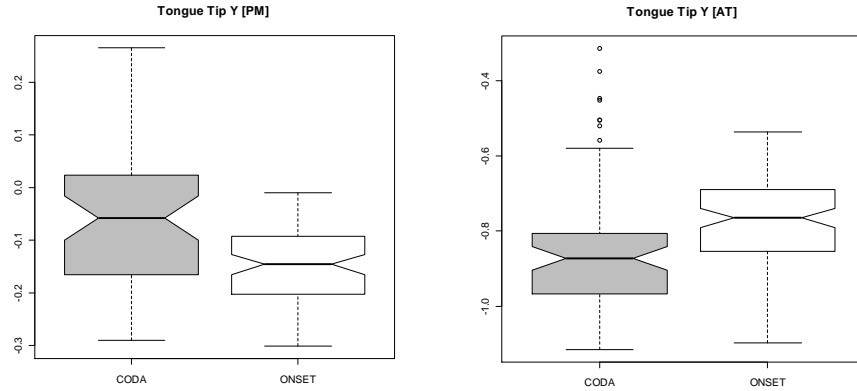


Figure 2 – TTy displacement for /l/ in onset and coda (PM and AT).

A second research goal of this study was to measure the **intergestural timing** between tongue tip and tongue dorsum movements in the two syllable positions. For both speakers (except for one vowel), the results revealed positive values for coda /l/ and negative values for onset /l/. Positive value means that the achievement of the TD gesture occurred earlier than the TT gesture, while a negative value indicates that TT preceded TD (Fig2).

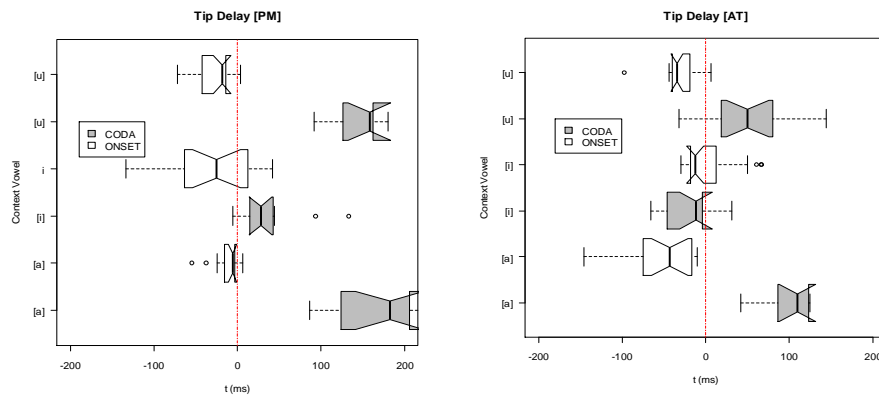


Figure 3 - Tip delay for the onset and coda /l/ (PM and AT).

Conclusion

Using EMA data, this paper investigated differences in articulatory properties of /l/ as function of syllable position. The preliminary results

reveal some changes in the spatiotemporal parameters of the tongue gestures (TD and TT) for the onset and coda /l/.

Evidence that syllable-final /l/ is associated with lower positions of the tongue tip than syllable-initial /l/ was observed only for one subject. Although little differences in the tongue body magnitude of the two varieties of /l/ were expected, our results showed that the tongue dorsum tended to be further back in the syllable-final than syllable-initial productions. The “tip lag” tendency predicted for /l/ on the basis of Sproat & Fujimura’s (1993) findings is also apparent in the present data.

The temporal findings are in general agreement with those obtained for languages with clear/ dark allophony (e.g. English). However, the reduction in magnitude of the TT gesture was not consistent across subjects. Since our study is limited to two speakers, to fully understand these patterns and to bring some light into the question of allophonic variation in EP /l/, we need more information. A systematic examination of the roles of the different articulators (using MRI) and acoustic consequences is now being carried out.

Acknowledgements

This work was supported by FCT (grant SFRH/BD/18058/2004 and HERON project). Many thanks are due to Solange Rossato and Christophe Savariaux (GIPSA-Lab), for help with the EMA recordings.

References

- Andrade, A. 1999. On /l/ velarization in European Portuguese. In Ohala, Y. Hasegawa, M. Ohala, D. Granville, A. Bailey (eds.), 14th International Congress of Phonetic Sciences, 543-546, San Francisco.
- Browman, C., Goldstein, L. 1995. Gestural syllable position effects in American English. In Bell-Berti, F., Raphael, L.J. (eds.) 1995, *Producing Speech: Contemporary Issues: For Katherine Safford Harris*. Woodbury, AIP Press.
- Gick, B. 2003. Articulatory correlates of ambisyllabicity in English glides and liquids. In Local, J., Ogden, R., Temple, R. (eds.), *Papers in Laboratory Phonology VI: Constraints on Phonetic Interpretation*, 222-236. Cambridge University Press.
- Giles, S., Moll, K. 1975. Cinefluorographic study of selected allophones of English /l/. *Phonetica* 31, 206-227.
- Mateus, M.H.M., d’Andrade, E. 2000. *Phonology of Portuguese*. Oxford University Press.
- Nakamura, M. 2009. Exploring Vocalization of /l/ in English: an EPG and EMA study. In *Proceedings of Interspeech 2009*. Brighton, UK.
- Narayanan, S., Alwan, A., Haker, K. 1997. Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals. *Journal of the Acoustical Society of America* 101, 1064-2007
- Sproat, R., Fujimura, O. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21, 291-311.

A complexity science perspective on language spread

Michał B. Paradowski

Institute of Applied Linguistics, University of Warsaw, Poland

Abstract

Linguistic creativity is a manifestation of innovativeness. Investigating the diffusion of linguistic innovation requires an interdisciplinary approach, pooling competences from sciences humanistic, social, and strict (in particular mathematics and computational methods). The initial results will be presented of an empirical project analysing the social spread of neologisms in a microblogging site.

Key words: linguistic creativity, neologisms, innovation diffusion, social networks

A complex system: more than a complicated system

A complex system must be distinguished from a complicated one. While no unanimously agreed definition exists of what constitutes a complex system, most researchers agree that in order to be classified thus, a system has to share most of the following properties (Paradowski 2009b):

- a large number of elements interacting via simple local rules;
- the emergent global behaviour of the system is not a simple product of the sum of the behaviours of its components (“more is different”);
- dynamic and evolving over time, with emergent properties;
- topological diversity, heterogeneity (and directedness);
- strong self-organisation (without an orchestrator), resulting in
- resistance to damage, resilience to failure and high flexibility due to
- learning and natural adaptability to changing conditions;
- ostensibly negligible perturbations and rare events on the local level lead to significant changes in the behaviour of the whole system;
- bifurcation and phase transitions;
- sensitivity to initial conditions;
- stability and multistability;
- hysteresis (with the consequence that the system is nondeterministic);
- scale-free – self-similarity;
- heavy-tailed distribution scaling like a power law;
- entropy (many possible ways of evolution);
- typically impossible to solve/verify/predict analytically – hence the fundamental role played by numerical simulation.

Examples of complex systems can be found at every step. In biology, ontogeny and phylogeny, the spot character of a cheetah, structure of the leaf, ant trails, termite mounds, flashing fireflies, chirping crickets, and

predator-prey ecosystems are all instantiations of complex systems. So are communities and social behaviours, from cooperation to conflict and riots. Complex are epidemics, nervous and immunological systems, neural networks, the climate and earthquakes, financial markets and price fluctuations, logistics and traffic jams. In the sphere of technology, the Internet, telecommunications infrastructure and power grids, and the World Wide Web instantiate complex networks. Last, but not least, the field where it all started – physics – has long been grappling with analyses and descriptions of non-equilibrium thermodynamics, crystals, boiling liquid, or the nuclear fission reaction.

In linguistics, within an individual, perceptual dynamics and categorisation in speech or word and sentence processing; across society, variations and typology, the rise of new grammatical constructions, semantic bleaching, language evolution in general, and the spread and competition of both individual expressions, and entire languages, are all inherently complex systems. Over 80 papers have been published dealing with language simulations. However, the vast majority of the *in silico* experiments carried out are inadequate to the scenery of the 21st c. (Paradowski 2009a). The models:

- only allow for Euclidean relationships (whereas nowadays more and more of our linguistic input covers immense distances);
- are ‘static’ (while mobility is not exclusively a 20th or 21st-c. phenomenon, as evidenced by warriors, refugees, missionaries, or tradespeople);
- assume a limited, identical number of ‘neighbours’ for every agent;
- presuppose identical perception of a given individual’s prestige by each of its neighbours, as well as
- invariant intensity of interactions between different agents;
- most fail to take into account multilingual agents;
- have no memory effect, and
- zero noise (while noise may be a mechanism for pattern change).

To address these limitations, rather than take a modelling outlook, we can tap into one repository of language data nearly perfectly suited to large-scale dynamic linguistic analyses – the Internet. After all, this medium stores data which is virtually unregulated, essentially uncensored, spontaneous, being immediately registered, interconnected, and amenable to relatively easy search and analyses with the use of statistical and concordancing tools.

Language on the Internet

Hitherto research on language evolution and change focused on large time-scales, typically spanning at least several decades. Nowadays, observable changes are taking place much faster. According to the Global Language

Monitor a new English word is born roughly every 98 minutes. As a handful of popular recent expressions recall *alcopop*, *Ardi*, *bangster*, *brown state*, *choice mom*, *death panel*, *defriend*, *deleb*, *e-vampire*, *freemium*, *fundoo*, *funemployed*, *green washing*, *intexticated*, *jai ho*, *mobama*, *n00b*, *octomom*, *quendy-trendy*, *recessionista*, *seatmate of size*, *sexting*, *slumdog*, *teabagger*, *tramp stamp*, *wonderstar*, or *zombie bank*... Particularly useful for multi-angle analyses of language phenomena are Web 2.0 services, with content (co)generated by the users, especially the ones which allow enriching analyses with information concerning the structure of the connections and interactions between the participating users. Our empirical research project has set out to investigate how mutual relations and communication between Internet users impact the social diffusion of neological tags (semantic shortcuts) in a Polish microblogging site (since diffusion of linguistic innovation is a function of social structure; Rogers 1962/2003; Watts 2007).

Table 1. The microblogging site in numbers (at time of data dump)

Users	20k, over half logging on daily
Users in the giant component	5.5k (density 0.003)
Relations	110k
Tags	40k
Tagged statuses	720k

Predictably, the frequency distribution of tags scales like a power law, both for short (24-hr) and long (1-yr) timescales. The most interesting is tracking those tags which are idiosyncratic to the system (i.e., not used in offline discourse), where the only way to pick them up is from a neighbour in the site. This allows plotting the dynamics of the spread of the items in a closed circuit, relative to its structure and size.

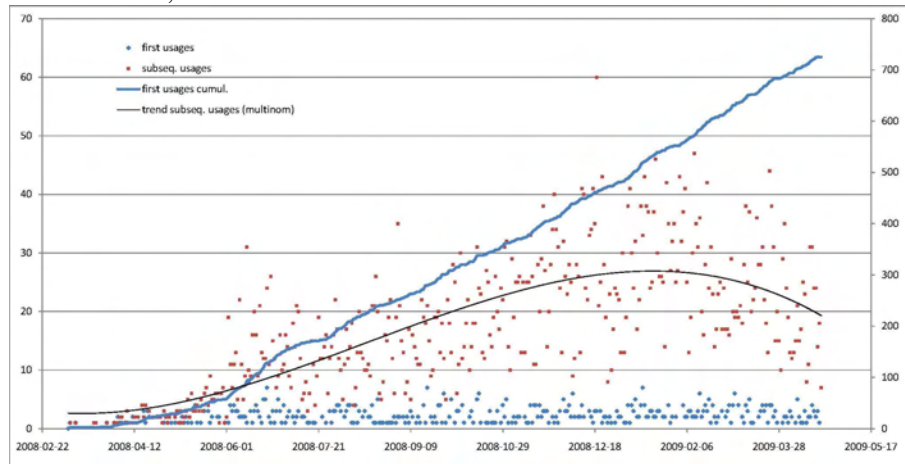


Figure 1. Usage of an idiosyncratic tag (Paradowski et al. 2010).

Social influence and diffusion

The general innovativeness of Internet users scales not like a power law, but a hump-shaped unimodal (Paradowski, Jonak & Kuscsik 2009). The exposure thresholds specifying what proportion of one's neighbours must take up a certain activity for the individual to be likely to pursue it (in line with Granovetter's (1978) threshold model of collective behaviour and Valente's (1995) theory of network exposure) concentrate at low values, suggesting more susceptibility to social influence than may erstwhile have been expected, high innovativeness of the users, and 'virality' of tags. It may also mean that the innovative users are the ones with a greater number of subscriptions (Jonak & Paradowski 2009).

Acknowledgements

The research has been supported by a grant from the Polish Society for Social Psychology, Agora SA, and Polskie Badania Internetu.

References

- Granovetter, M. 1978. Threshold Models of Collective Behavior. *The American Journal of Sociology* 83(6), 1420–43.
- Jonak, Ł., Paradowski, M.B. 2009, Sep 16. Ekspansja słowa – analiza dynamiki rozpowszechniania się neologizmów leksykalnych i frazeologicznych oraz wzajemnych interakcji w społecznościach internautów. Invited presentation, VI scientific rally of the PSPS "Psychologia społeczna: teorie i badania", Lublin.
- Paradowski, M.B. 2009a, Jun 25. Complexity science and language phenomena – a successful marriage or a mésalliance? Invited talk, Complexity Science Seminar, Inst Math Sci, Imperial Coll London.
- Paradowski, M.B. 2009b, Oct 10. Applying a complexity science approach to analysing & modelling language phenomena. Invited lecture, Higher English Language Seminar, Dept English, Stockholm Univ.
- Paradowski, M.B., Jonak, Ł., Kuscsik, Z. 2009, Nov 23. The evolution of speech – towards network models of language development and spread. Poster, Intl Workshop "150 Years after Darwin: From Molecular Evolution to Language". Inst Cross-Disciplinary Phys & Complex Syst, Palma de Mallorca.
- Paradowski, M.B., Jonak, Ł., Kuscsik, Z. 2010, Jan 27–8. Follow the white rabbit, or how complexity science and linguistics can inform AI. 4th Intl Conference on Cognitive Systems, Inst Robotics and Intell Syst, ETH Zürich.
- Rogers, E.M. 2003. *Diffusion of Innovations*. New York, Free Press.
- Valente, T.W. 1995. *Network Models of the Diffusion of Innovations*. Cresskill, NJ, Hampton Press.
- Watts, D.J. 2007. The Accidental Influentials. *Harvard Business Review* 85, 22–3.

***Inda mbu, nambu or ine ti pu?* Acquiring complex structures in Cypriot Greek**

Natalia Pavlou

Department of English Studies, University of Cyprus, Cyprus

Abstract

The current study investigates the hypothesis that certain Cypriot Greek (CG) *wh*-phrases show similar paths of acquisition with cleft sentences based on syntactic complexity. That is, *wh*-phrases of the *inda (mbu)*-type in CG can be decomposed in an abstract way as clefts, giving *ine ti pu* ‘what is it that’. Targeting a comparison between these two structures, *wh*-questions with *inda (mbu)* along with subject- and object-clefts were included in a production elicitation experiment given to children aged 2;7-6;4. The preliminary results raise several issues ranging from syntactic complexities and movement difficulties in the two structures to sociolinguistic factors and implications for the possible diglossic situation in the Greek Cypriot community.

Key words: acquisition, diglossia, clefts, *wh*-questions

Why clefts?

Several issues in CG have been explored through the syntactic and pragmatic environment of clefts, which are widely used by Greek Cypriot speakers. Fotiou (2009) gives a first insight to the pragmatic role of the clefts and discusses the notion of focus and its different interpretation in CG clefts. With regard to *wh*-questions, Grohmann et al. (2006) suggest the ‘*embu*-strategy’ for the Cypriot-specific element *embu*, which is analyzed by them as a cleft: *en(m) pu* ‘(it) is that’.

More generally, a cleft-strategy is assumed in this study to be the underlying form of *inda (mbu)*-questions in CG, which are restricted to *wh*-objects, true adjuncts, and D(iscourse)-linked *wh*-phrases (1). It should be noted that *mbu* is assumed to be a variant of *embu* and appears in contexts where *embu* is not allowed (Grohmann et al. 2006, Pavlou 2010), and *inda* ‘what’ (excluded from the experiment because it is marginally accepted and used in Cyprus) emerged from an interrogative pronoun.

- (1)a. *Inda mbu vasta* o andras?
 what hold-3SG the man-NOM
 b. *Inda mbu fonazi* o andras?
 why shout-3SG the man-NOM
 c. *Inda klei* o andras?
 why cry-3SG the man-NOM

d. Inda vivlio θkiavazi o andras?
which book read-3SG the man-NOM

The observable decomposition of *inda (mbu)* as *ine ti pu* (or *ti ine pu*) ‘what is (it) that’ is expected to have some effects on the ‘special’ acquisition of the CG clefts, *wh*-phrases and questions, and consequently their allomorphs (Pavlou 2010).

The Guess What Game (GWG)

The GWG is designed to function as a production elicitation game for *wh*-questions and clefts with the use of pictures. The children are introduced to the GWG and requested to ask two supposedly ignorant puppets some questions. Each picture has two subjects and two items, with only one subject performing an action with one item. While the child is asking questions, s/he and the puppets are involved in a competition game, where the one who gives the most correct answers gets a chick from a guarded coop. If the puppets give wrong answers to the child’s questions, then the child is awarded with a chicken.

The six different structures are ranging from the 4 *inda*-questions, given in (1), subject- and object-clefts, and are all randomized in 6 blocks, with two trials in the warm-up session and 4 test sentences.

The subjects were taken from kindergartens in the urban area of Limassol, after teacher and parental consent. Both parents of all children tested were Greek Cypriot. The children did not have any other learning difficulties or language disorders characteristics according to their teachers.

Results

Concentrating more on the type of the responses, the main results of this study will be presented here in three groups, the first dealing with *wh*-questions and the issue of diglossia in Cyprus. The relevance of this linguistic situation in Cyprus played a role to the lexical choice of the *wh*-phrases used. As it appears in the results, the children replaced the target *wh*-phrases with the corresponding ‘Standard Modern Greek (SMG)-like’ *wh*-phrases. This could be an immediate effect of the low status of CG and the code-switch to the ‘high’ variety, which is SMG, as well as the perception of the researcher as a stranger to the children.

The production of clefts was characterised by the development of different strategies such as the replacement of both subject and object clefts with SVO declaratives (2a), normally expressing focus with the fronting of the subject in a focus-position. Other types of responses show lack of complete acquisition of clefts. Reduced clefts, as in (2b), were used for subject clefts

and elliptical clefts (2c–d) for both cases. Focus in-situ (2e) was also met, especially for object clefts. In all cases, it is clear that children comprehend the pragmatic function of a cleft, which is focus, and develop different strategies to express it by avoiding the complexity involved by the clefts.

- (2) a. O andras krata aftokinitaki
 the man-NOM hold-3SG car-ACC
 b. En o andras
 is-3SG the man-NOM
 c. *O andras pu krata aftokinitaki
 the man-NOM that hold-3SG car-ACC
 d. *Aftokinitaki pu krata o andras
 car-ACC that hold-3SG the man-NOM
 e. O andras krata to KOKKINO aftokinitaki
 the man-NOM hold-3SG the RED car-ACC

Overall, the production of clefts started at the age group 4;0–5;0 with very low percentages. Interestingly, both subject- and object- clefts decrease in 5-year-old children and then increase again in the 6-year-old age group.

Thirdly, common errors were widely found in D-linked *wh*-questions and divided in two categories. The percentages of errors in complex *wh*-questions were relatively high in 4- and 5-year-old children. Errors which are marked as (–) pied-piping represent responses like (3a), whereas NP-omission was used for sentences like (3b). In both of these, the target question was ‘Which car is the man holding?’ Complex *wh*-questions are generally perceived more difficult, hence assumed to be acquired late. However, there were systematic errors, showing the movement of the *wh*-word without the NP which was left in its base position. These errors hold the same percentages in 4- and 6-year-old children and appear less in 3- and 5-year-old children. NP omission, and consequently inversion of referential to argumental *wh*-questions, similarly to Greek (Stavrakaki 2006), was very strong in 5-year-old children, while also appearing in 4- and 6-year-old children.

- (3) a. # Inda (or ti) krata aftokinitaki o andras?
 which hold-3SG car the man-NOM
 b. # Ti (or inda) krata o andras?
 which hold-3SG the man-NOM

A determining factor for explaining these errors was the calculation of *wh*-word alternation between CG and ‘SMG-like’ *wh*-words. The problematic

responses in 2- and 3-year-olds appear low with CG 'which', whereas both types of errors are greater in number in 4-, 5-, and 6-year-old children with the use of the 'SMG-like' *wh*-phrases.

Some First Thoughts

The combination of different structures has given a 'report' character to this ongoing study by providing some first data for certain complex structures. The sociolinguistic factors affecting the results, as well as the syntactic complexities observed have set the need for further and more specific research in these topics. Greek Cypriot children perform differently in complex *wh*-questions from Greek children (Stavrakaki 2006), that is, from (mainland) Greece speaking SMG, and the *wh*-word alternations between CG and SMG observed show that the use of SMG in pre-school years may start earlier than assumed in kindergartens.

The Cyprus Acquisition Team (CAT) is currently investigating the exact time and effects of SMG instruction and use in both the pre-school and school years through specific structures in an attempt to identify the pre-school variety used by Greek Cypriot children. Presumably, young speakers develop their linguistic awareness for the difference between CG and SMG very early and thus can make decisions for it, which in this case appeared as non-CG lexical choices for the aforementioned *wh*-phrases. However, first results indicate that conscious decisions may not be the only or even most relevant factor for CG-SMG alternations, but the overall use of SMG can affect certain structures and their acquisition.

References

- Fotiou, C. 2009 Focusing strategies in Cypriot Greek. In K.K. Grohmann and P. Panagiotidis (eds.), *Selected Papers from the 2006 Cyprus Syntaxfest*, 63-91. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Grohmann, K.K., Panagiotidis, P., Tsiplakou, S. 2006. Properties of *wh*-question formation in Cypriot Greek. In M. Janse, B.D. Joseph, A. Ralli (eds.), *Proc. 2nd Intern. Conference on Modern Greek Dialects and Linguistic Theory* (Mytilene, Greece: 30 September–3 October 2004). Patras: University of Patras, 83-98.
- Papadopoulou, E. In progress. *Acquisition of wh-questions: Evidence from Cypriot Greek*. University of Essex, Colchester: PhD dissertation.
- Pavlou, N. 2010. MBU! On *wh*-objects and true adjuncts of Cypriot Greek. In a. Ralli, B.D. Joseph, M. Janse, A. Karasimos (eds.), *Proc. 4th Modern Greek Dialects and Linguistic Theory Conference*. University of Patras, Department of Philology, Linguistics Division, 155-178. (electronic publication)
- Stavrakaki, S. 2006. Developmental perspectives on specific language impairment: Evidence from the production of *wh*-questions by Greek SLI children over time, *Advances in Speech-Language Pathology* 8, 384-396.

Two languages – one brain

Maija S. Peltola^{1,2}, Henna Tamminen^{1,2}, Laura Salonen¹, Heidi Toivonen³,
Teija Kujala⁴, Risto Näätänen^{4,5,6}

¹Department of Phonetics, University of Turku, Finland

²Centre for Cognitive Neuroscience, University of Turku, Finland

³Department of Logopedics, Åbo Academy University, Finland

⁴Department of Psychology, CBRU, University of Helsinki, Finland

⁵Centre of Functionally Integrative Neuroscience, University of Århus, Denmark

⁶Department of Psychology, University of Tartu, Estonia

Abstract

Studies on bilingual speech processing have implied that bilinguals may either have two separate or one intertwined system. These findings have been obtained by multiple methods using various types of bilinguals. Our study compared monolinguals and two types of bilinguals. We used the same methods for all groups, i.e. we measured attentive identification scores and preattentive discrimination. Our results show that bilinguals process speech sounds differently from monolinguals, and more importantly, that there is a difference between the two types of bilinguals. We suggest that dominant bilinguals have two separate phonological systems, while balanced bilinguals have one uniform system.

Key words: speech perception, bilingualism, mismatch negativity.

Introduction

Earlier research has revealed that speech sound processing is language-specific (Näätänen et al. 1997). This is seen in the mismatch negativity (MMN) component of the event-related potentials (ERPs), which has an extensive amplitude and short latency when a native language category boundary is crossed. The acquisition of new non-native speech sounds may also be studied by using the MMN response. Representations may emerge in immigration (Winkler et al. 1999) and in early exposure (Peltola et al. 2005), while tentatively not in classroom (Peltola et al. 2003). Bilingual speech processing studies have also considered the neural organisation of the two languages: Winkler et al. (2003) showed that immigrants had similar kinds of MMN-responses to both their languages irrespective of the linguistic context. The “one-store model” was also apparent in Chee et al. (1999) study on early bilinguals (pre 7 years). In contrast, Peltola et al. (2005) suggested that classroom learners have two separate phonological systems, which is in accordance with the “two-store-model” shown by Paulesu et al. (2000). One view is also that there is a language “switch” as indicated by the presence of additional neural control (Hernandez et al. 2001). Methodology may explain some of the contradictions, since procedures either keep the languages apart,

or they demand switching. While subjects may all be bilinguals by one definition, their linguistic backgrounds vary. We performed the same protocol on two types of Finnish-Swedish bilinguals and monolingual Finns. The results indicate the significant effect of the linguistic background on neural phonological organisation.

Methods

Three groups of right handed young adult subjects with normal hearing participated in this study: Group 1: 10 native speakers of Finnish (Monolinguals), Group 2: 12 balanced Finnish-Swedish bilinguals (Balanced), who had acquired both languages from birth and Group 3: advanced Finnish University students of Swedish (Dominant).

The stimuli consisted of 18 synthetic vowels from the closed vowel continuum divided into two categories in Finnish, /y/-/u/, and three in Swedish, /y/-/u/-/u/. The identification test showed the boundary location and steepness for each subject. We selected a stimulus pair individually so that it crossed the boundary in Finnish, but was within the Swedish /u/. Monolinguals participated once, while the bilingual groups twice (Finnish and Swedish researcher). The hypothesis was that, if the two phonological systems are distinct, the MMN to the same stimulus pair could have an extensive amplitude and a short latency in the Finnish context (phonemic), but a reduced amplitude and a long latency in the Swedish context (non-phonemic).

We registered EEG in the oddball paradigm using Synamps amplifier. We analysed the MMN peak latency (150-300ms) from one electrode and the mean amplitude from Fz, Cz, F3, F4, C3, and C4 with two Time windows (180-230ms, 230-280ms). The data was statistically analysed with Repeated measures ANOVA. The category boundary location and steepness were determined with logit transformation using SPSS.

Results

When the Monolingual and Balanced bilingual data were compared, the latency analysis showed the main effect of Group ($f(1,20)=4.753, p=0.041$) revealing the later response of the Balanced bilinguals. This result was supported by the amplitude analysis, which showed the interaction Group x Time window ($f(1,20)=8.399, p=0.009$). The main effect of Group ($f(1,20)=4.834, p=0.040$) suggested an overall reduced amplitude in the responses by the Balanced bilinguals, see Figure 1.

When the two types of bilinguals were compared, the latency analysis revealed the main effect of Group ($f(1,19)=11.381, p=0.003$) implying a longer latency of the MMN by the Balanced bilinguals. This was also

apparent in the amplitude analysis showing the Group x Context language x Time window ($f(1,19)=5.639, p=0.028$) interaction. Most importantly, the Dominant bilinguals showed the main effect of Context language ($f(1,8)=8.843, p=0.018$), while there was no such effect in the Balanced bilinguals (Group x Context language ($f(1,19)=5.369, p=0.032$)). This difference is demonstrated by the almost complete lack of a response by the Dominant bilinguals to the stimulus pair in the Swedish context (non-phonemic) in comparison with the extensive MMN in the Finnish context (phonemic). The Balanced bilinguals have similar kinds of slow latency responses irrespective of the linguistic context. This result was also supported by the ID-results, which showed that while the Finnish category boundary locations matched, the steepness value was lower in the Balanced bilinguals, implying hesitation (Group x Context language ($f(2,18)=4.737, p=0.022$)).

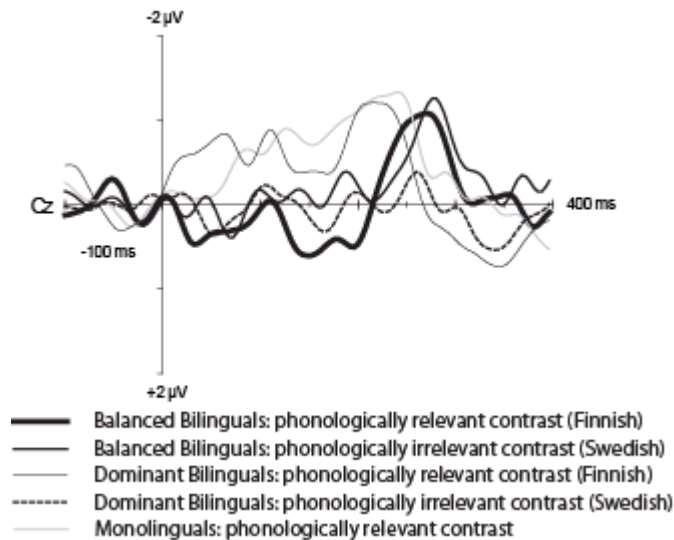


Figure 1. The MMN responses of the Monolingual, Balanced and Dominant bilingual groups to the same vowel pair.

Conclusions

Our results clearly indicate that monolinguals perceive speech sounds differently from bilinguals, which is evident from the lateness of the Balanced bilingual MMN response in comparison with that of the Monolingual group. More importantly, preattentive discrimination is dependent upon the type of bilingualism, which is obvious from the massive difference in the MMN of the bilingual groups in the perception of the pair

in the Swedish context, as well as from the behavioural transfer from Swedish to Finnish in Balanced bilinguals. This is further supported by the similarity of the Dominant bilingual and the Monolingual MMN to the native vowels. Therefore, Balanced bilinguals process speech sounds similarly irrespective of the language of the immediate vicinity, while Dominant bilinguals are context-sensitive.

We argue that Balanced bilinguals do not dissociate between their two languages even at the preattentive level of perception, which suggests that they have one intertwined phonological system. In contrast, Dominant bilinguals are able to block their maternal language in preattentive perception, which implies two separate phonological systems.

Acknowledgements

Supported by the Academy of Finland (project number 206352).

References

- Chee, M.W.L., Caplan, D., Soon, C.S., Sriram, N., Tan, E.W.L., Thiel, T., Weekes, B. 1999. Processing of visually presented sentences in Mandarin and English studied with fMRI. *Neuron* 23, 127-137.
- Hernandez, A.E., Dapretto, M., Mazziotta, J., Bookheimer, S. 2001. Language Switching and language representation in Spanish-English bilinguals: an fMRI study. *NeuroImage* 14, 510-520.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huottilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., Alho, K. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432-434.
- Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S.F., Cotelli, M., Cossu, G., Corte, F., Lorusso, M., Pesenti, S., Gallagher, A., Perani, D., Price, C., Frith, C.D., Frith, U. 2000. A cultural effect on brain function. *Nature Neuroscience* 3, 91-96.
- Peltola, M.S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., Näätänen, R. 2003. Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters* 352, 25-28.
- Peltola, M.S., Kuntola, M., Tamminen, H., Hämäläinen, H., Aaltonen, O. 2005. Early exposure to non-native language alters preattentive vowel discrimination. *Neuroscience Letters* 388, 121-125.
- Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csépe, V., Ilmoniemi, R.J., Näätänen, R. 1999. Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36, 638-642.
- Winkler, I., Kujala, T., Alku, P., Näätänen, R. 2003. Language context and phonetic change detection. *Cognitive Brain Research* 17, 833-844.

Real and acted responses of distress: an auditory & acoustic analysis of extreme stress & emotion

Lisa Roberts

Department of Language and Linguistic Science, University of York, U.K.

Abstract

The study explores vocal responses of distress in authentic and acted forensic situations. Data is drawn from an archive of forensic recordings and comparable scenarios involving actors. The material is subject to acoustic and auditory-phonetic analyses. Results are presented on parameters including fundamental frequency (F_0), tempo, intensity, and vowel quality. Findings show that victims and actors in distress generally demonstrate an increase in F_0 and intensity overall, an increase in variability of F_0 and intensity, and a decrease in tempo. Actors' distress responses differ from those from real victims in vowel quality and their exaggerated increase in F_0 .

Key words: forensic phonetics, distress speech, acted responses.

Introduction

Previous studies investigating emotional speech have concerned speech technology applications (Erickson, 2005) and/or observed vocal cues of emotion in everyday speech using actors (Scherer, 2003). Few attempt to characterise extreme emotion and fewer use authentic data. The present study represents a first step towards results that might ultimately be used to substantiate forensic expert opinions in this area.

Data and Method

Data relates to victims of violent attack where an attack is imminent, completed or in progress. Authentic material was drawn from eight forensic cases (victims: 4 male, 4 female; aged 15-47). All but one were calls to the emergency services. 9 actors (6 male, 3 female; aged 27-55) were chosen from filmed material, e.g. *Reservoir Dogs*, *A Clockwork Orange* etc. Injuries common to both victims and actors included stabbings, shootings, burnings and being chased. All material was digitised (44.1kHz, 16 bit depth). Auditory and acoustic analyses were conducted using *SoundForge* and *Praat*.

Findings

Due to the high amount of inter- and intra-individual variation, a four-way taxonomy was employed to compare like productions across all data.

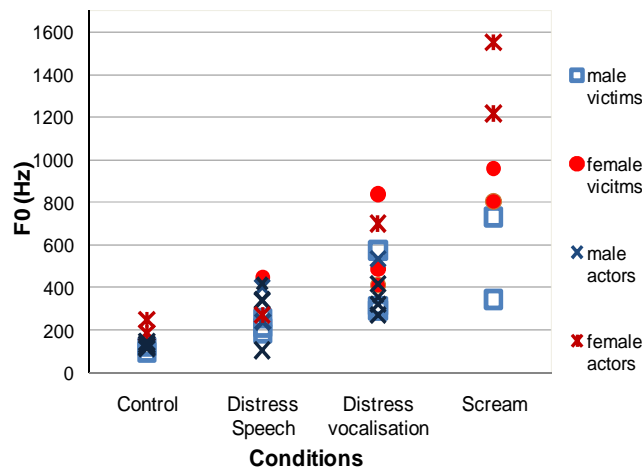
Table 1. Summary of data classification system.

Category	Control data	Distress speech	Distress vocalisation	Scream
Features	non-emotionally aroused speech, e.g. regular conversation	intelligible, emotionally aroused speech	emotionally aroused production, questionable intelligibility & linguistic content	emotionally aroused production, devoid of linguistic content

F_0

For both acted and authentic data, distress productions typically had a higher F_0 than the individual's control data/expected range, with productions increasing in F_0 as individuals progress through the taxonomy. As expected, females tend to have a higher F_0 than males, though productions classified as vocalisations and screams tend to be beyond expected ranges for both sexes.

Distress vocalisations and screams often had a rise-fall F_0 contour. These tokens contained voicing irregularities and breaks in phonation. Pitch jumps also occurred, e.g. from 510Hz to 1270Hz (vocalisation from a female victim) and from 559Hz to 810Hz (distress speech from a male actor).

Figure 4: Graph showing average F_0 (Hz) data for victims and actors.

Intensity

On the whole, levels of intensity mimicked F_0 contours, presumably as a result of increased F_0 . A plateau effect was observed in some recordings though this is more likely attributable to the limitations of the recording equipment (e.g. automatic gain control) rather than the individuals' physiology. However, results were variable across the data, for two possible reasons: first, in the authentic data, the distance and orientation of the

individual from the microphone is unknown; secondly, in the acted data, it is uncertain if sound editing took place post-production.

Vowel quality

Measurements of F1 and F2 were taken where vowels in the distress speech condition could be compared with control data, i.e. in all acted data but in only two cases of the authentic data. In the former, a contraction of the vowel space was seen; in the latter, distress vowels tended to show a shifting of the vowel space towards the front dimension. Both patterns can be associated with Lombard speech.

Tempo

Articulation rates decreased as individuals progressed through the categories of the taxonomy. Male actors had a slower AR than male victims whereas female actors tended to have a faster AR than female victims.

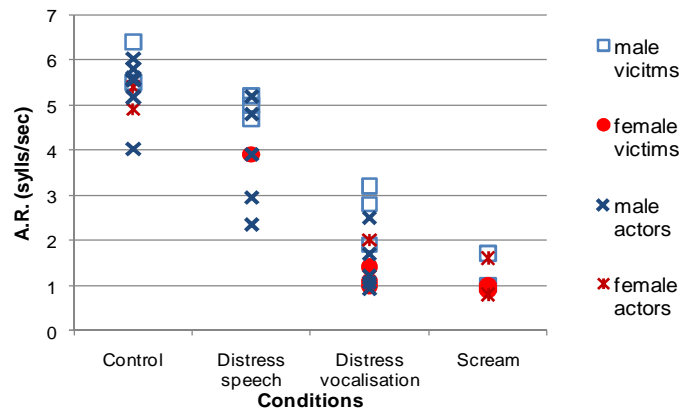


Figure 4: Articulation rates for all individuals (victims and actors).

Discussion

Acoustic changes in control and distress conditions may in part be caused by unconscious physiological responses to stress, namely changes in respiration rates and muscle tension (Hansen & Patil, 2007), giving rise to a faster tempo, an increase in intensity, and a tensing of the vocal folds with a consequent increase in F0.

Sources of variation between the data sets may be both external, e.g. due to different injuries and circumstances surrounding the victim/actor, or internal, since individuals may appraise sources of stress differently (Andreassi, 2007).

A further consideration is to what extent are distress productions in victims voluntary? It is not always clear if linguistic content is present in the productions. Involuntary responses are presumably purely physiological with

little variation. Voluntary responses, on the other hand, will most likely be greater in range. Whereas distress responses are similar in both data sets, actors may be reflecting an awareness of both a physiological response and the presence of variation amongst victims. In some cases, they seek to exaggerate these features for the benefit of the audience. Where responses differ across the data sets, actors may themselves be unsure of what they are trying to emulate.

Summary and Conclusions

Tendencies common to both real and acted distress responses are an increase in overall F_0 and intensity, an increase in variability of F_0 and intensity, with a decrease in speech tempo. Acted distress responses tend to use a contracted version of the vowel space, whereas authentic responses see an increase in F_1 and F_2 leading towards a shift in the front dimension. Acted distress responses differentiate themselves not only in regard to vowel quality, but also in the extent in which the F_0 increases – acted responses may be viewed as exaggerated versions of authentic distress responses.

Acknowledgements

This research was supported by the UK Economic and Social Research Council (PhD +3 studentship). I am very grateful to Paul Foulkes, Peter French, and Dominic Watt for their contributions, and especially J. P. French Associates for authentic forensic audio access and advice.

References

- Andreassi, J.L. 2007. Psychophysiology: human behaviour and physiological response (5th ed.). Hillsdale, N.J., Lawrence Erlbaum Associates.
- Erickson, D. 2004. Expressive Speech: Production, perception and application to speech synthesis. *Acoustical, Science & Technology* 4, 317-325.
- Hansen, J.H.L., Patil, S. 2007. Speech under stress: Analysis, modelling and recognition. In C. Müller (ed.), *Speaker Classification* 1, 108-137. Springer: Berlin.
- Scherer, K.R. 1979. Nonlinguistic vocal indicators of emotion and psychopathology. In Izard, C.E. (ed.), *Emotions in personality and psychopathology*, 495-529. New York: Plenum.
- Scherer, K.R. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40, 227-256.

How context affects perception: judging distress & linguistic content in forensic audio recordings

Lisa Roberts, Peter French

Department of Language & Linguistic Science, University of York, UK.

J. P. French Associates, York, UK.

Abstract

In order to assess the role of context on the perception of speech sounds, extracts of speech material from authentic forensic cases were played to two groups of listeners (experienced vs inexperienced forensic phoneticians) in two conditions (with and without sequential context). Listeners were asked to categorise each extract according to a four-way scale of perceptible distress and rate each extract according to a five-point scale of perceptible linguistic content. Findings show that extracts played without context are perceived as reflecting higher degrees of distress and lower levels of linguistic content. Experienced phoneticians performed more consistently as a group and were less prone to changing their responses across the two conditions.

Key words: perception, distress, forensic, context, linguistic content.

Introduction

Forensic phoneticians are often asked to transcribe and interpret recordings that are brief and of poor quality. To avoid preconceptions entering into the transcription, a 'bottom up' approach is encouraged (Fraser, 2003). This entails the phonetician undertaking the task 'blind' in the first instance, i.e., without context information and working only from the recorded sound. Some phoneticians working in this way have considered their task to have been impeded by the lack of context information (Hirson & Howard, 1994). Others, however, have written reports that may give the impression that that the significance or linguistic content of an unclear utterance was resolved mainly or wholly from examining its internal phonetic and acoustic properties (Rose 2009). Furthermore, forensic phoneticians may be asked to make assessments of the psychological states of victims in recordings. Such evaluations are currently not advocated by the International Association of Forensic Phonetics and Acoustics (IAFPA), most likely owing to a lack of sufficient research evidence to inform such an opinion (IAFPA Code of Practice, clause 9). This study examines the influence of context (sequential context together with background information) on forensic phoneticians' perceptions of the sounds uttered by victims undergoing violent attacks.

Experiment

Stimuli

Thirty-two stimulus utterances containing productions from victims experiencing violent attack were extracted using Sony Sound Forge from eight authentic forensic recordings. All victims originated from England/Wales and spoke English as their native language.

Subjects

Sixteen members of the Forensics Research Group in the Department of Language and Linguistic Science, University of York, participated in the experiment. Six participants (four male, two female) had extensive forensic casework experience and had all acted as expert witnesses (the 'experienced group'); ten participants (three male, seven female) were postgraduate student forensic phoneticians with no casework experience (the 'inexperienced group'). The inexperienced group was aged between 22 years and 25 years, whereas the ages of the experienced group ranged from 27 years to 57 years.

Method

Subjects were presented with brief stimulus sounds of which twenty four (the experimental trials) were played under two conditions: (i) without context - the sound was heard in isolation; (ii) with context – the sound was heard with preceding and following phonetic context, and background information concerning the circumstances of the recorded attack was provided. A further eight extracts acted as controls and were played devoid of context. They were repeated towards the end of the experiment to check subjects' behaviour did not change with continued exposure to the experiment. The experiment was delivered via PowerPoint Presentation 2007 and audio files were played through closed cup Sanako Tandberg Educational headphones (headset model SLH-07). Each listener was asked to: (a) categorise each excerpt in terms of a 4-way classification corresponding to the degree of distress they perceived the stimulus as representing; (b) rate the excerpt on a 5-point scale specifying the degree to which they perceived the stimulus material as having linguistic content. Two experiments were run, varying the order of presentation, in control for a possible order effect.

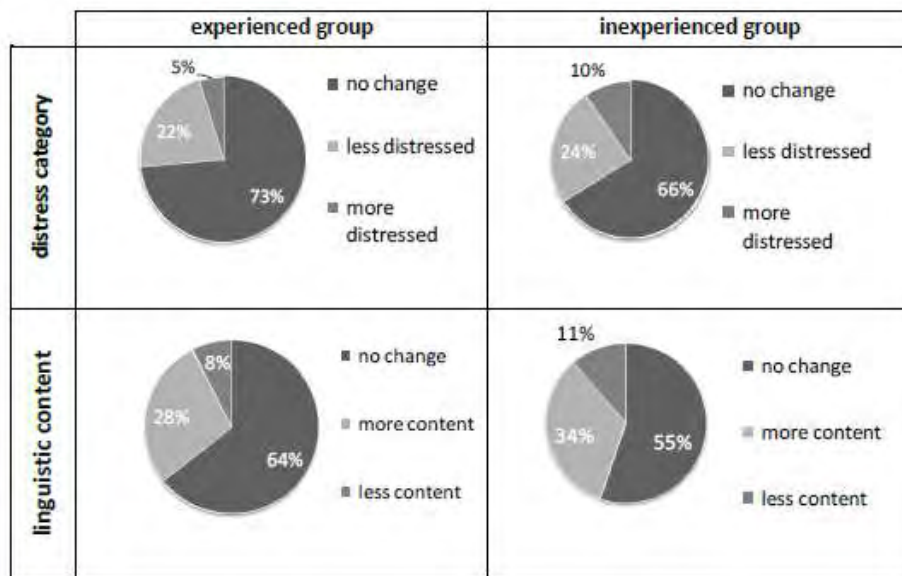
Findings

This section reports only on the experimental trials.¹ ANOVA showed a significant effect with respect to both the distress categorisations and the linguistic content ratings for context and experience, with context exercising the stronger effect. The distress categorisations showed: for context:

$F(1,707) = 6.29$, $p < .012$; and experience: $F(1,707) = 4.58$, $p < .033$). The linguistic content ratings showed: for context: $F(1,714) = 11.94$, $p < .001$; for experience: $F(1,714) = 5.24$, $p < .022$). The interaction between experience and context was not significant for either of the two scales. Owing to the small and unbalanced number of participants in each group; age, sex and native language were not included in the statistical analysis. The effect of order, i.e. whether the extract was first heard with or without context, was not significant.

As illustrated in the pie charts below, where participants altered their responses across the two conditions, extracts were categorised as reflecting lower degrees of distress and rated as having higher degrees of linguistic content when heard with context. This pattern is evident for both groups of forensic phoneticians, although the experienced group was less prone to changing responses across the different conditions.

Table 1: Pie charts showing the change in direction of listener responses from ‘without context’ trials to ‘with context’ trials.



Conclusion

These findings highlight the fact that the attribution of linguistic content to a victim's brief utterance, scream or vocalisation is unlikely to be the product of a consideration of the internal properties of the sound itself. Higher order information – including sequential context and background story – may play a pivotal part in listeners' decisions that a person is attempting to

communicate linguistic information. Similarly, the impression of distress is only partly conveyed by the sound itself; contextual information is again an important factor. The resources used by forensic phoneticians in interpreting difficult recordings may share more with those of lay-listeners (c.f., Jefferson 1978) than they might be aware.

Notes

1. ANOVA did not show a significant effect for changes in categorisation and rating of extracts from participants' control trials, thus the possibility that changes in participants' responses were due to increased exposure of the experiment is discounted.

Acknowledgements

We are very grateful to Paul Foulkes, Tamar Keren-Portnoy, and Eytan Zweig for their assistance with this research, and especially the Forensics Research Group for their participation in the experiment.

References

- Fraser, H. 2003. Issues in transcription: factors affecting the reliability of transcripts as evidence in legal cases. *International Journal Of Speech Language And The Law*, 10(2)..
- Hirson, A. and Howard, D.M. 1994. Spectrographic analysis of a cockpit voice recorder tape. *Forensic Linguistics*, 1(1): 59–69.
- International Association of Forensic Phonetics and Acoustics. Code of Practice – section 9. Accessed March 30 2009 at <<http://www.iafpa.net/code.htm>>.
- Jefferson, G. 1978. What's in a 'nyem'? *Sociology*, 12(1), 135-139.
- Rose, P. 2009. Evaluation of disputed utterance evidence in the matter of David Bain's retrial. Report prepared for court case, 23 February 2009. Accessed 25/03/10 from << <http://forensic-voice-comparison.net/publications.html> >>.

Intonational strategies for backchanneling in Italian Map Task dialogues

Michelina Savino

Department of Psychology, University of Bari, Italy
IfL-Phonetics, University of Cologne, Germany

Abstract

This paper investigates intonational strategies for backchanneling in Bari Italian task-oriented dialogues. Results show that whilst giving feedback, listeners indicate their intention to take the floor or not by using a falling or rising intonation contour respectively. However, this general tendency can be reversed if added paralinguistic meanings are conveyed.

Keywords: Backchannels, Prosody, Italian language, Map Task dialogues

Introduction

The crucial role played by “acknowledge tokens” or “continuers” (Schegloff 1982, Jefferson 1983) in verbal communication is widely recognised. These include lexical and non-lexical tokens (like ‘yes’, ‘yeah’, ‘mm’, ‘uh uh’, ect) which can be used for backchanneling, i.e. for signalling that the listener is attending to the speaker and prompting her/him to go on. Past research work on a number of languages has preliminarily ascertained that continuers are generally characterised by a rising contour as opposed, for example, to yes-reply tokens which normally have a falling F0 shape (for example Jurafky et al 1998, Caspers 2000, Cerrato and D’Imperio 2003, Benus et al 2007). The aim of this study is to contribute to investigating the role of prosody in the pragmatic interpretation of backchanneling in Italian.

Materials and method

Analysed speech data consist of 5 Bari Italian dialogues elicited with a modified version of the Map Task methodology (Carletta et al. 1997), in that participants were not informed about discrepancies between the two maps (how this feature plays a role in speakers’ intonational strategies of backchanneling is discussed later). Dialogues were pragmatically analysed using both orthographic transcriptions and audio files. All “acknowledge” and “positive reply” tokens realised with ‘sì’ (yes), ‘mm’, ‘eh’, and ‘okay’ were selected and intonationally analysed with Praat. For the pragmatic annotation of backchannels, an additional criterion with respect to the broad category of “Acknowledge” move proposed in the Map Task coding scheme (Carletta et al. 1997) was adopted, in which turn-taking dynamics was taken into account. Following Jurafky et al. (1998) work (inspired by theoretical assumptions in Jefferson 1983), we distinguished continuers reflecting

“Passive Reciprocity” (henceforth PR), i.e. those acknowledging the other speaker still holds the turn, from continuers signalling “Incipient Speakership” (henceforth IS), i.e. indicating the intention of taking the floor. Decisions on whether an acknowledge token was to be classified as an example of PR or IS were based on whether or not a change of speaker was involved immediately after the token. In order to avoid possible circularity in the pragmatic interpretation of backchannels, our initial analysis relied on this one criterion. This constitutes a sort of “bottom-up” approach, where a finer pragmatic distinction of backchanneling sub-categories can be derived *a posteriori* by combining analysis of prosody and of communicative context. In order to compare the intonational realisations of backchannels with the same token types used as positive replies, we also analysed all target tokens in “Reply-Y” moves, i.e. positive replies to any yes-no questions as realised in “Query-YN”, “Check” and “Align” moves (Grice & Savino 2003).

Results

Continuers marking PR are mainly realised with the lexical token ‘sì’, followed by ‘mm’ and ‘okay’ (Figure 2, left panel). These continuers are intonationally characterised predominantly by a rising pitch, where a wide range of contour choices is available: rise, fall-rise, “stylised rise”, and also high rise. The most widespread F0 pattern is the falling-rising (Figure 2, right panel): in the dialogues, this is typically used by Instruction Followers (IFs) for conveying not only understanding and agreement about the instructions being received, but also for signalling they are compatible with the presence and position of landmarks on her/his map, i.e. for confirming the current status of shared background knowledge. The stylised rise is only found in ‘mm’ tokens, and in such cases they are produced as bisyllabic, as also observed in English backchannels (Ward 2004). The stylised rise is realised as a pitch step-up from the first to the second syllable, where the pitch excursion is not wide. Examples of these 3 contour types in ‘mm’ are shown in Figure 1. A number of high rising contours are also found, typically used by IFs for conveying her/his understanding and agreement that the end of a (sub)set of instructions for completing a (sub)task is approaching. This contour has already been described in Bari Italian for signalling pre-finality (i.e. marking the antepenultimate item in a sequence) and phonologically analysed with H*H-H% (Savino 2001;2004). Even though backchannels marking PR are generally characterised by a high pitch, we found a large number of these continuers as realised with a falling (or sometimes flat) F0 contour instead. This pattern is usually encountered in the dialogues where participants have not discovered yet that the two maps are different, and it is typically produced by IFs when receiving instructions for drawing the route

which are not compatible with presence and position of landmarks on her/his own map. Interestingly, even though the use of backchannels normally implies signalling understanding and agreement, in these cases a falling contour seems to convey *disagreement* (or uncertainty) about what is assumed by the interlocutor to be the currently shared background knowledge.

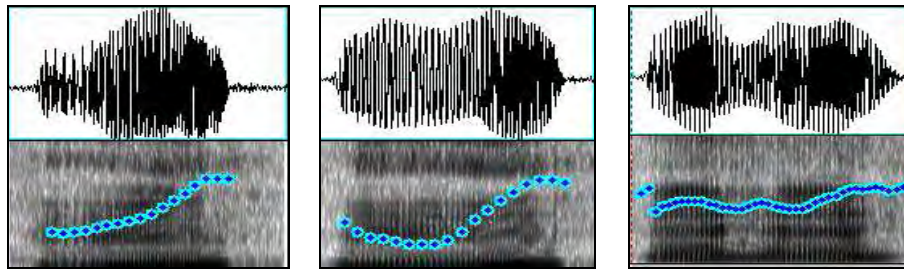


Figure 1. Intonational realisations of ‘mm’ continuers marking PR produced by the same speaker: rising (left), falling-rising (middle), stylised rise (right). In the second case the token is realised as bimoraic, in the third as bysillabic.

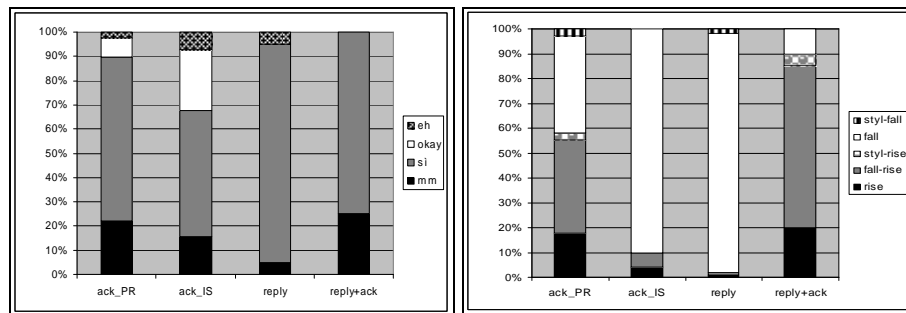


Figure 2. Distribution of token types (left) and F0 contour types (right) as a function of 4 pragmatic classes (acknowledge_Passive-Reciprocity, acknowledge_Incipient-Speakership, yes-reply, yes-reply+acknowledge).

Moreover, among backchannels with falling/flat contour, a wide range of phonetic variation (in terms of pitch slope, amplitude, voice quality, duration) is observed (discussion of these cases is not possible here for lack of space). Acknowledge tokens marking IS (i.e. those involving a change of speaker) have predominantly a falling contour, or a “stylised fall” (Figure 2, right panel). Therefore, the intention of taking the floor or not in backchanneling is conveyed intonationally by using a falling vs a rising contour. In general, acknowledge tokens marking IS are not distinguishable from the corresponding tokens used in yes-replies (Figure 2, right panel), as they both share the same F0 contour. However, especially in positive

answers, there was a great deal of phonetic variation conveying added paralinguistic meanings like surprise, involvement or uncertainty. There was also a number of cases involving a “mixed” pragmatic strategy for conveying yes-answering and backchanneling simultaneously. Here, tokens produced for replying to yes-no questions are intonationally realised with the same contours typically used in continuers marking PR, namely fall-rise, rise or stylised rise (Figure 2, right panel).

Conclusions

This paper presents an investigation on intonational strategies for backchanneling in Bari Italian task-oriented dialogues. Results show that whilst giving feedback, listeners indicate their intention to take the floor or not by using a falling or rising intonation contour. However, this general tendency can be reversed if added paralinguistic meanings are conveyed.

References

- Benus, S., Gravano, A., Hirschberg, J. 2007. The prosody of backchannels in American English. In Proc. of ICPhS 2007, 1065-10-68, Saarbruecken, Germany.
- Carletta, J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G., Anderson, A. 1997. The Reliability of a Dialogue Structure Coding Scheme. *Computational Linguistics* 23 (1), 13-32.
- Caspers, J. 2000. Melodic characteristics of backchannels in Dutch Map Task dialogues. In Proc. of ICSLP 2000, Beijing, China.
- Cerrato, L., D’Imperio, M. 2003. Duration and tonal characteristics of short expressions in Italian. In Proc. of ICPhS 2003, 1213-1216, Barcelona, Spain.
- Grice, M., Savino, M. 2003. Map Tasks in Italian: asking questions about given, accessible and new information. *Catalan Journal of Linguistics* 2, 153-180.
- Jefferson, G. 1983. Notes on Systematic Deployment of the Acknowledgment Tokens “yeah” and “mm hm”. TILL, Tilburg University, The Netherlands.
- Jurafsky, D., Shriberg E., Fox, B., Curl, T. 1998. Lexical, Prosodic, and Syntactic Cues for Dialogue Acts. In Proc. ACL/COLING 98 workshop on Discourse Relations and Discourse Markers, 114-120.
- Savino, M. 2001. Non-finality and pre-finality in Bari Italian intonation: a preliminary account. In Proc. of Eurospeech 2001, 939-942, Aalborg, Denmark.
- Savino, M. 2004. Intonational cues to discourse structure in a variety of Italian. In Gilles, P., Peters, J. (eds.), *Regional variation in intonation*, 145-159, Tuebingen: Niemeyer.
- Schegloff, E. 1982. Discourse as an interactional achievement: some uses of ‘uh huh’ and other things that come between sentences. In Tannen, D. (ed.), *Analysing Discourse*, 71-93. Washington DC, Georgetown Univ. Press.
- Ward, N. 2004. Pragmatic functions of prosodic features in non-lexical utterances. In Proc. of Speech Prosody 2004, Nara, Japan.

Improving intelligibility of synthesized speech in noise with emphasized prosody

Sunil R. Shukla

Department Electrical and Computer Engineering, Georgia Institute of Technology, USA

Abstract

The performance of current high quality concatenative text-to-speech (TTS) systems is limited under noisy environments. This paper investigates whether or not the intelligibility of synthesized speech in noise can be improved by emphasizing the prosody. Additionally, the paper presents a method that can effectively emphasize the prosody of units in existing TTS databases. The circular linear prediction (CLP) model is combined with the constant-pitch transform (CPT) to perform pitch and duration modifications to concatenative TTS units with little impact to the subjective quality. Test utterances are generated using the method and compared to reference utterances synthesized by a high quality TTS engine. The subjective test results demonstrate a preference for emphasized prosody in the majority of the test cases.

Key words: TTS, speech synthesis, linear prediction, prosody, noisy speech.

Background

Spoken dialogue interfaces have become highly common in numerous applications including telecommunications, vehicle navigation, and vehicle command and control. These applications are often used in environments that are prone to significant background noise. It is well-known that noisy environments can significantly impact the intelligibility of synthesized speech (Langner and Black 2005). Even the better performing of the currently available text-to-speech (TTS) products are, on average, 22% less intelligible than human voice under comparable noisy conditions (Venkatagiri 2003). This study explores the hypothesis that a system capable of adding emphasis to key words or syllables in the utterances can improve intelligibility. The current top performing high quality, concatenative TTS systems avoid signal processing for synthesis. Prosody realization, for these systems, is achieved by selecting the closest matching units from a large, prosodically rich database. Hence, emphasized prosody realization would require a significantly larger database. This study presents a novel approach for realizing emphasized prosody, and investigates its efficacy in improving intelligibility of concatenative TTS in “real world” noisy environments.

The CLP/CPT Representation

Although current speech models are effective, and highly useful (especially for speech coding), they are inherently inaccurate due to the quasi-stationary characteristic of speech signals. Current TTS systems are motivated to avoid signal processing for prosody realization due to the potential for audible artefacts caused by these modelling inaccuracies.

Circular linear prediction (CLP) is a highly accurate parametric model for speech (Shukla et al 2002). It improves upon classic LPC techniques (i.e. residual-excited LP) by providing a windowless method for calculating the LP coefficients while retaining the stable filter coefficients property. This reduces the modelling errors attributed to applying a window to each analysis frame and overlap-add during synthesis. In this method, each LP analysis frame is exactly the length of the local pitch period. In this case, each pitch period can be analyzed as an infinitely periodic signal. As in other LP-based methods, the parameters for the segment database are the residual signal, the CLP coefficients, and the pitch period of each frame.

The effectiveness of the model is based on the assumption that the exact pitch period of voiced speech can be detected. This is satisfied by performing the analysis with fractional resolution by oversampling the signal. Note that for unvoiced speech, the assumption becomes irrelevant. In practice, it was found that an oversampling factor of 10 provides sufficient resolution for accurate modelling without audible artefacts (Ertan 2004). Pitch period detection of the oversampled signal is conducted iteratively, by choosing the frame length that maximizes the CLP gain. The complexity of this algorithm is not an issue since analysis for TTS is not conducted in “real-time”.

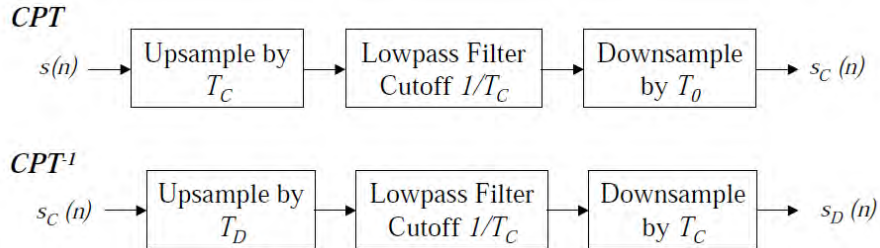


Figure 1: Block diagrams of the constant-pitch transform and the inverse.

The constant-pitch transform (CPT), as shown in figure 1, is a method for modifying the period, T_0 , of every frame of speech, $s(n)$, to a constant length, T_C (Shukla et al 2002). Following CLP analysis, the CPT is applied to every frame to result in a TTS segment database with uniform pitch. Effectively, this technique removes phase discontinuities, which aids in reducing the complexity of segment concatenation and pitch modifications. The

CLP/CPT representation provides a novel technique for realizing parametric modifications to speech segments during synthesis, while having little effect on the synthesis quality.

Prosody Modification and Synthesis

The CLP/CPT representation reduces the relative complexity of prosody modification and segment concatenation, since all units are phase-matched. Modifying the frames segment to a desired pitch, T_D , is achieved by the inverse CPT (see figure 1). Phoneme durations are modified by repeating or deleting frames and parameters. Segment concatenation does not require overlap-add or interpolation, assuming that the database has been normalized. Note that voiced/unvoiced transition regions require additional considerations, which are detailed by Shukla and Barnwell (2007).

Emphasized Prosody

To emphasize the prosody of key words in an utterance, this paper assumes that the locations of key words of the utterance have been marked. In this case, a simple algorithm has been employed for adding emphasis:

- (1) increase the target pitch and duration of the voiced phoneme of the primary syllable by 10-12%;
- (2) lower the pitch of the voiced phonemes of the adjacent syllables by 5-7%.

The above algorithm is applied to the prosody curve generated by an existing high quality, unit-selection synthesis TTS engine.

Subjective Testing

To compare the speech intelligibility improvements, this study implements the CMU Communicator, a spoken dialogue interface that includes a high quality, limited domain, concatenative TTS engine (Rudnicky et al 1999). A subjective test was designed by creating a simulated conversation between the CMU Communicator system and a user in a noisy environment such that the intelligibility of the reference synthesized utterances is degraded.

Method

The CLP/CPT analysis method was applied to the CMU Communicator synthesis database. The test utterances were generated by re-synthesizing the reference utterances with the key words emphasized. Road and traffic noise at highway speeds was recorded and added to the synthesized reference and test utterances. The test is set up such that the subjects are observers to a conversation between a user and the CMU Communicator. The user requests the synthesizer to repeat the response. The subjects then select whether they would prefer the CMU Communicator to repeat the utterance with the same or emphasized prosody. A total of 19 sets of

reference and test utterances were synthesized, and each set were presented to the subjects in random order.

Results and Analysis

Of the 285 total selections made by 15 different subjects, the emphasized responses were preferred 59% of the time, with high statistical significance. The confidence interval for the preference was greater than 99.5%. For some of the cases in which the unmodified speech was preferred, it was noted that the emphasized speech contained audible artefacts. Additionally, in certain cases, the test utterances sounded either overemphasized or not having enough emphasis. These issues are attributed to a suboptimal algorithm for determining the target prosody values and a limitation to the extent of modifications the CLP/CPT method can achieve.

This study concludes that the emphasized prosody is generally preferred in noisy environments. An effective method for realizing emphasized prosody for existing unit-selection TTS synthesizers is presented. The method can be utilized to either modify the prosody during synthesis, or increase the richness of the uni-selection database. During implementation, subjective tests revealed that the range of modifications for the CLP/CPT method is limited to 10% - 15%.

References

- Ertan, A.E., Shukla, S., Barnwell, T. 2002. Circular LPC modeling and constant pitch transform for accurate speech analysis and high quality speech synthesis. ICASSP, 269-272, Orlando, USA.
- Ertan, A.E. 2004. Pitch-synchronous processing of speech. Georgia Institute of Technology: School of Electrical and Computer Engineering, Ph.D. Thesis.
- Langner, B., Black, A. 2005. Improving the understandability of speech synthesis by modeling speech in noise. ICASSP, IV-265-268, Philadelphia, USA.
- Rudnicky, A., Thayer, E., Constantinides, P., Tchou, C., Shern, R., Lenzo, K., Xu W., Oh, A. 1999. Creating natural dialogs in the Carnegie Mellon Communicator system. Proc. of Eurospeech, vol. 4, 1531-1534, Budapest, Hungary.
- Shukla, S., Barnwell, T. 2007. Improving high quality TTS using circular linear prediction and constant pitch transform. ICASSP, 681-684, Honolulu, USA.
- Venkatagiri, H.S. 2003. Segmental intelligibility of four currently used text-to-speech synthesis methods. The Journal of the Acoustical Society of America, 113, 2095-2104.

There is more to chat discourse than written conversation

Cláudia Silva

Centre of Linguistics of University of Porto, Portugal

Abstract

In this paper, we aim at comparing some deviations from standard writing present in Portuguese and English informal chats in order to reflect on the linguistic knowledge that lies beneath these innovative spellings: phonetic, phonological or orthographic.

Key words: chats, spelling, deviations, phonetic and phonological knowledge

Introduction

In this study, we intend to analyse chat discourse in Portuguese and English to verify whether this type of text is merely a conversation which is reproduced in written form or whether it can reveal the speakers' linguistic knowledge.

It is generally agreed that the barriers between oral speech and writing are attenuated in chat discourse, a type of online synchronous communication which is often designated as written conversation (Crystal 2001).

In fact, as communication develops in real time, there is the need to write quickly in order to accompany the rhythm of the interaction. Consequently, chatters use some mechanisms in order to make writing more informal and closer to oral speech. These mechanisms include elision of letters which represent silent sounds, change of grapheme to reproduce the sounds actually heard, repetition of graphemes to suggest the lengthening of vowels or consonants typical of oral speech, absence of capital letters or punctuation marks, which are writing conventions, and use of paralinguistic elements to show emphasis or the attitudes and feelings of chatters. This suggests the relevance of phonetics to chat discourse.

However, it is our belief that chat discourse can reveal more about the speakers' knowledge than their ability to accurately reproduce the pronunciation of words. In Portuguese, at the phonetic level, the deletion of schwa creates sequences of up to five consonants, which highly exceed the two generally accepted as onset for this language (Mateus & Andrade 1998). We could thus say that European Portuguese syllable structure is more complex at the phonetic level, and this could make it closer to syllable structure in English (cf. Hammond 1999), which admits sequences of three or four consonants in word-initial or final position, respectively (even though

it is believed that the excessive consonants are extrasyllabic). In fact, it should be noted that both Portuguese and English are described as possessing a stress-timed rhythm, and, consequently, they both present features like the use of schwas and the compression of unstressed syllables (for overview, see Migliori 2007).

Thus, as chat discourse is said to reproduce oral speech, we could presuppose that the same type of deviations would prevail both in European Portuguese and English, aiming at reproducing the phonetic forms of words.

Therefore, to verify if chat users are really being influenced by their phonetic knowledge, or rather by their phonological and even orthographic knowledge, we are conducting a corpora-based study, which is described below.

Corpus-based study

Sessions of both Portuguese and English chats were recorded. Concerning Portuguese, the chat sessions were taken from “bláblá” (aeiou), whereas English chats belong to Chatterbox (UK). We intended to record sessions from chats used by native speakers in order to collect data that reflected the speakers’ intuitions of their mother tongue.

These two types of chats share similar features although they also show some divergences, according to some of the parameters defined by Herring (2007). As far as the similarities are concerned, they both possess a casual tone, they allow the use of nicknames freely chosen by users, they have a system that posts messages in the order in which they are received and they do not possess an official group topic. Concerning the main differences, the Portuguese chat enables interaction to happen from many to many whereas Chatterbox enables one-to-many interaction. Moreover, although the Portuguese chat studied is composed of text only, the English one enables the use of colourful emoticons. However, we believe that these differences do not significantly interfere with the type of language used in both.

All the words were divided into categories according to the types of deviations to standard writing or even to the respect of the orthographic form. Some examples of the categories include:

representation of the sound: of letters (‘u’ to “you”), of numbers (‘2’ to “to”), of the pronunciation of words (‘xeio’ to “cheio – *full*; ‘wat’ to “what”);

changes at the syllable level: suppression of the onset (‘kes’ to “queres” – *you want*; ‘ovely’ to “lovely”); whole rime (‘ctg’ to “contigo” – *with you*; ‘bldy’ to “bloody”), nucleus (‘tecls to “teclas – *you write*; ‘gd’ to “good”) or coda (‘goto’ to “gosto” – *I like*; ‘wi’ to “with”);

changes in sequences of words, namely junction of words (‘pergunteit’ to “perguntei-te: *I asked you*; ‘infront’ to “in front”).

The results which we are going to present are only preliminary as this is an ongoing study. Consequently, we prefer to refer to general tendencies rather than to quantitative results.

With regard to phonetic knowledge, we can verify that, in Portuguese chats, there are some deviations from standard orthography that may reflect a trend towards the reproduction of pronunciation, such as change of grapheme to imitate the sounds produced, deletion of silent graphemes, junction of words and adding of graphemes in order to imitate the lengthening of sounds typical of oral speech. Likewise, English chatters make use of some strategies in order to (creatively) imitate sound, like the deletion of silent graphemes and sequences of graphemes, the junction of words, and the reproduction of the sound of numbers and letters, a shortening device which is absent in Portuguese chats. These phonetic strategies seem to be more productive in English than in Portuguese, and we consider that this may be due to the fact that English orthography is more opaque than Portuguese spelling. Therefore, seeing that orthographic knowledge does not reflect phonological representations in a transparent way, it may be often disregarded by chat users in English. Moreover, there are more monosyllables in English, which can be easily replaced by a letter or number homophone.

As far as syllable structure is concerned, it is possible to see that there are more changes affecting the constituents of this prosodic unit in Portuguese than in English. The tendency in Portuguese is to delete the whole rime and to maintain the onset represented. If the goal was only to reproduce the phonetic forms of words, the unit would not be the whole rime but only the schwas, which do not correspond to a phonetic realization. Inversely, schwas are not always deleted, especially in less frequent words. We share Bybee's (2001) conviction that frequency (repetition) leads to reduction of form and that it influences the storage and processing of linguistic forms, which can be why schwas are only deleted in more frequent words. Furthermore, there is a tendency towards syllable regularization in some cases, which implies turning a CCV or CVC syllable into the format CV (more unmarked and cross-linguistically frequent according to Blevins 1995). In English chats, however, the suppression of graphemes seem to affect more the ones which do not correspond to an actual pronunciation, thus not affecting the syllable constituents but only involving the words as visual units. Moreover, the fact that there is not a tendency towards the formation of a CV syllable format may suggest that phonological intuitions in English admit a more complex syllable structure as unmarked. In fact, Archangeli (1997) refers to the low position that the restriction "No Complex" possesses in English, within an Optimality Theory approach. We could presuppose, therefore, that intuitions and markedness may be language-particular.

Concerning orthographic knowledge, we can verify that conventions are preserved in different manners. On the one hand, many words are written according to norm. Although many diacritics disappear, others are maintained or are replaced by a letter expressing that function. Some words in Portuguese are influenced by the speakers' orthographic knowledge since they are graphed with a letter that usually represents that sound but not in that context, which suggests that they are not accurately analysing sound but are rather being influenced by spelling.

To sum up, not only is the phonetic knowledge important to chat users, but also their phonological and orthographic knowledge. Consequently, it is our opinion that chats are not merely a form of conversation that happens to be written but that this type of text should rather be studied as a form of access to the speakers' linguistic knowledge.

References

- Archangeli, D. 1997. Optimality Theory: An Introduction to Linguistics in the 1990s. In Archangeli, D., Langendoen, D. (eds.) 1997, *Optimality Theory*, 1-32. Oxford and Massachusetts, Blackwell Publishers.
- Blevins, J. 1995. The Syllable in Phonological Theory. In Goldsmith, J. (ed.) 1995, *The Handbook of Phonological Theory*, 206-244. Cambridge (Mass.)/Oxford, Blackwell.
- Bybee, J. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Crystal, D. 2001. *Language and the Internet*. Cambridge: Cambridge University Press.
- Hammond, M. 1999. *The Phonology of English – A Prosodic Optimality-Theoretic Approach*. Oxford: Oxford University Press.
- Herring, S. 2007. A faceted classification scheme for computer-mediated discourse. *Language@Internet*, 4, article 1. Retrieved June 2, 2009, from <http://www.languageatinternet.de>.
- Mateus, M., Andrade, E. 1998. The syllable structure in European Portuguese. In Mateus, M. 2002. *A Face Exposta da Língua Portuguesa*, 281-297. Lisboa, Casa da Moeda.
- Migliorini, L. 2007. Ritmo Silábico/ Ritmo Acentual: Divergências quanto à definição de Ritmo do Português do Brasil. In *Estudos Lingüísticos XXXVI* vol. 1, 47-51. Retrieved April 5. 2010 from <http://www.gel.org.br/estudoslinguisticos/>.

Visual attention during L1 and L2 sounds perception: an eye-tracking study

Bianca Sisinni¹, Mirko Grimaldi¹, Elisa Tundo¹, Andrea Calabrese²

¹Department of Philology, Linguistics and Literature, University of Salento, Italy

²Department of Linguistics, University of Connecticut, Massachusetts, USA

Abstract

Visual information affects speech perception as demonstrated by the McGurk effect (McGurk & McDonald, 1976): when audio /ba/ is dubbed with a visual /ga/, what is perceived is /da/. This study aims at observing how visual information, intended as articulatory orofacial movements, is processed by eye, i.e., if gaze is related to articulatory information processing. The results indicate that visual attentional resources seem to be higher during multisensory (AV) than unisensory (A; V) presentation. Probably, higher visual attentional resources are needed to integrate inputs coming from different sources. Moreover, audiovisual speech perception seems to be similar across languages (e.g., Chen & Massaro, 2004) and not language-specific (Ghazanfar et al., 2005).

Key words: Audiovisual speech, multisensory integration, native and non-native perception.

Introduction

Visual information affects speech perception as proved by the McGurk effect (McGurk & McDonald, 1976) that shows that when audio /ba/ is dubbed with a face saying /ga/, what is perceived is /da/.

There are conflicting results on the audiovisual (AV) non-native perception. Navarra and Soto Faraco (2007) suggested that the addition of visual information about the articulatory gestures enhanced the ability to discriminate sounds in second language and that the integration of visual-gestural plus auditory information can produce a specific improvement in phonological processing. Hazan et alii (2002) found that non-native speech perception seemed not to be affected by the addition of visual information and concluded that the sensitivity to acoustic and visual cues in L2 acquisition process can be not so evident especially in the early stage of the process and that this sensitivity can largely vary across learners.

Our aim is to observe if gaze behaviour changes according (i) to the speech stimuli presented, i.e., native (L1) and non native (L2) stimuli, and (ii) to the modality of stimuli presentation, i.e., audio (A), video (V) and audio-video (AV), during an almost no-task experiment.

Method

Participants

Nine subjects (1 male) participated (mean age: 19). They all were native speakers of Salento Italian, a variety spoken in the south part of Apulia, and they all attended the first year at Salento University, Faculty of Foreign Language and Literature. In a questionnaire, they reported a formal knowledge of English (mean English education: 12 years) and also declared to have a normal or corrected to normal vision and normal hearing. They were informed about the experimental procedure, approved by the ASL/LE Ethical Committee, and gave their written consent.

Stimuli

Stimuli were produced by two female speakers differing in native language, i.e., Italian (L1) and American (L2). The Italian speaker produced /a/, /e/, /i/, /u/ and the American one produced /æ/, /ɜː/, /ɪ/ and /ʊ/ vowels. Speakers were filmed in a soundproof room by a the camera Camescope Canon MV 960 KIT and their productions were recorded by an SM58 microphone by means of Computer Speech Lab (sampling rate: 22.05 kHz). Video and audio stimuli have been then synchronised. Each stimulus presented the head and the higher part of speakers' shoulders and had a total duration equal to the vowel duration and 0,15 sec before and after the production of the vowel, with the speakers preserving a neutral facial expression.

The L1 and L2 stimuli were randomly presented 15 times in the A, V and AV modalities for a total of 3240 stimuli (8 stimuli x 15 repetitions x 3 modalities x 9 subjects). Subjects were simply asked to watch the computer screen where stimuli were presented. A question slide about what subjects heard (or saw) was randomly presented in order to be sure that they were actually performing the task. The audio of the stimuli was presented throughout loudspeakers at a comfortable level.

Procedure

The experiments were performed in a quiet dark room. Subjects were seated in front of the pc monitor Hp 1702 at a distance of 80 cm. They positioned their head in a headrest and chinrest in order to prevent movements. Right eye movements were monitored by means of ASL Eye-trac 6000, by an infrared camera Canon VC-C50i (sampling rate 120 Hz). A 9-point calibration was used. Spatial error between true gaze position and the computer measurement was less than 1 degree and the precision on a point was better than 0.5 degree. A region of interest (ROI) has been studied in this work, i.e., the mouth. In order to detect this ROI, single frames of the two speakers producing, respectively, the vowels /a/ and /æ/, have been overlapped by a MATLAB script and a ROI compatible with both the two

frames have been considered. This ROI “mouth” was delimited by an semi-horizontal visual angle of $1,9^\circ$ and a semi-vertical visual angle of $1,5^\circ$.

The gaze behaviour has been quantified in terms of (i) the percentage of time spent gazing in the ROI, (ii) the number of fixations (i.e., a fixation last 100ms at least) in the ROI, and (iii) the average duration of fixations in the ROI. These dependent variables have been analysed in series of ANOVA investigating separately the effect of the Condition of stimuli presentation (A, V, AV) and of Language (L1 and L2). Only the significant results ($p < 0,05$) are reported.

Results

When comparing each of the three dependent variables on the basis on the Condition, the results showed that, in both languages, there was a significant difference [$F(2,1619) = 21,320$ to $43,584$ $p < 0,05$]. The post hoc test showed that the dependent variables were higher in AV than in V than in A, as visible in Table 1.

Table 1. The percentage of time spent gazing (% time), the number of fixations (N° fix) and the average duration of fixations (A.d.(s)) in the “mouth” ROI, in each condition (A, V, AV) for each language (L1, L2). Standard deviations are in parenthesis.

	A L1	A L2	V L1	V L2	AV L1	AV L2
% time	2,48 (5,95)	3,56 (8,43)	5,64 (11,27)	5,45 (11,60)	8,02 (12,57)	8,29 (13,15)
N° fix	0,41 (0,97)	0,48 (1,12)	0,84 (1,56)	0,66 (1,34)	1,13 (1,75)	1,01 (1,51)
A.d.(s)	0,037 (0,07)	0,043 (0,08)	0,05 (0,09)	0,05 (0,09)	0,09 (0,11)	0,081 (0,10)

In the comparison of the dependent variables between each language in each single condition, there was a significant difference for the percentage of time fixation in A [$F(1,1079) = 5,898$ $p < 0,05$], since this variable was higher in L2 than in L1, and a significant difference in V [$F(1,1079) = 4,012$ $p < 0,05$] for the number of fixations, higher in L1 than in L2 (Table 1). On the whole, there were no other relevant differences among the dependent variables in the two languages.

Discussion and conclusion

The first most significant finding in our results is the concentration of the gaze on the mouth in the AV condition. The fact that visual attention—i.e., the variables related to gaze behaviour—is higher in AV condition than in V or in A alone, indicate that visual attentional resources are higher during

multisensory presentation than during unisensory presentation. Our results can be considered in line with the results of Tiippana et al. (2004) which found that visual attention is required to combine visual and auditory speech features since when it is disrupted, audiovisual integration is less efficient.

The other significant finding is that gaze behaviour was not different when processing L1 and L2 sounds. Namely, visual information during L2 vowel speech perception seems not to be processed in a different way than the L1 vowel speech perception. Navarra and Soto Faraco (2007) found that vowel discrimination was not different between L1 and L2 speakers when visual information was added. They argued that in order to interpret linguistic speech gestures, the listener does not need to be familiar being the interpretation of visual gestures not language specific. Accordingly, our data seem to provide further evidence for this hypothesis. The lack of differences in visual processing during perception of L1 and L2 sounds could be accounted for by assuming that the underlying process of audiovisual speech perception is similar across languages (e.g., Chen & Massaro, 2004), so that no special visual processing is needed to perceive foreign sounds. It follows that the higher visual attentional resources that are observed during multisensory presentation—another finding of our study—are needed to integrate sensory inputs from different sources as part of a perceptual process, which is not language-specific and is based on a general process of perception also found in non-human primates (Ghazanfar et al., 2005).

References

- Chen, T., Massaro, D.W. 2004. Mandarin speech perception by ear and eye follows a universal principle. *Perception and Psychophysics*. 66, 820–836.
- Ghazanfar, A.A., Maier, J.X., Hoffman, K.L., Logothetis, N.K., 2005. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *The Journal of Neuroscience*. 25(29), 5004–5012
- Hazan, V., Sennema, A., Faulkner, A. 2002. Audiovisual perception in L2 learners. In *Proc. of the Intern. Conference for Spoken Language Processing*, 1685–1688.
- McGurk, H., McDonald, J. 1976. Hearing lips and seeing voices. *Nature*, 265, 746–748.
- Navarra, J., Soto Faraco, S. 2007. Hearing lips in a second language. *Psychological Research*, 71, 4–12.
- Tiippana, K., Andersen, T.S., Sams, M. 2004. Visual attention modulate audiovisual speech perception. *European J. of Cognitive Psychology*. 16, 457–462.

Pause ratio effect in emotion ascribing

Eszter Szabó

Department of Cognitive Science, Budapest University of Technology and Economics (BME), Hungary

Abstract

In this study we investigated how pause length and pause ratio influence listeners in ascribing emotional states to the speaker, if nothing but the length of the pauses changes. We did this by taking emotionally neutral speech samples and manipulating the length of pauses. Our results show that the same speech samples were perceived to be happier, more positive, more heated, and concurrently less sad, less disgusted and less scared when pauses were shorter.

Key words: emotion, speech, pause ratio

Introduction

The relationship between emotions and speech is analyzed from many aspects. Speech signal is often analyzed phonetically. One way to investigate the relationship is through inference studies, which look at the cues that produce certain types of emotion inferences in listeners.

The most commonly measured parameters are intensity, tempo or fundamental frequency (review: Johnstone & Scherer 2000 or Scherer 2003). In most of the research only one-to-two word long, or at longest, one-sentence long speech samples are analyzed. Due to the short sample lengths in these studies, pause length and pause ratio are not relevant. Juslin and Scherer (2005) stressed that although pauses and rhythm seem to be important there remains no standardized measure for this.

However, results of some recent studies suggest that these parameters can vary in different emotional states. Deppermann and Lucius-Hoene (2005) analyzed trauma narratives in TV interviews with conversation analytic methods and noted that speech became deeper, more quiet and powerless, and lost its intonation while people were telling sad stories. They also detected more pauses, even within syntactical units. In our previous study (Szabó 2008) sad and happy emotional states were induced by music and autobiographical recall and half minute long speech samples of the recalls were analyzed. It was found that there were significant differences in average duration of silent pauses and in pause ratio (rate of pauses to the whole speech in percentage) between sad and happy states. Jovicic and his collaborator (2004) designed a Serbian emotional speech database. They asked six actors to repeat single words, short and long sentences and a passage while imitating different emotional states. Total speech segment

duration, total duration of pauses between speech segments and speech/pause ratio were measured. It was found that pause duration in passages is a more discriminative parameter between emotions than speech duration. Speech/pause ratio is a good indicator for speech rhythm, and distinguishes sadness, anger, happiness, fear, and neutral emotional states.

In the present research we asked how pause length and pause ratio influence listeners in ascribing emotional states to the speaker if nothing but the length of the pauses change.

Method

Participants

Fifty university students (16 male and 34 female) participated in the experiment. They had a mean age of 20.12, (SD 1.41). All of them were native speakers of Hungarian. Participants received university credits for their participation.

Material

In the experiment we used five emotionally neutral speech samples of the BEA Hungarian speech database (Gósy, 2008) and a previous research (Szabó, 2008). We recorded women describing themselves, their work and an ordinary day in their lives. Speech samples were taken one by one for approximately one minute each (52-66 seconds). The content was emotionally neutral and didn't contain emotional words (like "*it was a very happy part of my life*"), or laughter. The speech samples used were all monologues and were cut out from longer interviews.

The one-minute long speech samples were taken as "original" samples, and each was modified with the program "Audacity" in four different ways. All of the pauses that were longer than 100 msec were elongated by 18 or 50 percent, or were shortened by 21 or 50 percent. Modifications with 21 and 18 percent were based on our previous results (Szabó, 2008), however in a pretest listeners couldn't perceive any differences so we also increased modifications by 50 percent. Finally we had five variants of each speech sample: the original version, pauses elongated by 18 or 50 percent, and pauses shortened by 21 or 50 percent.

Procedure

Participants were asked to judge five speech samples by filling out a questionnaire. After listening to each speech sample, subjects were to rate how angry, sad, disgusted, happy, surprised, scared, positive, and heated the speaker seemed on a scale of 1-to-6.

In all cases we explained the task to the participants in detail and emphasized that subjects should rate the *whole* speech sample and not decide

too early on. They were instructed to assign a rating in the minute long pauses between the speech samples, only after listening to the full sample.

All listeners participated in the experiment alone, with the speech samples played by a computer via a headset.

Arrangement

In the experiment we had five experimental groups, each with 10 participants. Each participant judged each speaker only once, and all of the speech samples he or she listened to were modified with a different pause ratio (see Table 1.). Participants in group A listened to speech *a* with 50 % abbreviated pauses, speech *b* with 21 % abbreviated pauses, speech *c* in original version etc..

Participants belonging to the same group listened to speech samples in an arrangement called “Balanced Latin Square” (Williams, 1949), where effect of the order of targets is balanced.

Table 1. Speech sample (a–e) and pausing (0,5–1,5) in the different experimental groups.

Group A	Group B	Group C	Group D	Group E
a 0.5	a 1.5	a 1.18	a 1	a 0.79
b 0.79	b 0.5	b 1.5	b 1.18	b 1
c 1	c 0.79	c 0.5	c 1.5	c 1.18
d 1.18	d 1	d 0.79	d 0.5	d 1.5
e 1.5	e 1.18	e 1	e 0.79	e 0.5

Results

We used Spearman’s correlation to measure whether pause length has an effect on listeners’ judgments of how sad, happy, positive etc. they found the speaker. We found that as the length of pauses increased, participants rated the speaker as more sad ($r=0.204$, $p<0.01$), more disgusted ($r=0.174$, $p<0.01$), more scared ($r=0.131$, $p<0.05$). With longer pauses, the speaker was perceived as less happy ($r= -0.179$, $p<0.01$), less positive ($r= -0.153$, $p<0.05$), and emotionally less heated ($r= -0.126$, $p<0.05$).

Discussion

We demonstrated that the length of silent pauses influences listeners in attributing emotional states to the speaker. Results show that with shorter pauses, the same speech samples were perceived to be happier, more

positive, and more heated and concurrently less sad, less disgusted and less scared. Although correlations are significant, it is important to stress that the rate of correlation is not exceptionally high. This is not surprising because the content of all the monologues was emotionally neutral.

Since all of the speech samples were manipulated with all of the pause ratios, and listeners listened to all of the speakers only once, they did not have the opportunity to observe differences, or compare pause ratios between the samples. It might therefore be possible that people have an unconscious conception about the way a speaker might speak in a neutral emotional state, and by perceiving differences in pause ratios, listeners can ascribe emotional states to the speaker. If a pause pattern perception like this does exist, it may be learned and used differently in childhood.

Also, if we know that pauses are relevant for emotion attribution, there are then important implications for speech synthesis.

References

- Deppermann, A., Lucius-Hoene, G. 2005. Trauma erzählen – kommunikative, sprachliche und stimmliche Verfahren der Darstellung traumatischer Erlebnisse. *Psychotherapie und Spezialwissenschaft. Zeitschrift für Qualitative Forschung und klinische Praxis*, 1. 35–73.
- Gósy, M. 2008. Magyar spontán beszéd adatbázis – BEA. *Beszéd kutatás* 2008. 116–128.
- Johnstone, T., Scherer, K. R. 2000. Vocal communication of emotion. In Lewis, M., Haviland-Jones, J. (eds.) 2000, *Handbook of Emotions*, Second Edition, 220–235, New York: Guilford Press.
- Juslin, P.N., Scherer, K.R. 2005. Vocal expression of affect. In JHarrigan, J.A., Rosenthal, R. and Scherer, K.R. (eds.) 2005. *The new handbook of methods in nonverbal behavior research*, 65–135, New York: Oxford University Press.
- Scherer, K.R. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40, 227–256.
- Szabó, E. 2008. A szomorú és a vidám érzelmi állapot megjelenése a beszédben. *Magyar Pszichológiai Szemle*, 63, 4, 651–668.
- Williams, E.J. 1949. Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Scientific Research*, Ser. A 2, 149–168.

Prosodic evidences to sentential negation in Brazilian Portuguese

Lílian Teixeira de Sousa

Department of Linguistics, University of Campinas, Brazil

Abstract

Brazilian Portuguese (BP) permits at least three kinds of sentential negation - Neg1 (Não VP), Neg2 (Não VP Não) and Neg3 (VP Não). The theme of this paper is the nature of this variation. We will try to demonstrate that at least in some contexts Neg2 exhibit properties associated with presupposed denial, reflecting fundamental differences in relation to canonical negation (Neg1). Some contexts show the incompatibility of Neg2 and Neg1. We believe this incompatibility can be explained by presupposition and scope. The aim of this paper is try to show the relation between the negation structures in BP and the pragmatic content.

Key words: Negation, Brazilian Portuguese, pragmatic, prosody

Introduction

Brazilian Portuguese (BP), as (1) shows, permits at least three kinds of negation structures (cf. Alkmim 1999, Schwenter 2005, Cavalcante 2007):

- (1) a. Eu *não/num*¹ gosto de doces
I Neg₁ like of candy-pl
'I don't like candies.'
- b. Eu *não* gosto de doces *não*.
I Neg like of candy-pl Neg₂
- c. Gosto de doces *não*.
Like-1ps of candy-pl Neg₃

The theme of this paper is the nature of variation on sentential negation in BP. We will try to demonstrate that at least in some contexts Neg₂ and Neg₃ in BP exhibit properties associated with presupposed denial, reflecting fundamental differences in relation to canonical negation (1a). By analyzing some contexts in which the incompatibility of Neg₂, Neg₃ and canonical negation (Neg₁) can be explained by presupposition and the notion of focus as theoretical basis, our analysis captures various aspects that support the identificational focus construction hypothesis to Neg₂ and Neg₃.

Another goal of these study is the identification of an analytic model for the interaction between syntax and phonology in the context of the

expression of the informational structure, trying to capture the way syntax influences the prosodic component of the grammar and vice-versa as well as the composition of the meanings in the expression of the informational structure, either universally or parametrically.

On Syntax

*Não*₂ in (1b-c) is different from *não*₁ since it is phonologically autonomous while *não*₁ can surface as a clitic, *num* (cf. Cavalcante 2007, contra Martins 1997 and Fonseca 2004), i.e. *não*₂ is stronger than *não*₁ as only *não*₁ can be reduced:

- (2) *Ele num comprou a casa (não₂/*num)*
 he not.CL bought the house NEG
 ‘‘He has not bought the house’’

As shown follow Neg3 is ungrammatical in embedded sentences, in sentences with subject and in interrogative sentences:

- (3) a. **Eu sei que livro é esse não.*
 I know which book is this not
 b. **O João acha que você deve viajar não*
 the João think th you should travel.INF not
 c. **O João comprou cigarro não*
 the João buy.PAST cigarette not
 d. **Quem você conheceu não?*
 Who you know not

On pragmatics

In BP, *não*₂ necessarily contributes a particular interpretive effect when it occurs. It cannot surface in simple, unmarked declaratives – this is the domain of *não*₁ (4) (cf. Schwenter 2005, Cavalcante 2007).

- (4) a. *A Maria não₁/num vai no teatro* [= simple declarative]
 the Mary not not.CL go in-the theater
 ‘‘Mary is not going to the theatre’’
 b. %*A Maria não₁/num vai no teatro não₂* [≠ simple declarative]
 c. %*A Maria vai no teatro não₂* [≠ simple declarative]

Neg2 occurrence depends nevertheless of common ground, inferable contexts. The following examples show some instances of Neg2.

(5) [Speaker sees interlocutor blowing on some soup]

Não está quente não!

'It's not hot!'

In (5), the speaker sees the interlocutor blowing on the soup and infers that the interlocutor assumes the soup to be hot, but he/she knows that it is not hot and tells it to the interlocutor. Using Neg2, the speaker is correcting a presupposition and introducing new information at the same time.

Now, pay attention to the example below. In this case, the whole sentence bears out the new information, there is no common ground and no presupposition being denied and Neg2 is not suitable.

(6)[speaker going to the school and suddenly remembers she forgot to do her homework]

Nossa! Eu não fiz os exercícios (#não)!

'Oh My! I didn't do the (pl) homework (#not)!'

Interesting data is shown below.

(7)[same situation as (6)]

A: Você fez os exercícios, né?

'You did the homework, right?

B1: Não fiz não!

'No, I didn't!'

B2: (Não) eu não consegui (#não).

'(No.) I couldn't.'

In these data, there are two distinct answers to the question "Você fez os exercícios, né?" (You did the homework, right?). In B1, the presupposition that she did the homework is denied, and Neg2 is ok. In B2, new information is introduced by the sentence "Eu não consegui", in this case only the Neg1 is employed.

Some Results

The methodological analysis is composed by prosodic analysis of sentences produced by four native speakers. We paid special attention to pitch contour. Prosodic structures also reinforce the contrast in the negation possibilities. The entoacional analysis shows that in Neg₂ (1b) nothing receives prominence, the clause-final *não* has the ordinary final sentence intonation pattern. Neg₁ item can receive prominence or not, but if *não* receives prominence the interpretation is informational focus with no presupposition.

Neg₃ is never prominent but, unlike Neg₂, this item can be phonologically separated of the sentence by pause or boundary tone.

Proposal

According Kiss (1998), the identificational focus involves the syntactic projection of the Focus head. In the framework assumed here, this mean that in identificational focus construction, the Focus head acts as a probe and undergoes the Agree operation with the focused expression. Taking those assumptions, we consider that the Neg₃ unlike Neg₂ is not an integrated item, act as a focus head in left periphery attracting the VP to Spec/Foc.

Notes

1. In BP, the pre-verbal negative item can be pronounced by *não* or by the clitic *num*. But the post-verbal negative element can only be the form *não*.

References

- Alkmim, M.G.R. de. 2001. As Negativas Sentenciais no Dialeto Mineiro: uma Abordagem Variacionista, 260 p. Tese de Doutorado, UFMG.
- Careno, M.F., Peter, M.M.T. 1994. Observações sobre o uso da estrutura negativa. *Papia*. V3, n.2. p 98-102.
- Chomsky, N. 2000. Minimalist inquiries: the framework. In *Step by step*, ed. R. Martin, D. Michaels and J. Uriagereka, 89-156. Cambridge, Mass.: MIT Press.
- Furtado Da Cunha, M.A. 1996. Gramaticalização dos Mecanismos de Negação em Natal. IN: Martelotta, Votre & Cezário (org.) *Gramaticalização no Português do Brasil: uma abordagem funcional*. R.J.: Tempo Brasileiro. p. 167-189.
- Jackendoff, R.S. 1972. *Semantic interpretation in generative grammar*. Cambridge, Mass: MIT Press.
- Roncarati, C.N. 1997. A negação no Português falado. In: Macedo, A.T. et al. (Orgs.). *Variação e Discurso*. RJ: Tempo Brasileiro. p.65-102.
- Schwegler, A. 1983. Predicate Negation and word-order change – A problem of multiple causation. *Lingua* 61: s97-334.
- Schwenter, S.A. 2005. The pragmatics of negation in Brazilian Portuguese. *Lingua* 115. p. 1427-1456.

Focus prominence and tonal alignment in Athenian and Cypriot Greek

Charalambos Themistocleous, Stelios Kyriacou
Phonetics Laboratory, University of Athens, Greece

Abstract

This study examines the effects of focus prominence production on tonal alignment in Athenian Greek and Cypriot Greek. Information focus has been elicited by using a Wh-Question in utterances with varying number of syllable constituents—from twelve syllables to eighteen—following the constituent under focus prominence. The results indicate significant effects of speech variety on the alignment of tonal targets; utterance length on the other hand showed poor effects on tonal alignment.

Key words: focus, intonation, Athenian Greek, Cypriot Greek.

Introduction

This study examines the effects of focus prominence on tonal alignment, in utterances with varying length, in Athenian Greek (henceforth AG) and Cypriot Greek (henceforth CG). Focus is manifested in speech as a post-lexical prominence realized intonationally by the exploitation of nuclear pitch accents and phrase accents (c.f. Botinis, Bannert, Tatham, 2000). Importantly, the exact alignment of tonal targets comprising the nuclear pitch accents plays a significant role in the phonetic realization of nuclear pitch accents (c.f. Bruce, 1977). Two main questions are addressed: (a) what is the effect of utterance length on the alignment of nuclear pitch accents, and (b) how do the two varieties of Greek differ in the tonal alignment of nuclear pitch accents.

Methodology

Information focus on a prepositional phrase constituent functioning as a verbal compliment has been elicited by the use of Wh-Questions, specific for each language variety. Specifically, the constituent under focus prominence was produced in three different contexts which varied in the number of syllables following the constituent under focus prominence (see Table 1). The materials were produced by ten educated speakers in total, one male and four female speakers for each variety. AG speakers were brought up in Athens and CG speakers in the main urban areas of Cyprus. All speakers were in their early twenties. All ten speakers produced focus prominence in 3 sentence carriers x 5 repetitions, a total of 150 utterances.

Table 1. The Cypriot Greek material with varying number of syllables following the constituent with focus prominence.

Utterances	Syllables
a. H Nína μιλά [με την Νάνα]F δυνατά. [i 'nina mi'la me tin 'nana ðina'ta] “Nina speaks loudly with Nana”.	[12] Twelve
b. H Nína μιλά [με την Νάνα]F δυνατά στην αυλή. [i 'nina mi'la me tin 'nana ðina'ta stin a'vli] “Nina speaks loudly in the backyard with Nana”.	[15] Fifteen
c. H Nína μιλά [με την Νάνα]F δυνατά στην αυλή του Μάνου. [i 'nina mi'la me tin 'nana ðina'ta stin a'vli tu 'manu] “Nina speaks loudly in the backyard of Manos with Nana”.	[18] Eighteen

The utterances were recorded using *Praat*, with 22050 Hz sampling frequency. The test words were manually segmented and labelled, by using simultaneous inspections of waveforms and wide-band spectrograms following the segmentation criteria of Peterson & Lehiste (1960). The following measurements were obtained from the data (see Figure 1): the distance of local minimum (L) and local maximum (H) (a) from the onset of the first syllable of the word “nana” {n}, (b) from the beginning of the nucleus of the first syllable of the word /'nana/ {v} and (c) from the right edge of the word /'nana/ {we}. The distance of turning point (TP) from the {we} was also measured. A two way ANOVA (language variety x utterance category) was performed for each variable.

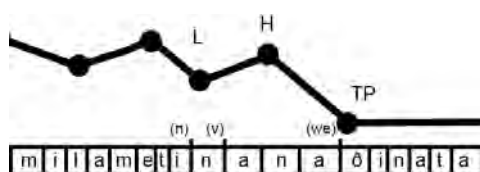


Figure 1. Stylized contour of the utterance /'mila me ti 'nana ðina'ta/ and its associated segmental string. The dots represent tonal targets while the solid lines the interpolated pitch contours. L stands for local minimum, H for local maximum, TP for turning point, {n} for the onset of the syllable (na), {v} for nucleus beginning and {we} for word /'nana/ right edge.

Results

AG and CG productions show apparent similarities in the realization of focus prominence (see Figure 2). A rise and a fall are associated with the stressed syllable of the lexical constituent in focus. The local minimum of the F0 contour aligns at the onset of the stressed syllable, while the local maximum of the F0 contour aligns at the onset of the post-stressed syllable

(see Figure 2). The turning point (TP) aligns at the coda of the lexical constituent in focus prominence.

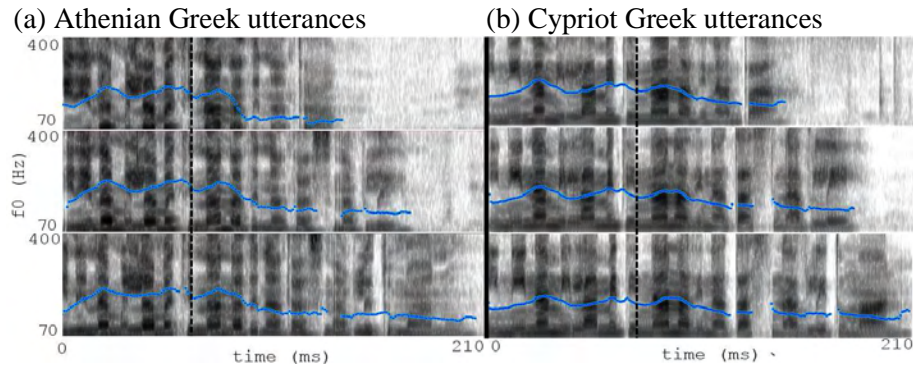


Figure 2. Panel (a) shows the productions of a female AG speaker and Panel (b) shows the productions of a female CG speaker for the utterances a-c. The dashed line indicates the onset of the constituent in focus ['nana].

Language variety however led to different productions in tonal alignment (see Figure 3). ANOVA tests for the effects of variety and utterance type on the distance between L from the onset of the stressed syllable [na] were significant for variety ($F(1,144) = 10.286$, $p < .05$, $r = .36$) and for utterance category ($F(2,144) = 3.157$, $p < .05$, $r = .2$). Also, the distance between H from the onset of [na] was significant for variety ($F(1,144) = 15.567$, $p < .0001$, $r = .39$) and utterance category ($F(2,144) = 6.268$, $p < .05$, $r = .3$). Language variety affected significantly the distance between L ($F(1,144) = 25.622$, $p < .0001$, $r = .5$) and H ($F(1,144) = 28.489$, $p < .0001$, $r = .5$) from the beginning of syllable nucleus {v} and the distance between TP from {we} ($F(1,144) = 28.489$, $p < .0001$, $r = .5$). All other effects were non-significant.

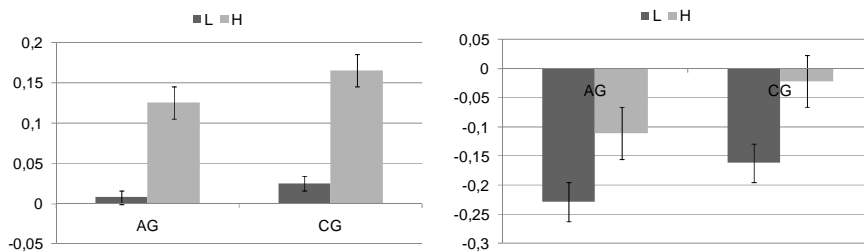


Figure 3. The mean distance in seconds of local minimum (L) and local maximum (H) from the onset of the first syllable of the word “nana” (panel A) and from the nucleus of the first syllable of the word “nana” (panel B) is shown for each variety. Error bars show SE.

Discussion

This study examined the phonetic realization of nuclear pitch accents signalling information focus on a prepositional phrase constituent functioning as a verbal compliment. Importantly, both language varieties show apparent similarities in the realization of the nuclear pitch accent; a rise and a fall are associated with the stressed syllable of the lexical constituent in focus. The local minimum of the F0 contour aligns at the onset of the stressed syllable, while the local maximum of the F0 contour aligns at the onset of the post-stressed syllable. A turning point following local maximum aligns at the coda of the final syllable of the lexical constituent in focus prominence and functions as the left anchoring point of a level declining contour that terminates at the utterance's right edge (c.f. Baltazani, Jun, 1999, Botinis et al. 2000).

The results show significant effects of language variety on the alignment of tonal targets of nuclear pitch accent, indicating systematic differences in the phonemic and phonetic realization of nuclear pitch accents in the two language varieties. Both language varieties show structural consistency in the alignment of tonal targets with the segmental string as manifested by the poorly significant effects of utterances' length on tonal alignment. Further research is needed so as to gain specific understanding of the interaction between different focus categories and their phonemic and phonetic manifestation.

Acknowledgements

The present study has been supported by the University of Athens in the framework of the programme "Kapodistrias", project "Text prosody in Greek". The authors would like to thank the participants of the experiment as well as Antonis Botinis, Marios Fourakis and Spyros Armostis for comments and much useful feedback.

References

- Baltazani, M., Jun S. 1999. Topic and focus intonation in Greek. Proceedings of the XIVth International Congress of Phonetic Sciences, vol. 2, 1305-1308.
- Botinis, A., Bannert R., Tatham M. 2000. Contrastive tonal analysis of focus perception in Greek and Swedish. In Botinis, A. (ed.), *Intonation: Analysis, modelling and technology*. Dordrecht: Kluwer Academic Publishers, 97-118.
- Bruce, G. 1977. *Swedish word accents in sentence perspective*. Lund: Gleerup.
- Peterson, G.E., Lehiste, I. 1960. Duration of Syllable Nuclei in English. *The Journal of the Acoustical Society of America*, 32(6), 693-703.

Narratives in Cypriot Greek mono- and bilingual children with SLI

Eleni Theodorou, Kleanthes K. Grohmann
Department of English Studies, University of Cyprus, Cyprus

Abstract

Narrative ability is one of the most promising ways to assess communicative competence in children: Coordination of lexical, morphosyntactic, phonological, and pragmatic elements is needed to produce narrations. Narratives might even serve as a diagnostic tool for Specific Language Impairment, since children with Specific Language Impairment are reported to have difficulties producing (oral) narratives. Different approaches are available for narrative analysis. In the present study, the Renfrew Bus Story Test was used to investigate narrative ability in mono- and bilingual children aged between 5 and 10 years with typical and impaired language development acquiring Cypriot Greek. The findings suggest that narratives can indeed be a useful tool to identify and assess language-impaired children.

Key words: Cypriot Greek, mono- and bilingual first language acquisition, narrative ability, Renfrew Bus Story Test, Specific Language Impairment.

Introduction

It is well known among linguists and clinicians that coordination of language and a cognitive ability is essential to comprehend and produce narrations. All aspects of language (e.g. lexicon, morphosyntax, phonology, pragmatics) and high function abilities such as attention, processing, and memory need to be coordinated in order to tell an adequate and a tangible story.

Narratives are found across different contexts, cultures, and times, conveying significant information. Narratives form the bases of social interactions contributing to individuals' self-concept (McCabe 1996). It was described by Labov and Waletzky (1967) as the temporal organization and sequencing of past experience into a linguistic device available to speakers. A large body of literature exists which describes the oral narrative skills of young children with typical language development as well as children with language impairment, including Specific Language Impairment (SLI), across different languages. However, there are no such studies on Cypriot Greek, neither with typically developing (TD) mono- or bilingual children nor with impaired children of any kind, in particular SLI.

Justice et al. (2009) suggest that analysis of children's oral narratives represents an informative approach to language assessment. The development of oral narrative is an important achievement, while storytelling plays a significant and critical role in academic success. The

relationship between academic success and narrative ability has been further amplified by data demonstrating that children with literacy difficulties show persistent difficulties over time in narrative production (Feagans and Short 1984).

SLI is a term applied to children who exhibit a significant deficit in language ability yet display normal hearing, age-appropriate scores on tests of non-verbal intelligence, and no obvious signs of neurological damage. SLI causes difficulties in one or more aspects of language: word finding, phonology, morphology, syntax, semantics, and pragmatics (Leonard 1998), which is why SLI is characterized by lack of homogeneity. SLI has been estimated to affect approximately 7% of children, with boys more affected than girls (Tomblin et al. 1997).

Regarding the relation between SLI and narrative ability, previous research, especially in the area of language pathology, shows that children with SLI retell and produce narratives that are less competent and immature with respect to length, lexical diversity, sentence complexity, and content (Gillam and Pearson 2004). Moreover, details that make the story more complete, such as cohesive devices or markers are missing in children with SLI (Leonard 1998). Narrative ability is considered a strong predicting factor for academic success for children with SLI since proficiency in storytelling contributes toward academic performance (Crais & Lorch 1994, Bishop and Edmundson 1987), particularly the acquisition of literacy (Paul et al. 1996). Another important indication for the relation between narratives and SLI refers narrative skill as an area of oral language in which delay is less likely to resolve over time (Girolametto et al. 2001).

With regard to the analysis of narratives, researchers have differentiated between micro- and macro-structural analyses (Hughes, McGillivray, and Schmidek 1997): Micro-structures concern semantic and syntactic complexity of narratives, while macro-structures focus on coherence and organization of the narrative.

Current practice in the field of speech therapy employs narratives in assessment.

Method

Participants

In the present study, the Renfrew Bus Story Test (BST) was used to investigate narrative ability in three groups of children. Ten monolingual children with SLI acquiring Cypriot Greek (aged 5;3–9;3), 4 bilingual children with SLI with Cypriot Greek as one of the two languages (7;6–9;3), and, as the control group, 12 monolingual TD children acquiring Cypriot Greek (5;4–9;11) participated in the study. TD children were recruited randomly from public pre-primary schools in Limassol; children with SLI

were recruited from speech and language therapists in public and in private settings, both upon written parental consent. Participants were assessed on a large testing battery, including the Developmental Verbal IQ Test, the Peabody Picture Vocabulary Test, and the Phonological Phonetic Test, among others. Children showed no evidence of hearing loss, obvious neurological lesion, social problems, or psychiatric syndromes, and they had a non-verbal IQ score higher than 80 measured by the Raven Coloured Progressive Matrices.

Procedures

The BST is a screening test of verbal expression that examines story retelling with picture support. It can demonstrate difficulties with verbal expression, as well as phonological, semantic, grammatical, and sequencing problems. The experimenter told each child a short story about a red bus, while the child looked through a book of pictures illustrating the story. The child then retold it, using the pictures as prompts. The narrations were recorded, transcribed, and evaluated with respect to three measures the BST suggests: the amount of original information included (Inf. Total in Table 1 below), the number of subordinate clauses (Sub. Cl.) and the mean sentence length of the longest five sentences (A5LS).

Results

The table below displays the mean scores for the three measures that were evaluated. The scores in all three measures are found to be different between the two experimental groups (moSLI and biSLI) and the control group (TD).

	TD (n=12)	moSLI (n=10)	biSLI (n=4)
Mean Age	6;4y	6;3y	8;8y
Inf. Total (mean)	33.4	24.1	25.5
Sub. Cl. (mean)	4.7	2.0	2.25
A5LS (mean)	9.2	5.9	7.15

Looking through the mean scores, monolingual and bilingual children with SLI (moSLI and biSLI, respectively) performed worse in all three measures than the TD children did in Cypriot Greek. Looking into individual scores, which are not displayed here for space reasons, differences among individuals in the same group are noticed. However, these differences cannot underestimate the lower score of SLI children, mono- and bilingual.

Discussion

This study confirmed previous research findings regarding narrative ability in mono- and bilingual children with SLI concerning use of complex syntax and mean length of utterances. Our preliminary findings indicate that the BST and narrative assessment can generally provide clinicians with enough information regarding language skills. In addition, information such as narrative abilities may be useful for assessing and monitoring language development within the language intervention program.

References

- Bishop, D.V.M., Edmundson, A. 1987. Language-impaired four year olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders*, 52, 156-173.
- Crais, E., Lorch, N. 1994. Oral narratives in school aged children. *Topics in Language Disorders*, 14, 13-28.
- Dunn, L., Dunn, L. 1981. Peabody Picture Vocabulary Test. Circle Pines, MN, American Guidance Service.
- Feagans, L., Short, E.J. 1984. Developmental differences in the comprehension and production of narratives by reading disabled and normally achieving children. *Child Development*, 55, 1727-1736.
- Gillam, R.B., Pearson, N. 2004. Test of narrative language. Austin, TX: Pro-Ed.
- Girolametto, L., Wiigs, M., Smyth, R., Weitzman, E., Pearce, P.S. 2001. Children with a history of expressive language delay: Outcomes at 5 years of age. *American Journal of Speech-Language Pathology*, 10, 358-369.
- Hughes, D., McGillivray, L., Schmidek, M. 1997. Guide to narrative language procedures for assessment. Eau Claire, WI, Thinking Publications.
- Justice, L.M., Bowles, R.P., Kaderavek, J.N., Ukrainetz, T.A., Eisenberg, S.L., Gillam, R.B. 2006. The index of narrative microstructure: A clinical tool for analysing school-age children's narrative performances. *American Journal of Speech and Language Pathology*, 15, 155-191.
- Leonard, L. 1998. Children with Specific Language Impairment. Cambridge, MA, MIT Press.
- McCabe, A. 1996. Chameleon readers: Teaching children to appreciate all kinds of good stories. New York, NY, McGraw-Hill.
- Paul, R., Hernandez, R., Taylor, L., Johnson, K. 1996. Narrative development in late talkers: early school age. *Journal of Speech and Hearing Research* 39, 1295-1303.
- Tomblin, J., Records, N., Buckwalter, P., Zhang, X., Smith, E., O'Brien, M. 1997. Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language and Hearing research*, 40, 1245-60.

English communication skills for prospective English teachers in Japan

Yuichi Todaka

Department of Intercultural Studies, Miyazaki Municipal University, Japan

Abstract

The objectives of this study are twofold: (1) to promote necessary English oral skills; and (2) to provide prospective English teachers with a necessary roadmap towards achieving competency in English oral communication skills. Prospective English teachers at 11 Japanese universities participated, and we tested our text materials in terms of the degree of motivation, and of the progress on their English oral communication skills. Based upon the obtained results, our text materials helped our participants improve their motivation towards English study and their English oral communication skills. Various modifications that have to be made to the text materials for our next project were also found, however.

Key words: listening, speaking, motivation, self-assessment checklist

Introduction

Todaka (2009) and Todaka and Misono (2009) reviewed the major articles on second language acquisition, English education, and English phonetics published both in Japan and in overseas during the last two years to pinpoint the theoretical constructs, techniques and methods that are common among the articles reviewed. Based on their findings, various exercises focusing on those constructs were developed and customized to help our students develop necessary English oral communication skills.

In addition, another significant factor considered in this project is the issue of motivation. Dornyei and Csizer (2002, 2006) promulgated the notion of L2 Motivational Self System based on their findings. Its core constructs are (1) ideal self, which is defined as the attributes someone would ideally like to possess, (2) out-to self, attributes that one believes one need to have, and (3) L2 learning experience.

Thus, Todaka (2009) formulated a questionnaire assessing Japanese EFL learners' concept of their Ideal L2 Self and motivated learning behavior based on Dornyei and Csizer' (2006) study. However, the number of participants in his study was limited so that only tentative conclusions could have been made. Thus, we utilized the same questionnaire in the present project to reconfirm if the same or similar tendencies could be found.

In addition, the ideal self concept described above needs to come as part of an imagery component and a repertoire of appropriate plans, scripts, and self-regulatory instructional avenues, which are methodological in nature.

We therefore focused our attention on the significance of long-term commitment to L2 learning, as the amounts of time EFL learners engage in L2 learning in class are limited, so that EFL learners need training to become autonomous learners.

Methodology

345 prospective English teachers participated in the present study. 12 to 15 sessions, dependant of the curriculum of each university, for this project were conducted once a week from April through July in 2009. Each session lasted 90 minutes.

A self-assessment checklist was created based upon essential phonetic features, and a questionnaire was administered twice, during the outset of the class and at the end of the class to examine how well the participants had understood each item. In addition, we also formulated a questionnaire aiming to assess factors in L2 learning. The objective of this questionnaire was to evaluate our participants' concept of 'Ideal L2 Self' and their motivated behaviour.

Furthermore, we had the participants take the same TOEIC listening test at the outset and at the end of the semester to examine their improvements in their English listening skill. In addition, we selected 25 out of 345 subjects to assess their improvements in their English speaking skill. The speaking tasks were formulated based upon a TOEFL speaking test and the assessment scale was developed based upon UCLA oral proficiency test for non-native TAs (Celce-Murcia et al, 1996: 404).

Results and discussion

The constructs of instrumentality, language anxiety, classroom anxiety, language learning attitude, attitudes towards L2 community, Out-to L2 self, and ideal L2 self were supported by our participants. Furthermore, the lack of motivated learning behaviour and linguistic self-confidence were also found. However, the participants indicated that they had experienced varying degrees of anxiety in L2 communication, and that they did not have confidence in their linguistic abilities. Furthermore, they also indicated that they did not consider themselves to be diligent language learners.

Second, the results of our checklist criteria were examined. Scores of all the checklist criteria improved, which in turn indicates significant improvements in the understanding of the text materials. Nonetheless, the sections of sentence stress and of English intonation in our textbook materials were found to need further modifications.

Third, we examined the TOEIC test scores taken at the outset and at the end of the training sessions. 345 students took both the pre-and post TOEIC

tests. Based upon an ANOVA analysis, the difference in scores between the pre-and the post TOEIC tests was found to be significant at the 99% confidence level ($p < .0001$, $F = 24.43$). However, the same test was utilized twice; thus, the obtained results might have been different if different tests were administered.

Finally, we examined the results of the speaking performance ratings of 25 randomly selected subjects, and the subjects had to answer the questions on the following four tasks: (1) personal preferences; (2) choices; (3) school matters; and (4) academic matters.

It became apparent that Task 4, which required the subjects to understand the given listening material and to summarize what had been described in the listening material, was found to be most difficult. Although practice exercises were tailored to promote our participants' oral communication skills, the number of carefully planned and customized exercises had not been enough.

Having examined all the data, we can say that (1) enjoyment derived from language learning in the classroom and learning how to study on their own by exploiting the available L2 learning opportunities are the keys to helping Japanese college students relieve their learning anxiety and to helping them realize the significance of making efforts and of having strong persistence in L2 learning to realize their ideal L2 self-images; (2) Japanese prospective English teachers are interested in improving their oral communication skills; (3) our participants listening skill significantly improved based on TOEIC listening test scores; and (4) significant improvements in our participants' speaking skill were found. However, the following factors were found to be important for reconsiderations for our next project: (1) some of the exercises in the textbook were not suitable for our participants in terms of the degree of difficulty and usefulness. This was especially true of the exercises in the sections of sentence stress and of intonation; (2) different TOEIC tests need to be utilized to truly understand the efficacy of our pedagogical approach to improving our participants' English listening skill; (3) more spoken samples need to be examined to understand the efficacy of our pedagogical approach to improving our participants' speaking skill; and (4) much more care needs to be taken for recordings, as some of the recordings by American and Australian English speakers misread their readings.

Conclusion

This is the second year of our three-year longitudinal study of helping prospective English teachers improve their English oral communication skills. We created various exercises to justify one of the aims; that is, developing oral communicative competency, mentioned above. The results we obtained seem promising in the sense that our participants' listening and

speaking skills significantly improved and that our checklist criteria helped our participants understand the essential constructs indispensable to improving their oral communication skills. Nonetheless, various modifications to our original text material, based upon the participants' responses, have to be developed for our next project.

We believe that our understanding of developing effective L2 oral communication skills will open up new avenues for empirical inquiry in this domain as we test the limits of our pedagogy, and stretch beyond further boundaries.

References

- Celce-Murcia, M., Brington, D., Goodwin, J. 1996. Teaching pronunciation: A reference book for teachers of English to speakers of other languages. NY: Cambridge University Press.
- Dornyei, Z., Csizer, K. 2002. Some dynamics of language attitudes and motivation: Results of a longitudinal nationwide survey. *Applied Linguistics* 23 (4): 421-462.
- Dornyei, Z., Csizer, K., Nemeth, N. 2006. Motivation, language attitudes and globalization: A hungarian perspective. Clevedon: Multilingual Matters.
- Todaka, Y. 2009. Daini gengo shutokuriron no matome: kokusaihen. *Journal of English Phonetics*, 13: 395-406.
- Todaka, Y., Misono, K. 2009. Daini gengo shutokuriron no matome: kokunaihen.. *Journal of English Phonetics*, 13: 407-418.

Morphological productivity in Maltese verbs

Alina Twist

University of Maryland Center for Advanced Study of Language, USA

Abstract

This paper describes an experiment that was designed to determine the morphological productivity of two possible verb formation strategies in Maltese: root and pattern on the one hand, and suffixation on the other. Native Maltese speakers created novel words in response to nonce stimuli. The stimuli ranged from phonotactically and prosodically acceptable, but non-existent nonce forms to those that contained segments and/or prosodic patterns typically found in English or Italian, but not native Maltese words. The results show that speakers are able to utilize both non-concatenative and concatenative strategies of word formation.

Key words: Maltese, morphological productivity, psycholinguistics, borrowing

Borrowed verbs in Maltese

Most Semitic languages allow the borrowing of verbs only through intermediate forms of nouns or, less commonly, adjectives. From these forms, denominal verbs are formed using standard verbal patterns. While Maltese allows for this type of borrowing, it also provides a different strategy, allowing verbs to take on a suffix (-*ja*) and subsequent declension class instead of forcing them to conform to regular Semitic verbal patterns. Although loan verbs of both types, root-extracted and suffixed, occur in Maltese, the way in which foreign verbs have been integrated into Maltese has shifted over time. Suffixed forms are considered the most productive current loan verb integration strategy (Mifsud 1995).

Morphological productivity is the combination of two related factors (Bauer 2004). Availability determines whether or not a particular process can be used to form new words. Profitability refers to the extent to which the process is actively used. The suffixation strategy is so prevalent in modern Maltese that Hoberman and Aronoff (2003) were prompted to speculate that root and pattern morphology is no longer productive in Maltese. They argued that the lack of adherence to templatic constraints in borrowed verbs indicates that non-concatenative morphology is not actively functioning in Maltese, suggesting that the profitability of root and pattern morphology is low or non-existent.

Stolz (2003) classified Maltese as a type of mixed language with two available morphological systems. This approach puts more weight on availability of production than profitability, so that the mere fact that it is possible to use a particular strategy contributes to its productivity, even if it

is seldom used. Under this approach, the fact that Maltese exhibits little or no verb borrowing with root extraction is less important than whether or not this is a potential strategy available to native speakers. The experiment described below was designed to explore the factors driving the use of verb formation strategies among native speakers of Maltese by creating a controlled and artificial borrowing situation.

Experiment

Participants

Forty-nine native Maltese speakers between the ages of 18 and 37 participated in the experiment on the University of Malta campus at Msida. All participants were bilingual in Maltese and English, and two were also native speakers of Italian. Participants received monetary compensation for their participation.

Methods

A total of 160 items were presented to participants in Maltese orthography. Half of the items were real nouns of Maltese, and half were possible but unattested nonce forms. The nonce forms were constructed to resemble real nouns. Both the words and non-words were split into two groups, one representing words of Semitic origin, the other words of English origin. The term *origin* is applied to both words and non-words in this study, though technically the non-words have no origin at all. The classification of non-words was determined by a number of structural properties that influenced their construction, including consonant identity and distribution, vocalic sequences, and sonority sequencing.

Within each stimulus origin group, the items were divided equally into two sets: more acceptable and less acceptable. These divisions were based on the general prosodic structure and segmental inventory of each item and verified by native speakers.

Table 1. Examples of test item categories.

	Real Words		Non-words
Semitic, more acceptable	<i>toqba</i>	“a hole”	xesna
Semitic, less acceptable	<i>felliek</i>	“bad steering”	paffien
English, more acceptable	<i>spid</i>	“speed”	klid
English, less acceptable	<i>drill</i>	“drill”	braff

Participants were instructed to respond verbally to each item, naming a verb associated with the item. For the non-word items, participants were

instructed to create a novel verb. Responses were transcribed and coded for lexicality and structural features.

Results

The first step in data analysis was to establish that participants were able to perform the task accurately with real word stimuli. A 2-level analysis of variance (ANOVA) showed no significant difference in the number of acceptable responses to stimuli of different origins ($F_1 < 1$; $F_2 < 1$). In addition, results indicate that nouns of Semitic origin prompted regularly formed Semitic verbs, and English origin nouns prompted suffixed verbs.

Turning to the non-word items, another ANOVA was conducted, adding item acceptability as an additional 2-level factor. The interaction of word origin and acceptability was not significant in by-subjects ($F_1(1, 34) = 1.22$, $p < .05$) or by-items ($F_2(1, 56) = 1.63$, $p > .05$) analyses. The main effect of stimulus origin was significant in both by-subjects ($F_1(1, 34) = 92.31$, $p < .01$) and by-items ($F_2(1, 56) = 26.34$, $p < .01$) analyses. The main effect of item acceptability was significant in the by-subjects analysis ($F_1(1, 34) = 13.36$, $p < .01$) but not in the by-items analysis ($F_2(1, 56) = 2.58$, $p > .05$). Further analysis revealed that the effect of item acceptability differed according to the stimulus origin. For non-words modelled after Semitic words, the effect of item acceptability was not significant in by-subjects ($F_1(1, 34) = 2.10$, $p > .05$) or by-items ($F_2 < 1$) analyses. However, for English origin non-words, both by-subjects ($F_1(1, 35) = 32.94$, $p < .01$) and by-items ($F_2(1, 34) = 7.38$, $p < .02$) analyses yielded significant results. Less acceptable items elicited more *-ja* suffixation than more acceptable items.

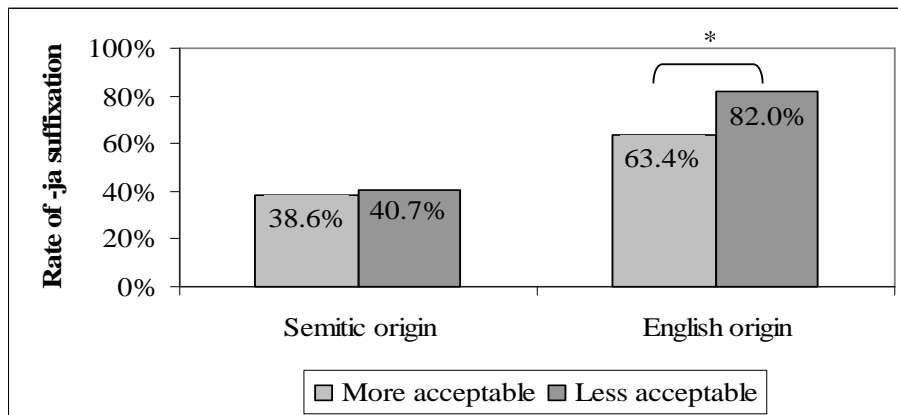


Figure 1. Rate of *-ja* suffixation is highest in less-acceptable verbs of English origin.

Discussion

These data show the availability of root and pattern morphology as a productive verb formation strategy in Maltese, in direct opposition to Hoberman and Aronoff's (2003) claim to the contrary. In some instances speakers chose this strategy even in response to real word stimuli with established loan verbs formed with concatenative suffixation. Responses to the nonce stimuli indicate that root and pattern morphology is a profitable strategy, at least in an experimental context devoid of influences beyond word structure, such as semantics or social context.

Although root and pattern morphology is available to Maltese speakers, a second verb formation strategy, *-ja* suffixation, is also productive. In order to decide which strategy to use to form new words, speakers must consider structural factors such as prosodic shape and phonological segments. However, these factors alone are not sufficient to explain all of the variation.

Acknowledgements

This material is based upon work supported by National Science Foundation Doctoral Dissertation Improvement Grant #0449757. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.


I am grateful to the faculty of the Institute of Linguistics at the University of Malta, Adam Ussishkin, and the anonymous participants in the study for their assistance.

References

- Bauer, L. 2004. A glossary of morphology. Edinburgh: Edinburgh University Press.
- Hoberman, R., Aronoff, M. 2003. The verbal morphology of Maltese. In Shimron, J. (Ed.), *Language processing and acquisition in languages of Semitic, root-based, morphology*, 61-78. Amsterdam: John Benjamins.
- Mifsud, M. 1995. *Loan Verbs in Maltese*. Leiden: E.J. Brill.
- Stolz, T. 2003. Not quite the right mixture: Chamorro and Malti as candidates for the status of mixed language. In Matras, Y., P. Bakker (Eds.), *The mixed language debate*, 271-316. New York: Mouton de Gruyter.

Index of authors

Agapitou, P., 1
Alexandris, Ch. 5
Alonso, P., 9
Anastassiou, F., 53
Andreou, G., 1, 53
Blochowiak, J. 13
Boudraa, M., 37
Calabrese, A., 169
Castelain, Th., 13
Castro, L., 17
Chaida, A. 21, 25, 125
Cichocki, W., 37
d'Apolito, S., 29
de Camargo, Z.A., 33, 105
Droua-Hamdani, G., 37
Eleni Theodorou, E., 185
Epistithiou, A., 97, 113
Fourakis, M., 41, 125
Fox, R.A., 45
Freitas, M., 17
French, P. 149
Fujio, M., 49
Galantomos, I., 1, 53
Grimaldi, M., 169
Grohmann, K.K., 73, 81, 185
Hatzivasiliou, Th., 57
Hatzopoulou, M., 57
Hatzopoulou, M., 89
Henriques, I., 61
Jacewicz, E., 45
Javier, F., 65
Johnston, L.H., 77
Kainada, K., 69
Kambanaros, M., 73
Karlsson, H. 41
Karpava, S., 81
Kliper, R., 85
Kujala, T., 145
Kyriacou, S., 181
Lazarou, E., 89
Lehtinen, M., 93
Leivada, E., 113
Leivada, E., 97
Li, B., 101
Madureira, S. 33, 105
Martin, Ph., 109
Martins, P., 133
Mavroudi, P., 113
Mavroudi, P., 97
Moeschler, J., 13
Moraes, J., 17
Müller, D. 117
Nikolaenkova, O., 121
Nirgianaki, E., 25, 125
Näätänen, R., 145
Oh, S., 129
Oliveira, C., 133
Panaretou, E., 25
Paradowski, M.B., 137
Pavlou, N., 141
Peltola, M.S., 145
Portuguese, S., 85
Roberts, L. 149
Salonen, L., 145
Savino, M., 157
Selouani, S.A., 37
Serridge, B., 17
Shriberg, L., 41
Shukla, S.R., 161
Silva, C., 165
Sisinni, B., 169
Szabó, E., 173
Tamminen, H., 145
Teixeira de Sousa, L. 177
Teixeira, A., 133
Themistocleous, Ch., 181
Theodorou, E., 73
Tilkens, Ch., 41
Toivonen, H., 145
Tundo, E., 169
Vaizman, Y., 85
Wayland, R., 101
Weinshall, D., 85
Vsevolod Kapatsinski, V., 77
Yuichi Todaka, Y., 189
Zhang, C., 101
Twist, E., 193



ExLing 2010

Ebook ISSN: 2529-1092

Ebook ISBN: 978-960-466-157-2