

International Speech Communication Association

Proceedings of ISCA Tutorial and Research Workshop

on

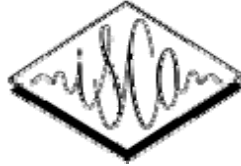
Experimental Linguistics

28-30 August 2006, Athens, Greece.

Edited by Antonis Botinis



University of Athens



International Speech Communication Association

Proceedings of ISCA Tutorial and Research Workshop on

Experimental Linguistics

28-30 August 2006, Athens, Greece.

Edited by Antonis Botinis



University of Athens

ISCA Experimental Linguistics

Antonis Botinis, editor

ISBN: 960-6608-57-3

Printed in Athens, Greece, by the University of Athens

5 Stadiou Ave

10562 Athens, Greece

© copyright 2006

ISCA and the University of Athens

Foreword

This volume includes the proceedings of the ISCA (International Speech Communication Association) Tutorial and Research Workshop on Experimental Linguistics held in Athens, Greece, 28-30 August 2006, under the auspices of the University of Athens, Greece, the University of Skövde, Sweden, and the University of Wisconsin-Madison, USA.

Our call had a significant appeal to the international scientific community and papers were submitted from different parts of the world. Thus, in accordance with the aims of the Workshop, the volume includes a variety of papers covering theoretical as well as experimental and interdisciplinary approaches. In addition to the merits of each and every paper, the ultimate objectives of the Workshop are to bring together scientists from different areas and boost interdisciplinary research and cooperation. A main issue for discussion is the use of experimental methodologies in order to produce linguistic knowledge. Another issue is the effect of each linguistic factor as well as the interactions between factors in relation to linguistic structures. A further issue is the relation between sound and meaning as a function of linguistic categories and structures.

We do not expect to have answers to many questions to be raised. However, we wish to approach language from different perspectives and discuss disciplinary methodologies and goals in relation to linguistic theory and linguistic knowledge at length. And, if more questions are to come along, that would be another cycle for renewed thoughts in the study of language. ISCA workshops and conferences are excellent opportunities for established as well as for young scientists to present their work at an international forum. Pursuing linguistic knowledge, we may face old problems in new ways and new problems in old ways. This cycle is necessarily based on the constant influx of young scientists who, equipped with experimental methodologies and laboratory expertise, may extend linguistic research beyond its current limits.

Our thanks to the contributors and the invited speakers Artemis Alexiadou, David Caplan, George Clements, Diane Kewley-Port, Anna Papafragou, Joseph Perkell and Niels Schiller as well as Anthi Chaida for the administration of the Workshop and the University of Athens for the publication of the present proceedings volume.

The Organising committee

Aikaterini Bakakou-Orphanou

Antonis Botinis

Christoforos Charalambakis

Tutorial papers

On the properties of VSO and VOS orders in Greek and Italian: a study on the syntax-information structure interface	1
Artemis Alexiadou	
Neurolinguistics	9
David N. Caplan	
Quantal phonetics and distinctive features	17
George N. Clements and Rachid Ridouane	
A new trainable trajectory formation system for facial animation	25
Oxana Govokhina, Gérard Bailly, Gaspard Breton and Paul Bagshaw	
Topics in speech perception	33
Diane Kewley-Port	
Spatial representations in language and thought	41
Anna Papafragou	
Sensorimotor control of speech production: models and data	45
Joseph S. Perkell	
Phonological encoding in speech production	53
Niels O. Schiller	

Research papers

Experiments in investigating sound symbolism and onomatopoeia	61
Åsa Abelin	
Prosodic emphasis versus word order in Greek instructive texts	65
Christina Alexandris and Stavroula-Evita Fotinea	
Gradiance and parametric variation	69
Theodora Alexopoulou and Frank Keller	
Stress and accent: acoustic correlates of metrical prominence in Catalan	73
Lluïsa Astruc and Pilar Prieto	
Word etymology in monolingual and bilingual dictionaries: lexicographers' versus EFL learners' perspectives	77
Zahra Awad	
Characteristics of pre-nuclear pitch accents in statements and yes-no questions in Greek	81
Mary Baltazani	
Effects of VV-sequence deletion across word boundaries in Spanish	85
Irene Barberia	

Production and perception of Greek vowels in normal and cerebral palsy speech	89
Antonis Botinis, Marios Fourakis, John W. Hawks, Ioanna Orfanidou	
Pre-glottal vowels in Shanghai Chinese	93
Yiya Chen	
Distinctive feature enhancement: a review	97
George N. Clements and Rachid Ridouane	
Where the wine is velvet: Verbo-pictorial metaphors in written advertising	101
Rosa Lúdia Coimbra, Helena Margarida Vaz Duarte and Lurdes de Castro Moutinho	
Measuring synchronization among speakers reading together	105
Fred Cummins	
Formal features and intonation in Persian speakers' English interlanguage wh-questions	109
Laya Heidari Darani	
The effect of semantic distance in the picture-word interference task	113
Simon De Deyne, Sven Van Lommel and Gert Storms	
Melodic contours of yes/no questions in Brazilian Portuguese	117
João Antônio de Moraes	
The phonology and phonetics of prenuclear and nuclear accents in French	121
Mariapaola D'Imperio, Roxane Bertrand, Albert Di Cristo and Cristel Portes	
The influence of second language learning on speech production by Greek/English bilinguals	125
Niki-Pagona Efstathopoulou	
Aspectual composition in Modern Greek	129
Maria Flouraki	
Investigating interfaces: an experimental approach to focus in Sicilian Italian	133
Raffaella Folli and Elinor Payne	
A corpus based analysis of English, Swedish, Polish, and Russian prepositions	137
Barbara Gawronska, Olga Nikolaenkova and Björn Erlendsson	
Evaluation of a virtual speech cuer	141
Guillaume Gibert, Gérard Bailly and Frédéric Elisei	
Broad vs. narrow focus in Greek	145
Stella Gryllia	

Incremental interpretation and discourse complexity Jana Häussler and Markus Bader	149
Dynamic auditory representations and phonetic processing: The case of virtual diphthongs Ewa Jacewicz, Robert Allen Fox and Lawrence L. Feth	153
Syntactic abilities in Williams Syndrome: How intact is ‘intact’? Victoria Joffe and Spyridoula Varlokosta	157
Experimental investigations on implicatures: a window into the semantics/pragmatics interface Napoleon Katsos	161
On learnability and naturalness as constraints on phonological grammar Hahn Koo and Jennifer Cole	165
Prosody and punctuation in <i>The Stranger</i> by Albert Camus Mari Lehtinen	169
An acoustic study on the paralinguistic prosody in the politeness talk in Taiwan Mandarin Hsin-Yi Lin, Kwock-Ping John Tse and Janice Fon	173
Analysis of stop consonant production in European Portuguese Marisa Lobo Lousada and Luis M. T. Jesus	177
Towards multilingual articulatory feature recognition with Support Vector Machines Jan Macek, Anja Geumann and Julie Carson-Berndsen	181
Prosody, syntax, macrosyntax Philippe Martin	185
Effects of structural prominence on anaphora: The case of relative clauses Eleni Miltsakaki and Paschalia Patsala	189
Speaker based segmentation on broadcast news-on the use of ISI technique S. Ouamour, M. Guerti and H. Sayoud	193
The residence in the country of the target language and its influence to the writings of Greek learners of French Zafeiroula Papadopoulou	197
The Acquisition of Epistemic Modality Anna Papafragou and Ozge Isik Ozturk	201
Towards empirical dimensions for the classification of aphasic performance	205

Athanassios Protopapas, Spyridoula Varlokosta, Alexandra Economou and Maria Kakavoulia	
Analysis of intonation in news presentation on television	209
Emma Rodero	
Templates from syntax to morphology: affix ordering in Qafar	213
Pierre Rucart	
Processing causal and diagnostic uses of <i>so</i>	217
Sharmaine Seneviratne	
Acoustics of speech and environmental Sounds	221
Susana M. Capitão Silva, Luis M. T. Jesus and Mário A. L. Alves	
Romanian palatalized consonants: A perceptual study	225
Laura Spinu	
Formal expressive indiscernibility underlying a prosodic deformation model	229
Ioana Suciu, Ioannis Kanellos and Thierry Moudenc	
What is said and what is implicated: A study with reference to communication in English and Russian	233
Anna Sysoeva	
Animacy effects on discourse prominence in Greek complex NPs	237
Stella Tsaklidou and Eleni Miltsakaki	
Formality and informality in electronic communication	241
Edmund Turney, Carmen Pérez Sabater, Begoña Montero Fleta	
All roads lead to advertising: Use of proverbs in slogans	245
Helena Margarida Vaz Duarte, Rosa Lúcia Coimbra and Lurdes de Castro Moutinho	
Perception of complex coda clusters and the role of the SSP	249
Irene Vogel and Robin Aronow-Meredith	
Factors influencing ratios of filled pauses at clause boundaries in Japanese	253
Michiko Watanabe, Keikichi Hirose, Yasuharu Den, Shusaku Miwa and Nobuaki Minematsu	
Assessing aspectual asymmetries in human language processing	257
Foong Ha Yap, Stella Wing Man Kwan, Emily Sze Man Yiu, Patrick Chun Kau Chu and Stella Fat Wong	
Supplement	
Toward a rich phonology	261
Robert Port	

On the properties of VSO and VOS orders in Greek and Italian: a study on the syntax-information structure interface

Artemis Alexiadou
Institute of English Linguistics, University of Stuttgart, Germany

Abstract

This paper deals with word order variation that relates to patterns of information structure. The empirical focus of the paper is a comparison of Italian and Greek word order patterns. The paper will address, however, issues of word order typology in general. The main line of argumentation is one according to which syntax directly reflects information structure, and variation is explained on the basis different movement parameters.

Introduction

The patterns in (1-3) are all found in Greek and Italian, two pro-drop languages known to allow several word order permutations.

1. SV(O)
2. VS (O)
3. VOS

In the recent literature a lot of attention has been devoted to the fact that these patterns reflect topic/focus relations. A possible description of the above patterns in terms of information structure is as follows:

- 1'. SV(O): subject is taken to be old information, i.e. it is a topic.
- 2'. VS(O): in the unmarked case all information is new.
- 3'. VOS: the subject is new information.

The patterns in 2 and 3 can be further subdivided into a number of sub-types depending on intonation, which will be discussed here in detail.

The existence of these patterns raises three questions: (i) how are properties of information structure reflected in syntax? (ii) are all these orders and interpretations equally available in both languages? If not, what explains this variation? (iii) how are the VSO and VOS patterns related to e.g. Celtic VSO and Malagasy VOS? Questions (ii) and (iii) are important for the comparative syntax perspective. First, as we will see immediately, Italian is

rather different from Greek. Second, intuitively there is a difference between e.g. Irish VSO and Malagasy VOS and the patterns discussed here. Importantly, in Greek and Italian the above are only some of a number of possible patterns and not the obligatory patterns, as is the case in the other languages and our syntactic theory should be able to explain this.

Here I focus on the VS(O) patterns and I briefly discuss VOS patterns.

Patterns

Some terminology

As the patterns to be discussed related to notions such as focus and topic, following Zubizarreta (1998: 10) and many others, I distinguish between contrastive focus and new information focus. There a number of criteria that can be used to tease them apart. Contrastive focus contrasts the subset of a set of alternatives with the complement subset. In this case, a background assertion is introduced by a statement. New information focus simply conveys new information. In this case, the background is introduced by wh-questions.

Different types of VS(O) patterns

The following patterns can be distinguished:

- (i) VS/VSO (with no particular intonation)
 - (ii) V#S and (cl)VS#O with comma intonation; in this case, the S and O are right-dislocated.
 - (iii) VS /VSO: in this case the subject bears contrastive focus and the object in the VSO case is de-accented but in situ (Zubizarreta 1998: 155f, see also Cardinaletti 2001).
- (ii-iii) are equally found in Italian, and Greek, while (i) is restricted/impossible in Italian.

- | | | | |
|-----|----|---|-------|
| (1) | a. | irthe o Janis
came John | Greek |
| | b. | irthe, o Janis
came John | |
| | c. | agorase o Janis tin efimerida
bought John the newspaper | |
| | d. | agorase o JANIS tin efimerida
bought John the newspaper | |
| | e. | tin agorase o Janis, tin efimerida
it bought John, the newspaper | |

- (2) a. ha parlato Gianni *Italian*
 has spoken John
 b. ha parlato, Gianni
 has spoken John
 c. L'ha comprato Maria, il giornale
 it bought Mary the newspaper
 b. Ha comprato MARIA, il giornale
 has bought Mary the journal

The position of the subject in VS(O)

The position of the arguments in VS and VSO orders are taken to be low in the IP area, in particular within the vP, as it follows adverbs that mark the vP edge (Alexiadou & Anagnostopoulou 1998, Belletti 1999).

- (3) a. ?Capirá bene Maria *Italian*
 will understand well Maria
 b. *Capirá Maria bene
 will understand Maria well
 (4) an ehi idhi diavasi_j [_{vP} kala [_{vP} o Petros t_j to mathima]] *Greek*
 if has already read well Peter the lesson
 If Peter has already read the lesson well

VS: differences between Italian and Greek

In Italian VS is marginal as an answer to the question ‘What happened?’:

- (5) irthe o Janis
 came John
 (6) e'entrata Beatrice
 is entered Beatrice
 (7) # e'impallidito Berlusconi
 is turned pale Berlusconi

According to Benincá (1988) and Pinto (1997:21), the example in (7) is not felicitous under a wide focus interpretation, but acceptable under a narrow reading on the subject. Such an interpretation is in general possible with VS orders (see also Belletti 1999). For this reason, VS orders are felicitous answers to the question ‘Who came?’:

- (8) irthe o Janis
 came John
 (9) e arrivato Gianni
 is arrived John

Thus we can conclude that Italian VS orders are generally characterized by new information focus on the subject. Only under special conditions can all information be considered new. Greek is not subject to these constraints.

Benincá (1988), Pinto (1997), Belletti (1999), Tortora (2001) and Cardinaletti (to appear) note that definite subjects can appear postverbally in Italian, if they satisfy the following two conditions:

- (10) a. the definite description identifies its referent in a unique way
- b. the definite description must bear new information (as the postverbal subject position is normally identified with focus)

Second, verbs that permit inversion with definite subjects in Italian differ in their lexical structure from those that do not permit inversion. In particular, the former contain a locative or temporal argument, which can be overtly or covertly realized, which is located in subject position. In particular, what occupies the preverbal position is a null locational goal argument of the unaccusative verb (Cardinaletti to appear). The aforementioned authors agree that when the locative remains implicit, it is interpreted deictically. Thus a sentence like (6) means that Beatrice arrived/entered here. That inversion is closely related to deixis in Italian is supported by the data in (11-12), from Pinto (1997: 130):

- (11) Da questo porto è partito Marco Polo
from this harbour left Marco Polo
- (12) *Dal porto è partita la nave
from the harbour left the ship

(12) is ungrammatical. According to Pinto, the reason for this ungrammaticality is related to the difference between the demonstrative *questo* 'this' and the determiner *il* 'the'.

V#S

In this pattern the subject is already given information, separated by comma intonation. So as an answer to the question 'What did John do?', we can find the examples in (14) and (15), where especially in Greek the use of the overt subject is like an afterthought:

- (14) efige, o Janis
left John
- (15) ha parlato, Gianni
has spoken John

Arguably the subject is in a right-dislocated position. According to Kayne (1994), Cardinaletti (2001, 2002), see also Georgiafendis (2001), in this case, the subject is generated in the complement of a functional projection whose specifier hosts the whole clause.

(16) [[efige] X° [o Janis]]

VSO

In VSO orders in Greek, all information is new, and the subject is VP internal, as the pattern can function as an answer to the question 'what happened?'

(17) molis espase o Janis tin kristalini lamba
 just broke the-John-NOM the crystal lamp
 'John just broke the crystal lamp'

Italian disallows VSO but (data from Belletti 1999), but allows for VSPP and VSO orders when the subject is a pronoun:

- (18) a. Ha telefonato Maria al giornale
 has phoned Mary to the newspaper
 b. *Ha telefonato Maria il giornale
 has called Mary the newspaper
- (19) a. Di quel cassetto ho io le chiavi
 of which drawer have I the keys
 b. *Di quel cassetto ha Maria le chiavi
 of which drawer has Mary the keys

Why is VSPP possible but VSO impossible? Alexiadou & Anagnostopoulou (2001) argued that intransitivity constraint on inverted orders of the type in (20); this is active in English and French, which do not permit inversion with transitive verbs.

(20) At Spell-Out the vP-VP should not contain more than one argument, at least one DP argument must check Case overtly

(20) can be violated in languages that permit clitic-doubling such as Greek and Spanish. That is VSO orders are permitted in languages that have a doubling configuration (the relationship between V and S is one of doubling). Italian does not have doubling of the type found in Greek, hence both arguments can remain VP-internally only when the second one is a PP, and hence it does not need to check Case. This means that V never checks the case of the subject in Italian. This help us understand why the pronominal subject

fares better. Pronouns target a position which is outside the VP. To the extent that such patterns are possible they indicate overt subject movement to a Case checking position (based on Belletti 1999). This is shown in (21) where the pronominal subject precedes the adverb marking the vP edge:

- (21) Di questo mi informeró io bene
of this myself I will inform better

VSO and (cl)VS#O

Both patterns are possible in Italian and Greek. Here the one pattern contains a clitic, the other not:

- | | | | |
|------|----|------------------------------------|----------------|
| (22) | a. | agorase o JANIS tin efimerida | <i>Greek</i> |
| | | bought John the newspaper | |
| | b. | tin agorase o Janis, tin efimerida | |
| | | it bought John, the newspaper | |
| (23) | a. | Ha comprato Maria, il giornale | <i>Italian</i> |
| | | has bought Mary the journal | |
| | b. | L'ha comprato Maria, il giornale | |
| | | it bought Mary the newspaper | |

Greek permits a further pattern.

- (22) c. tin agorase o Janis tin efimerida
it bought John the newspaper

It will be shown that the two patterns, the one with and the one without the clitic are different. The difference between (22c) and (22b) relates to the difference between clitic-doubling and clitic right dislocation.

The syntax of VOS

VOS

VOS is a possible word order and tends to be associated with new information and contrastive focus. The question here is how can we derive these patterns, and in addition explain the restrictions found with Italian VOS. I will argue that the marginality of VOS can be understood if Italian VOS involves VP internal scrambling.

VOS

In this case the object bears contrastive focus. For Italian, Cardinaletti (2001) argues that subject is right dislocated. Indeed in cases where the object bears contrastive focus the subject has been previously mentioned, and could be analysed as being right dislocated.

CIVOS

In this case cIIVOS belong to the 'known' part of the clause, and the subject receives new information. This is impossible in languages that have right dislocation only. In principle the syntax of CIVOS should not be different from that of VOS, but see Revithiadou & Spyropoulos (2002).

Two word order parameters

Two types of VSO languages

There are two types of VSO languages. Both are characterized by V-movement. But they differ as to whether they make another, non EPP-related vP external specifier available for the subject DP, like non pro-drop languages. This is present in Irish, but not in Greek, Alexiadou & Anagnostopoulou (1998).

Two types of VOS languages

There are two types of VOS languages differentiated by the XP vs. X° movement parameter. The languages discussed here have been all argued to have head movement. According to Pearson (2001), Malagasy lacks head movement and rather makes use of XP movement.

(23) Pearson's generalization

- a. languages with suffixal tense/aspect morphology seem to have Verb movement, if overt.
- b. language with prefixal tense/aspect morphology seem to have XP movement, if overt.

Greek instantiates pattern (a), while Malagasy instantiates pattern (b).

References

- Alexiadou, A. and Anagnostopoulou, E. 1998. Parametrizing Agr: Word Order, Verb-movement and EPP-checking. *Natural Language and Linguistic Theory* 16.3, 491-539.
- Alexiadou, A. and Anagnostopoulou, E. 2001. The subject in situ generalization, and the role of Case in driving computations. *Linguistic Inquiry* 32, 193-231.
- Belletti, A. 1988. The case of unaccusatives. *Linguistic Inquiry* 19, 1-34.
- Belletti, A. 1999. VSO vs. VOS: On the licensing of possible positions for postverbal subjects in Italian and Romance. Paper presented at the workshop on Inversion, May 1998, Amsterdam.
- Belletti, A. 2001. Aspects of the Low IP area. Manuscript, University of Siena.
- Benincà, P. 1988. L'ordine degli elementi della frase e le costruzioni marcate: soggetto postverbale. In L. Renzi (ed.) *Grande grammatica italiana di consultazione*, vol. 1, 115-191. Il Mulino.
- Cardinaletti, A. 1997. Subjects and Clause Structure. In L. Haegeman (ed.) *The New Comparative Syntax*, 33-63. London, Longman.
- Cardinaletti, A. 1999. On Italian Post-verbal Subjects. Ms., University of Venice.
- Cardinaletti, A. 2001. A second thought on emarginazione: destressing vs. right dislocation. In G. Cinque and G. Salvi (eds) *Current studies in Italian Syntax*, 118-135. Amsterdam, Elsevier.
- Cardinaletti, A. 2002. Against optional and null clitics: right dislocation vs. marginalization. *Studia Linguistica* 56, 39-58.
- Cardinaletti, A. To appear. Towards a cartography of subject positions.
- Georgiamentis, M. 2001. On the properties of the VOS order in Greek. *University of Reading Working Papers in Linguistics* 5: 137-154.
- Kayne, R. 1994. *The antisymmetry of syntax*. Cambridge, Mass., MIT Press.
- Pearson, M. 2001. *The clause structure of Malagasy: a minimalist approach*. Ph.D. dissertation, UCLA.
- Pinto, M. 1997. *Licensing and interpretation of inverted subjects in Italian*. Doctoral dissertation, University of Utrecht.
- Revithiadou, A. and Spyropoulos, V. 2002. Trapped within a phrase: effects of syntactic derivation of p-phrasing. Ms. University of the Aegean.
- Tortora, C. 2001. Evidence for a null locative in Italian. In G. Cinque and G. Salvi (eds) *Current studies in Italian Syntax*, 313-326. Amsterdam, Elsevier.
- Zubizarreta, M.L. 1998. *Prosody, Focus and Word Order*. Cambridge, Mass., MIT Press.

Neurolinguistics

David N. Caplan

Department of Neurology, Harvard Medical School, USA

Abstract

Neurolinguistics studies the relation of language processes to the brain. It is well established that the critical brain regions for language include the perisylvian association cortex, lateralized to the left in most right-handed individuals. It is becoming increasingly clear that other brain regions are part of one or more complex systems that support language operations. Evidence regarding the more detailed organization of the brain for specific language operations is accruing rapidly, due to functional neuroimaging, but has not clearly established whether specific language operations are invariantly localized, distributed over large areas, or show individual differences in their neural substrate.

Introduction

“Neurolinguistics” refers to the study of how the brain is organized to support language. It focuses on the neural basis of the largely unconscious normal processes of speaking, understanding spoken language, reading and writing.

Data bearing on language-brain relations come from two sources. The first are correlations of lesions with deficits, using autopsy material, magnetic resonance imaging (MRI), positron emission tomography (PET), direct cortical stimulation, subdural stimulation, and transcranial magnetic stimulation. The logic of the approach is that the damaged areas of the brain are necessary to carry out the operations that are deficient at the time of testing, and undamaged areas of the brain are sufficient to carry out intact operations. The second source of information is to record physiological and vascular responses to language processing in normal individuals, using event related potentials (ERPs), magnetoencephalography (MEG), cellular responses, positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). The logic behind this approach is that differences in the neural variable associated with the comparison of performance on two tasks can be related to the operation that differs in the two tasks. This approach provides evidence regarding the brain areas that sufficient to accomplish the operation under study. Functional neuroimaging studies in patients can reveal brain areas that are sufficient for the accomplishment of an operation that were not active prior to damage to the areas that usually support an operation.

The Gross Functional Neuroanatomy of Language

Beginning in the late nineteenth century, the application of deficit-lesion correlations based on autopsy material to the problem of the regional specialization of the brain for language yielded the important finding that human language requires parts of the association cortex in the lateral portion of one cerebral hemisphere, usually the left in right handed individuals. This cortex surrounds the sylvian fissure and runs from the pars triangularis and opercularis of the inferior frontal gyrus (Brodmann's areas (BA) 45, 44: Broca's area), through the angular and supramarginal gyri (BA 39 and 40) into the superior temporal gyrus (BA22: Wernicke's area) in the dominant hemisphere (Fig 1). For the most part, the connections of these cortical areas are to one another and to dorsolateral prefrontal cortex, lateral inferior temporal cortex, and inferior parietal lobe. These regions have only indirect connections to limbic structures (Geschwind, 1965). These areas consist of many different types of association cortex.

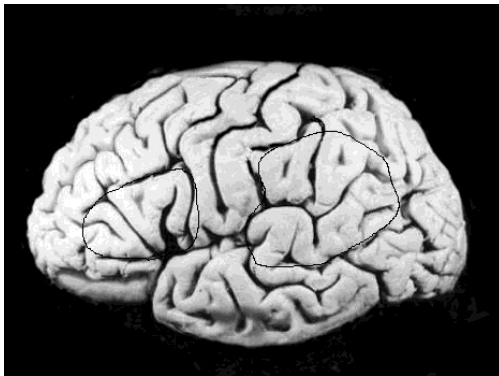


Figure 1. A depiction of the left hemisphere of the brain showing the main language areas.

Data from other sources – deficit-lesion correlations based on ante-mortem neuroimaging, functional neuroimaging – has provided evidence that regions outside the perisylvian association cortex also support language processing. These include the inferior and anterior temporal lobe, the supplementary motor cortex, subcortical nuclei such as the thalamus and striatum, the cingulate gyrus, and the cerebellum. Whether these areas are responsible for the computations of the language processing system or only support cortical areas in which these computations occur remains under study. These areas are connected by white matter tracts, in which lesions can produce language disorders.

The statistics regarding gross hemispheric dominance for language are now quite well established. In about 98% of right-handed individuals, the left hemisphere is dominant. About 60% - 65% of non-right-handed individuals are left-hemisphere dominant; about 15% - 20% are right-hemisphere dominant; and the remainder appear to use both hemispheres for language processing (Goodglass and Quadfasel, 1954). The relationship of dominance for language to handedness suggests a common determination of both, probably in large part genetic (Annett, 1985). The neural basis for lateralization was first suggested by Geschwind and Levitsky (1968), who discovered that part of the language zone (the planum temporale -- a portion of the superior temporal) was larger in the left than in the right hemisphere. Subsequent studies have confirmed this finding, and identified specific cytoarchitecturally defined regions in this posterior language area that show this asymmetry (Geschwind and Galaburda, 1987). Several other asymmetries that may be related to lateralization have been identified although the exact relationship between size and function is not known. The "nondominant" hemisphere is involved in many language operations, such as representing word meanings, and some language operations may be carried out primarily in the right hemisphere (e.g., revising inferences, interpreting non-literal language, and appreciating humor).

In summary, a large number of brain regions are involved in representing and processing language. The most important of the regions used to support the normal production and comprehension of literal propositional language appears to be the dominant perisylvian cortex. Ultimately, all areas interact with one another as well as with other brain areas involved in using the products of language processing to accomplish tasks. In this sense, all these areas are part of a "neural system" for language, but there is evidence, reviewed below, that many of these areas compute specific linguistic representations in particular tasks.

Models of Organization of the Brain for Language Processing

Two general models of the relationship of areas of the brain to components of the language processing system have been developed. Localizationist theories maintain that language processing components are localized in specific parts of the brain. "Holistic" theories maintain that linguistic representations and processes require broad areas of the brain. Five basic models, which capture the set of logically possible relations of brain areas to language processes, can be extracted from these two conceptualizations: invariant localization, variable localization, even distribution, invariant uneven distribution, and variable uneven distribution.

Invariant localization hypothesizes that only a small area of the brain supports a function. Variable localization hypothesizes that different small areas of the brain support a function in different individuals. Distribution hypothesizes that a large region of the brain supports a function. Traditional distributed models (e.g., Lashley, 1950, modelled by Wood, 1978) assumed an even distribution of distributed functions: all parts of the region contributed equally to the function. If a function is evenly distributed throughout a region, there can be no individual variability in its neural basis. If a function is unevenly distributed throughout a region, it may be distributed the same way in everyone (invariant uneven distribution) or differently in different individuals (variable uneven distribution). Other models are extensions of these basic five. Degeneracy is a variant of localization in which more than one structure independently supports a function (Noppeney et al, 2004); degeneracy can either be invariant (the same areas independently support the function in everyone) or variable (different areas independently support the function in different people). Variable localization could be constrained so that a function is localized more often in one area than another.

It is not possible to review all the areas of language whose neurological basis has been studied. I shall review work on comprehension at the lexical and syntactic levels, highlighting new concepts and examining the evidence that supports them.

Lexical Access and Word Meaning

Evidence from normal and impaired human subjects suggests that temporospectral acoustic cues to feature identity appear to be integrated in unimodal auditory association cortex lying along the superior temporal sulcus immediately adjacent to the primary auditory koniocortex (Binder, 2000). Some researchers have suggested that the unconscious, automatic activation of features and phonemes as a stage in word recognition under normal conditions occurs bilaterally, and that the dominant hemisphere is the sole site only of phonemic processing that is associated with controlled processes such as subvocal rehearsal and conscious processes such as explicit phoneme discrimination and identification, making judgments about rhyme, and other similar functions.

Based on functional neuroimaging results, activation of the long term representations of the sound patterns of words is thought to occur in the left superior temporal gyrus. Scott and her colleagues have argued that there is a pathway along this gyrus and the corresponding left superior temporal sulcus such that word recognition occurs in a region anterior and inferior to primary auditory cortex, and that word meanings are activated further along this pathway in anterior inferior temporal lobe bilaterally (Scott and Wise, 2004).

This pathway constitutes the auditory counterpart to the visual “what” pathway in the inferior occipital-temporal lobe.

Speech perception is connected to speech production, especially during language acquisition when imitation is crucial for the development of the child’s sound inventory and lexicon. On the basis of lesions in patients with repetition disorders known as “Conduction aphasia,” the neural substrate for this connection has been thought to consist of the arcuate fibers of the inferior longitudinal fasciculus, which connect auditory association cortex (Wernicke’s area in the posterior part of the superior temporal gyrus) to motor association cortex (Broca’s area in the posterior part of the inferior frontal gyrus). Recent functional neuroimaging studies and neural models have partially confirmed these ideas, providing evidence that integrated perceptual-motor processing of speech sounds and words makes use of a “dorsal” pathway separate from that involved in word recognition (Hickok and Poeppel, 2004).

Traditional neurological models maintained that the meanings of words consist of sets of neural correlates of the physical properties that are associated with a heard word, all converging in the inferior parietal lobe. It is now known that most lesions in the inferior parietal lobe do not affect word meaning and functional neuroimaging studies designed to require word meaning do not tend to activate this region. Evidence is accruing that the associations of words include “retroactivation” of neural patterns back to unimodal motor and sensory association cortex (Damasio, 1989), and that different types of words activate different cortical regions. Verbs are more likely to activate frontal cortex, and nouns temporal cortex for nouns, possibly because verbs refer to actions and nouns refer to static items. A more fine-grained set of distinctions has been made within the class of objects themselves. Both deficits and functional activation studies have suggested that there are unique neural loci for the representation of categories such as tools (frontal association cortex and middle temporal lobe), animals and foods (inferior temporal lobe and superior temporal sulcus), and faces (fusiform gyrus) (see Caramazza and Mahon, 2006, for review). Debate continues as to whether such divisions reflect different co-occurrences of properties of objects within these classes, or possibly innate human capacities to divide the world along these lines. At the same time as these specialization receive support, evidence from patients with semantic dementia and from functional neuroimaging indicates that a critical part of the semantic network that relates word meanings and concepts to one another is located in the anterior inferior temporal lobes.

Syntactic Comprehension

Syntactic structures determine the relationships between words that allow sentences to convey propositional information – information about thematic roles (who is initiating an action, who receiving it, etc.), attribution of modification (which adjectives are assigned to which nouns), scope of quantification, co-reference, and other relations between words. The propositional content of a sentence conveys a great deal of information beyond what is conveyed by words alone, and is crucial to many human intellectual functions. Propositions are the source of much of the information stored in semantic memory. Because propositions can be true or false, they can be used in thinking logically. They serve the purpose of planning actions. They are the basic building blocks of much of what is conveyed in a discourse.

Unlike models of the neural basis for lexical access and lexical semantic processes, a variety of models have been proposed regarding the neural basis for syntactic processing, ranging from localization, though distribution to variable localization.

Evidence supporting these models based on correlating deficits in syntactic comprehension to lesions is limited, both in terms of psycholinguistic and neural observations. Many patients have only been tested on one task, and we have found that there is virtually no consistency of individual patients' performances across tasks, raising questions about whether it is correct to say that a patient who fails on a particular structure in a single task has a parsing deficit. Lesions have usually not been analyzed quantitatively and related to performance using multivariate statistics.

We have just reported the most detailed study of patients with lesions whose syntactic comprehension has been assessed (Caplan et al, in press). We studied forty-two patients with aphasia secondary to left hemisphere strokes and twenty-five control subjects for the ability to assign and interpret three syntactic structures in enactment, sentence-picture matching and grammaticality judgment tasks. We obtained magnetic resonance (MR) and five-deoxyglucose positron emission tomography (FDG PET) data on 31 patients and 12 controls. The percent of selected regions of interest that was lesioned on MR and the mean normalized PET counts per voxel in regions of interest were calculated. In regression analyses, lesion measures in both perisylvian and non-perisylvian regions of interest predicted performance after factors such as age, time since stroke, and total lesion volume had been entered into the equations. Patients who performed at similar levels behaviorally had lesions of very different sizes, and patients with equivalent lesion sizes varied greatly in their level of performance. The data are consistent with a model in which the neural tissue that is responsible for the operations underlying sentence comprehension and syntactic processing is localized in different neural regions, possibly varying in different individuals.

Functional neuroimaging studies have led many researchers to articulate models in which one or another aspect of parsing or interpretation is localized in Broca's area, or in portions of this region, and some researchers have argued that "Universal Grammar," in Chomsky's sense (the innate capacity that underlies the ability to acquire the syntax of natural language) is localized in this region. However, most neuroimaging studies actually show that multiple cortical areas are activated in tasks that involve syntactic processing. Overall, the data are inconsistent with invariant localization, and suggest variation in the localization of the areas that are sufficient to support syntactic processing within the language area across the adult population, with perhaps some constraint on the areas in which processing is localized as a function of how proficient individuals are at assigning syntactic structure and determining the meaning of sentences (Caplan et al, 2003).

Final Notes

Human language is a unique representational system that relates aspects of meaning to many types of forms (e.g., phonemes, lexical items, syntax), each with its own complex structure. Deficit-lesion correlations and neuroimaging studies are beginning to provide data about the neural structures involved in human language. It appears that many areas of the brain are either necessary or sufficient for representing and processing language, the left perisylvian association cortex being the most important such region. How these areas act to support particular language operations is not yet understood. There is evidence for both localization of some functions in specific regions and either multi-focal or distributed involvement of brain areas in others. It may be that some higher-level principles operate in this area, such that content-addressable activation and associative operations are invariantly localized and computational operations are not, but many aspects of these topics remain to be studied with tools of modern cognitive neuroscience.

References

- Annett M. 1985. Left, right, hand and brain: the right shift theory. London: Erlbaum.
- Binder, J. 2000. The new neuroanatomy of speech perception, *Brain* 123: 2371-2372.
- Caplan, D., Hildebrandt, N. and Makris, N. 1996. Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain* 119: 933-949
- Caplan D., Waters G. and Alpert N. 2003. Effects of age and speed of processing on rCBF correlates of syntactic processing in sentence comprehension. *Human Brain Map* 19: 112-131.

- Caplan, D. Waters, G., Kennedy, D. Alpert, A., Makris, N., DeDe, G., Michaud, J., Reddy, A. (in press). A Study of Syntactic Processing in Aphasia II: Neurological Aspects, Brain and Language.
- Caramazza, A. and Mahon, B.Z. 2006. The organisation of conceptual knowledge in the brain: The future's past and some future directions. *Cognitive Neuropsychology*, 23: 13-38
- Damasio, A. 1989. Time-Locked multiregional retroactivation: A Systems-level proposal for the neural substrates of recall and recognition. *Cognition* 33: 25-62.
- Geschwind, N. 1965. Disconnection syndromes in animals and man. *Brain* 88: 237-294, 585-644.
- Geschwind, N. and Galaburda, A.M. 1987. *Cerebral Lateralization: Biological Mechanisms, Associations and Pathology*. Cambridge: MIT Press.
- Geschwind N. and Levitsky W. 1968. Human brain: left-right asymmetries in temporal speech region. *Science*. 12;186-7, 42:428-59, 421-52, 634-54.
- Goodglass, H. and Quadfasel, F.A. 1954. Language laterality in left-handed aphasics. *Brain* 77: 521-548
- Hickok, G. and Poeppel, D. 2004. Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition* 92: 67-99.
- Lashley, K.S. 1950. In search of the engram. *Society of Experimental Biology, Symposium* 4, 454-482.
- Noppeney, U., Friston, K. J. and Price, C. 2004. Degenerate neuronal systems sustaining cognitive functions. *Journal of Anatomy*, 205, 433.
- Scott, S.K. and Wise, R.J.S. 2004. The functional neuroanatomy of prelexical processing of speech. *Cognition* 92:13-45.
- Wernicke C. 1874. The aphasic symptom complex: a psychological study on a neurological basis. Kohn and Weigert, Breslau.
- Wood, C.C. 1978. Variations on a theme by Lashley: Lesion experiments on the neural model of Anderson, Silverstein, Ritz, and Jones. *Psychological Review* 85: 582-591.

Quantal phonetics and distinctive features

George N. Clements¹ and Rachid Ridouane^{1,2}

¹Laboratoire de phonologie et phonétique, Sorbonne Nouvelle, France

²ENST/TSI/CNRS-LTCI, UMR 5141, Paris, France

Abstract

This paper reviews some of the basic premises of Quantal-Enhancement Theory as developed by K.N. Stevens and his colleagues. Quantal theory seeks to explain why some articulatory and acoustic dimensions are favored over others in distinctive feature contrasts across languages. In this paper, after a review of basic concepts, a protocol for quantal feature definitions is proposed and problems in the interpretation of vowel features are discussed.

The quantal basis of distinctive feature

Though most linguists and phoneticians agree that the distinctive features of spoken languages are realized in terms of concrete physical and auditory properties, there is little agreement on exactly how they are defined. According to a tradition launched by Jakobson and his collaborators (for example, Jakobson, Fant and Halle 1952), features are defined mainly in the acoustic (or perhaps auditory) domain. In a second tradition initiated by Chomsky and Halle (1968), features are defined primarily in articulatory terms. After several decades of research, these conflicting approaches have not yet led to any widely-accepted synthesis.

In recent years, a new initiative has emerged within the framework of the Quantal Theory of speech, developed by K.N. Stevens and his colleagues (e.g. Stevens 1989, 2002, 2005). This theory maintains that the universal set of features is not arbitrary, but can be deduced from the interactions between the articulatory parameters of speech and their acoustic effects. The central claim is that there are phonetic regions in which the relationship between an articulatory configuration and its corresponding acoustic output is not linear. Within such regions, small changes along the articulatory dimension have little effect on the acoustic output. It is such regions of acoustic *stability* that define the articulatory inventories used in natural languages. In other words, these regions form the basis for a universal set of distinctive features, each of which corresponds to an articulatory-acoustic coupling within which the auditory system is insensitive to small articulatory movements.

A simple example of an acoustic-articulatory coupling can be found in the parameter of vocal tract constriction. Degrees of constriction can be ordered along an articulatory continuum extending from a large opening (as in

low vowels) to complete closure (as in simple oral stops). In most voiced non-nasal sounds, the passage along a scale of successively greater degrees of constriction gives rise to three relatively stable acoustic regions with separate and well-defined properties. Sounds generated with an unobstructed vocal tract constriction, such as vowels, semivowels, and liquids, are classified as sonorants. A sudden change in the acoustic output occurs when the constriction degree passes the critical threshold for noise production (the Reynolds number, see Catford 1977), giving rise to continuant obstruent sounds (fricatives). A further discontinuity occurs when the vocal tract reaches complete closure, corresponding to the configuration for noncontinuant obstruents (oral stops). These relations are shown for voiced sounds in Figure 1, where the three stable regions correspond to the three plateaux.

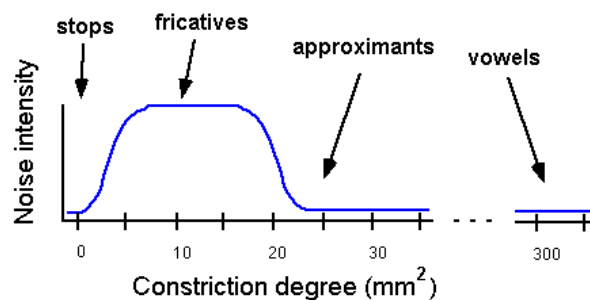


Figure 1. Continuous changes along the articulatory parameter “constriction degree” define three stable acoustic regions in voiced sounds.

In *voiceless* sounds, the falling slope in this figure shifts some distance to the right (to around 90 mm^2), and the region between the shifted and unshifted slopes (about 20 to 90 mm^2), corresponding to voiceless noise production, defines the class of approximant sounds (liquids, high semivowels, etc.), whose acoustic realization is noiseless when they are voiced but noisy when they are voiceless (Catford 1977).

Languages prefer to exploit articulations that correspond to each of the four stable regions defined in this way. These regions give rise to the features which define the major classes of speech sounds, as shown in Table 1. (The feature [+vocalic], used here to define vowels and semivowels, is equivalent to the classical feature [-consonantal]).

Table 1. The four major classes of speech sounds.

Vowels	stops	fricatives	approximants	vocoids
[continuant]	no	yes	yes	yes
[sonorant]	no	no	yes	yes
[vocalic]	no	no	yes/no	yes

These features are commonly used across languages. All known languages have stops and vowels, and most have fricatives and approximants as well.

A protocol for quantal feature definitions

A feature definition, if it is quantal, must identify an articulatory continuum associated with one or more acoustic discontinuities, and must specify the range within this continuum that corresponds to relatively stable regions in the related acoustic output. The range is the articulatory definition of the feature, and the associated output is the acoustic definition. A feature definition must also identify the stable region in terms specific enough to distinguish it from other regions, yet general enough to apply to all articulations within this region, allowing for observed crosslinguistic variation. It must effectively distinguish segments bearing this feature (e.g. /t^h/) from otherwise similar segments that do not (e.g. /t/). Finally, it must identify the classes of sounds in which the definition holds. This will usually be the class in which the feature is at least potentially distinctive.

As an example, consider a proposed definition of the feature [+consonantal], which distinguishes true consonants from vocoids (vowels, semivowels) and laryngeals: "The defining acoustic attribute for this feature is an abrupt discontinuity in the acoustic signal, usually across a range of frequencies. The defining articulatory attribute is the formation of a constriction in the oral cavity that is sufficiently narrow to create such an acoustic discontinuity. This description applies to both [-sonorant] and [+sonorant] consonants." (Stevens 2004, B79). This definition conforms to the protocol suggested above. It identifies an articulatory continuum (constriction degree) and identifies the range within this continuum ("narrow constriction") associated with a discontinuity -- specifically, a rapid drop in F1 frequency and amplitude, as further explained and illustrated in the extended discussion of this feature in Stevens (1998), 244-6. It will be noted that this definition is specific enough to distinguish [+consonantal] sounds from other sounds, yet general enough to apply to a variety of realizations, for example by the lips, tongue blade, or tongue body. Finally, the definition is general enough to hold across all consonants, including both obstruents and sonorants.

There are two general families of quantal feature definitions: a) *contextual* definitions, in which the acoustic or auditory cue to the feature can only be detected when the sound bearing the feature occurs in an appropriate context, and b) *intrinsic* definitions, in which the cue can be found within the segment itself. The feature [+consonantal] just discussed is an example of a contextual definition, as the discontinuity in question occurs when the consonantal sound occurs in the context of a nonconsonantal sound (as in *may* or *aim*). A strong advantage of contextual cues is that they are linked to "landmarks" in the signal often associated with phoneme boundaries. Such "landmarks" are perceptually salient and tend to be rich in feature cues. It is suggested that they may facilitate speech segmentation and lexical access (e.g. Huffman 1990, Stevens 2000, 2002).

An example of an intrinsic definition is the following, as proposed for the feature [±back] which distinguishes front vowels from central and back vowels. "[During the] forward displacement of the tongue body, the second natural frequency F2 of the vocal tract passes through the second natural frequency of the airway below the glottis, which we will call F2T, for the second tracheal resonance. For adult speakers, F2T has been observed to be in the range 1400 to 1600 Hz, and it is relatively constant for a given speaker. As F2 passes through F2T, the spectrum prominence corresponding to F2 often does not move smoothly, but exhibits a discontinuity or abrupt jump in frequency. Thus there tends to be a range of values of F2 within 100 Hz or so where the frequency of the spectrum prominence is unstable. It appears that languages avoid vowels with F2 in or close to this region... and put the F2 their vowels on one side or the other of this region; corresponding to [+back] vowels for lower F2 and [-back] vowels for higher F2. Thus there appears to be a dividing line between two regions with a low F2 for a backed tongue body position and a high F2 for a fronted tongue body position." (Stevens 2004, B79-80)

This definition again follows the protocol. The articulatory continuum is tongue fronting (assuming a central position at rest), and the two stable regions correspond to positions in which the associated F2 is either above or below F2T. The definition is specific enough to distinguish this feature from others, but general enough to apply to various types of front, central and back vowels as well as to the same vowel in different contexts. Finally, it identifies the class of sounds in which the definition holds (vowels). This definition is an *intrinsic* definition, since to apply it we need only examine the internal properties of the vowel. An advantage of using an intrinsic definition in this case is that it accounts for the fact that vowels can usually be identified as front or back in isolation. Another is that vowels typically occur next to consonants, in which F2 is less prominent or absent. (Landmark effects can be found in front-to-back vowel transitions, as in the transition

from [a] to [i] (Honda & Takano 2006), but vowels in hiatus are too infrequent in most languages to provide a primary basis for feature definition.).

Quantal acoustic-auditory relations

Further types of discontinuity can be found among certain acoustic-auditory relations (Stevens 1989). We consider an example involving vowels.

Vowels are often considered problematic for quantal analysis and it has been suggested that they may organize themselves instead according to an inherently gradient principle of maximal dispersion in perceptual space (e.g. Lindblom 1986). However, the fact that vowels pattern in terms of natural classes just as consonants do suggests that they are also organized in terms of features (see much phonological literature, as well as Schwartz et al. 1997: 281), raising the question of what these features are, and whether they are also quantal. A proposed quantal definition for the feature [\pm back] has been cited above, based on a region of F2 instability located in the mid-frequency range. Here we will examine evidence for the same feature from natural acoustic/auditory discontinuities.

Vowel-matching experiments have shown that vowel formant patterns are perceived not just on the basis of individual formant frequencies, but also according to the *distance* between formants. In such experiments, synthetic vowels with several formants are matched against synthetic one- or two-formant vowels. Subjects are asked to adjust the frequency of the only (or the higher) formant of the latter vowel so that it matches the former as closely as possible in quality. Results show that when two formants in the normal range for F1 and F2 are well separated, they tend to be heard as two separate spectral peaks, but when two formants approach each other across a certain threshold value, their mutual amplitude is strongly enhanced and they are perceptually integrated into a single peak whose value is intermediate between the two acoustic formants. The crucial threshold for this integration is usually estimated at a value around 3.5 bark (Chistovich & Lublinskaja 1979). The implication of these experiments is that some aspect of the response of the auditory system undergoes a qualitative change -- a discontinuity -- when the distance between two spectral prominences falls under a critical value.

Experiments with data involving Swedish vowels have confirmed this effect for higher formants as well (Carlson et al. 1970). In these experiments, synthetic vowels with five formants were matched against two-formant synthetic vowels. The first-formant frequency was the same for both vowels. Subjects were asked to adjust the second frequency F2' of the two-formant vowel to give the best match in quality to the corresponding five-formant vowel.

The results of the experiment are shown in Figure 2. Here, the frequencies of the first four formants in Hz are shown as lines and the F2' frequencies of the matching vowel are shown as rectangles. It is observed that when the spacing between F2 and F3 is less than about 3.0 bark, as it was for the front vowels (the first six in the figure), subjects place F2' at a frequency between F2 and F3 for all vowels except /i/. (In /i/, in which F3 is closer to F4 than to F2, they place F2' between F3 and F4.) In back vowels, in which higher formants have very low amplitude, F2' is placed directly on F2.

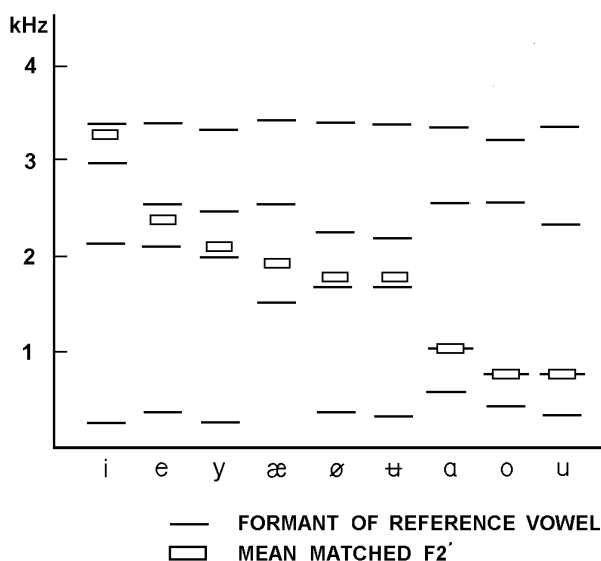


Figure 2. Results of a matching experiment in which subjects adjusted the frequency F2' of a two-formant vowel to give the best match in quality to each of nine Swedish five-formant vowels; only the four lowest formants are shown here. (After Carlson et al. 1970.)

These results indicate that there is a critical spacing of higher formants (F2, F3 and F4) leading to the interpretation of closely-grouped two-peak spectral prominences as single broad perceptual prominences. They give independent support for the view that the feature [\pm back] has a natural basis, in this case in terms of audition. We see that for [-back], but not [+back] vowels, the distance in Hz between F1 and the effective F2' is always greater than the distance between F1 and the acoustic F2. In other words, perception *magnifies* the front/back vowel distinction present in the acoustic structure.

While the difference between [-back] and [+back] vowels seems well-founded in quantal terms, it is much less clear that other features, such as those of vowel height and lip rounding, can be defined in these terms. For

example, there is no obvious discontinuity in the comparison of Swedish [+high] /u/ and [-high] /o/ in Figure 2. For reasons such as these, phoneticians usually tend to speak of quantal *vowels* rather than of quantal features. Quantal vowels are those in which two formants approach each other maximally, an effect known as focalisation (Schwartz et al. 1997). It is sometimes thought that /i/, /u/, /a/ and perhaps /y/ or /æ/ may constitute quantal vowels in this sense, though experimentally-based, multispeaker data bearing on this question is still rather scarce.

We do not propose, however, to abandon the search for nongradient definitions for vowel features. We tentatively suggest that features of vowel height -- setting aside the problematic feature [±ATR] -- may be defined in terms of the absolute boundary values set by the upper and lower range of each speaker. In this point of view, a vowel bearing the feature [+high] would be one whose perceived lowest prominence - let us call it P1 -- falls within an auditorily indistinguishable subrange of values at the bottom of a given speaker's total range of values for this prominence, while a [+low] vowel would be one whose perceived lowest prominence falls within the corresponding subrange at the top. A mid vowel, bearing the values [-high, -low], would be defined as falling within neither of these subranges. In other words, the speaker's total range of values for a given prominence P_n establishes the frame of reference with respect to which a given production is evaluated. While this account is not strictly quantal (as there appears to be no natural discontinuity as we pass up and down the vowel height scale), it has the advantage of tying the feature definition to a set of fixed reference points, defined in a way that is applicable to any speaker, regardless of the size and shape of their vocal tract. If it is true that vowel identification is more reliable as a vowel's values approach the periphery of the vowel triangle (see Polka & Bohn 2003), we can explain why distinctions among mid vowels (such as /e/ vs. /ɛ/) are much less stable across languages, in both historical and synchronic terms, than distinctions involving high vs. mid or mid vs. low vowels. These suggestions are quite tentative, of course, and we believe that future research should continue to seek possible quantal correlates of vowel height.

Summary

Our aim in this short tutorial has been to present a brief overview of a number of basic concepts of Quantal Theory, proposing a protocol according to which quantal feature definitions may be given. Quantal Theory offers a promising basis for redefining features in both articulatory and acoustic terms, overcoming the tradition competition between these two apparently incompatible approaches.

References

- Carson, R., Granström, B., and Fant, G. 1970. Some studies concerning perception of isolated vowels. *Speech Transmission Laboratory Quarterly Progress and Status Report* 2-3, 19-35. Royal Institute of Technology, Stockholm.
- Catford, J. C. 1977. *Fundamental Problems in Phonetics*. Bloomington, Indiana University Press.
- Chistovich, L.A. and V.V. Lublinskaja. 1979. The "center of gravity" effect in vowel spectra and critical distance between the formants : psychoacoustical study of the perception of vowel-like stimuli, *Hearing Research* 1, 185-195.
- Chomsky, N. and Halle, M. 1968. *Sound Pattern of English*. New York, Harper and Row.
- Honda, K. and Takano, S. 2006. Physiological and acoustic factors involved in /a/ to /i/ transitions. Invited talk, Colloquium on the Phonetic Bases of Distinctive Features, Paris, July 3.
- Huffman, M.K. 1990. Implementation of nasal: timing and articulatory landmarks. *UCLA Working Papers in Phonetics* 75, 1-149.
- Jakobson, R., Fant, C.M., and Halle, M. 1952. *Preliminaries to Speech Analysis*. Cambridge MA, MIT Press.
- Lindblom, B. 1986. Phonetic Universals in Vowel Systems. In J. J. Ohala and J. J. Jaeger (eds.), *Experimental Phonology*, 13-44. Orlando: Academic Press, Inc.
- Polka, L. and O.-S.Bohn. 2003. Asymmetries in vowel perception, *Speech Communication* 41, 221-231.
- Schwartz, J-L., Boë, L-J. Vallée, N. and Abry, C. 1997. The Dispersion-Focalisation Theory of Vowel Systems, *Journal of Phonetics* 25, 255-286.
- Stevens, K.N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17, 3-46.
- Stevens, K.N. 1998. *Acoustic Phonetics*. Cambridge, MA: MIT Press.
- Stevens, K. N. 2000. Diverse Acoustic Cues at Consonantal Landmarks. *Phonetica* 57, 139-51.
- Stevens, K.N. 2002. Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustic Society of America* 111, 1872-1891.
- Stevens, K. N. 2004. Invariance and variability in speech: interpreting acoustic evidence. *Proceedings of From Sound to Sense*, June 11 – June 13, 2004, B77-B85. Cambridge, MA, Speech Communication Laboratory, MIT.
- Stevens, K.N. 2005. Features in Speech Perception and Lexical Access. In Pisoni, D.E. and Remez, R.E. (eds.), *Handbook of Speech Perception*, 125-155. Cambridge, MA, Blackwell.

A new trainable trajectory formation system for facial animation

Oxana Govokhina^{1,2}, Gérard Bailly², Gaspard Breton² and Paul Bagshaw²

¹ Institut de la Communication Parlée, 46 av. Félix Viallet, F38031 Grenoble

² France Telecom R&D, 4 rue du Clos Courtel, F35512 Cesson-Sévigné

Abstract

A new trainable trajectory formation system for facial animation is here proposed that dissociates parametric spaces and methods for movement planning and execution. Movement planning is achieved by HMM-based trajectory formation. Movement execution is performed by concatenation of multi-represented diphones. Movement planning ensures that the essential visual characteristics of visemes are reached (lip closing for bilabials, rounding and opening for palatal fricatives, etc) and that appropriate coarticulation is planned. Movement execution grafts phonetic details and idiosyncratic articulatory strategies (dissymetries, importance of jaw movements, etc) to the planned gestural score.

Introduction

The modelling of coarticulation is in fact a difficult and largely unsolved problem (Hardcastle and Hewlett 1999). The variability of observed articulatory patterns is largely planned (Whalen 1990) and exploited by the interlocutor (Munhall and Tohkura 1998). Since the early work of Öhman on tongue movements (1967), several coarticulation models have been proposed and applied to facial animation. Bailly et al (Bailly, Gibert et al. 2002) implemented some key proposals and confronted them to ground-truth data: the concatenation-based technique was shown to provide audiovisual integration close to natural movements. The HMM-based trajectory formation technique was further included (Govokhina, Bailly et al. 2006). It outperforms both objectively and subjectively the other proposals. In this paper we further tune the various free parameters of the HMM-based trajectory formation technique using a large motion capture database (Gibert, Bailly et al. 2005) and compare its performance with the winning system of Bailly et al study. We finally propose a system that aims at combining the most interesting features of both proposals.

Audiovisual data and articulatory modelling

The models are benchmarked using motion capture data. Our audiovisual database consists of 238 (228 for training and 10 for test) French utterances

spoken by a female speaker. Acoustic and motion capture data are recorded synchronously using a Vicon© system with 12 cameras (Gibert, Bailly et al. 2005). The system delivers the 3D positions of 63 infra-red reflexive markers glued on the speaker's face at 120 Hz (see Figure 1). The acoustic data is segmented semi-automatically into phonemes. An articulatory model is built using a statistical analysis of the 3D positions of 63 feature points. The *cloning* methodology developed at ICP (Badin, Bailly et al. 2002; Revéret, Bailly et al. 2000) consists of an iterative Principal Component Analysis (PCA) performed on pertinent subsets of feature points. First, jaw rotation and protrusion (*Jaw1* and *Jaw2*) are estimated from the points on jaw line and their effects subtracted from the data. Then the lip rounding/spreading gesture (*Lips1*), the proper vertical movements of upper and lower lips (*Lips2* and *Lips3*), of the lip corners (*Lips4*) and of the throat (*Lar1*) are further subtracted from the residual data. These parameters explain 46.2, 4.6, 18.7, 3.8, 3.2, 1.6 and 1.3% of the movement variance.

The analysis of geometric targets of the 5690 allophones produced by the speaker (see Figure 2) reveals confusion trees similar to previous findings (Odisio and Bailly 2004). Consequently 3 visemes are considered for vowels (grouping respectively rounded [ʊψɔ̃], mid-open [iɛøɑ̃] and open vowels [æœœãẽ]) and 4 visemes for consonants (distinguishing respectively bilabials [pbm], labiodentals [fv], rounded fricatives [ʒ]) from the others).

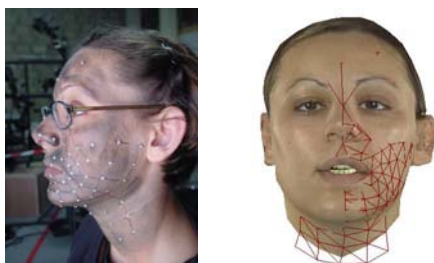


Figure 1: Motion capture data and videorealistic clone mimicking recorded articulation.

HMM-based synthesis

The principle of HMM-based synthesis was first introduced by Donovan for acoustic speech synthesis (Donovan 1996). This was extended to audiovisual speech by the HTS working group (Tamura, Kondo et al. 1999).

Training. An HMM and a duration model for each state are first learned for each segment of the training set. The input data for the HMM training is a set of observation vectors. The observation vectors consist of static and dynamic parameters, i.e. the values of articulatory parameters and their derivatives. The HMM parameter estimation is based on ML (Maximum-Likelihood) criterion (Tokuda, Yoshimura et al. 2000). Here, for each pho-

neme in context, a 3-state left-to-right model with single Gaussian diagonal output distributions and no skips is learned.

Synthesis. The phonetic string to be synthesized is first chunked into segments and a sequence of HMM states is built by concatenating the corresponding segmental HMMs. State durations for the HMM sequence are determined so that the output probability of the state durations is maximized. From the HMM sequence with the proper state durations assigned, a sequence of observation parameters is generated using a specific ML-based generation algorithm (Zen, Tokuda et al. 2004).

Note that HMM synthesis imposes some constraints on the distribution of observations for each state. The ML-based parameter generation algorithm requires Gaussian diagonal output distributions. It thus best operates on an observation space that has compact targets and characterizes targets with maximally independent parameters. We compared the dispersion of visemes obtained using different observation spaces: articulatory vs. geometric. Only lip geometry (aperture, width and protrusion) is considered. Despite its lower dimension, the geometric space provides less confusable visemes.

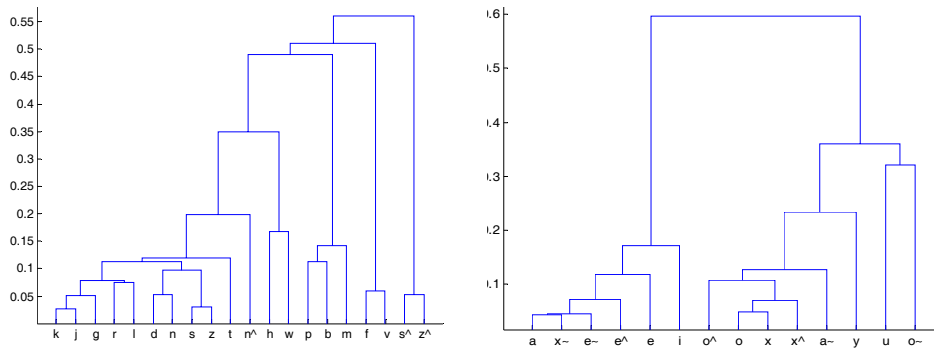


Figure 2. Grouping phonemes into viseme classes according to geometric confusability. Left: consonantal targets. Right: vocalic targets.

Detailed analysis

We compared phoneme-HMM with and without contextual information for selection. Table 1 summarizes our findings: anticipatory coarticulation is predominant, grouping context into visemes does not degrade performance. This contextual information enables the HMM system to progressively capture variability of allophonic realizations (see Figure 3). Syllable boundaries are known to influence coarticulation patterns. For this data, however, adding presence/absence of syllabic boundary does not improve the results (see bottom of Table 1). Sentence-internal (syntactic) pauses behave quite differ-

ently from initial and final pauses: initial pauses are characterized visually by prephonatory lips opening that reveals presence of initial bilabial occlusives if any; final pauses are characterized by a complete closure whereas the mouth often remains open during syntactic pauses especially when occurring between open sounds. We show that the viseme class immediately following the syntactic pause provides efficient contextual information for predicting lip geometry (see Table 2).

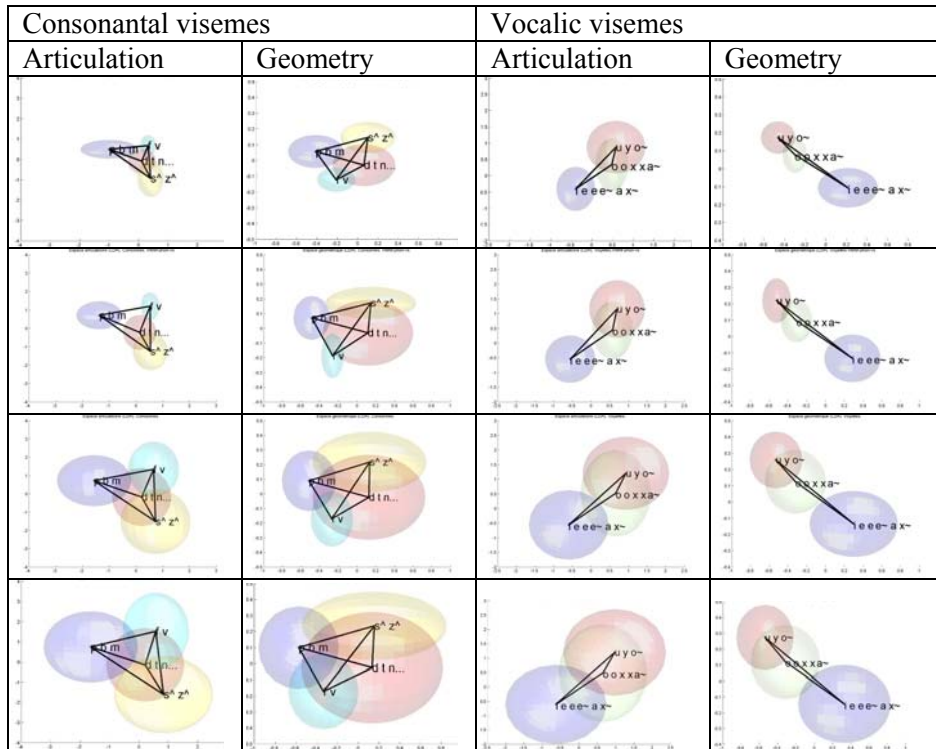


Figure 3. Projecting the consonantal and vocalic visemes on the first discriminant plane (set using natural reference) for various systems and two different parametric spaces: articulatory versus geometric. From top to bottom: phoneme HMM, phoneme HMM with next segment information, TDA and natural reference.

Table 1: Adding contextual information to an initial context-independent phoneme HMM. Mean correlation (\pm standard deviation) between observed and predicted trajectories using different phoneme HMM systems for geometric space; coverage (nb. of segments which number of samples is superior to ten divided by total nb. of segments) and mean nb. of samples (\pm standard deviations) are computed.

Phoneme HMM	Correlation	Coverage	Mean nb. of samples
Without context	0.77 \pm 0.07	1.00	164 \pm 112
Prev. phoneme	0.78 \pm 0.09	0.13	20 \pm 11
Next phoneme	0.83 \pm 0.06	0.13	20 \pm 11
Next viseme	0.83 \pm 0.07	0.23	31 \pm 35
adding syllable	0.84 \pm 0.06	0.12	28 \pm 26

Table 2: Mean correlations (\pm standard deviations) between the targets of the sentence-internal pauses and the targets of next (or previous) segment.

Target	Articulation	Geometry
Next	0.76 \pm 0.04	0.80 \pm 0.07
Previous	0.43 \pm 0.10	0.40 \pm 0.13

Table 3: Mean correlations (\pm standard deviations) between observed and predicted trajectories using different systems and representations.

System	Articulation	Geometry
Phoneme-HMM	0.61 \pm 0.11	0.77 \pm 0.07
Contextual phoneme-HMM	0.69 \pm 0.10	0.83 \pm 0.07
Concatenation of diphones	0.61 \pm 0.15	0.78 \pm 0.07
Concatenation with HMM selection	0.63 \pm 0.15	0.81 \pm 0.06
TDA	0.59 \pm 0.16	0.81 \pm 0.06

The proposed trajectory formation system

TDA (Task Dynamics for Animation), the proposed trajectory formation system, combines the advantages of both HMM- and concatenation-based techniques. The proposed system (see Figure 4) is motivated by articulatory phonology and its first implementation by the task dynamics model (Saltzman and Munhall 1989). Articulatory phonology put forward underspecified gestures as primary objects of both speech production and perception. In the task dynamics model, context-independent underspecified gestures first give spatio-temporal gauges of vocal tract constrictions for each phoneme. Then a trajectory formation model executes this gestural score by moving articulatory parameters shaping the vocal tract. In this proposal, the gestural score specifying the lip geometry (lip opening, width and protrusion) is first computed by HMM models. Then execution of this score is performed by a concatenation model where the selection score penalizes

segments according to their deviation from this planned geometry. The stored segments are thus characterized both by lip geometry for selection and by detailed articulation (jaw, separate control of upper and lower lips as well as rounding, etc) for the final generation.

Planning gestures by HMM synthesis. HMM-based synthesis outperforms both in objective and subjective terms concatenative synthesis and phoneme or diphone HMMs, when all these systems are trained to generate directly articulatory parameters. When trained on geometric parameters, these systems generate targets that are more discriminated and the correlation between original trajectories and those generated by all systems is substantially higher when considering geometry (see Table 3). This confirms previous studies that promote constrictions as the best characteristics for speech planning (Bailly 1998).

Executing gestures by concatenative synthesis. While diphone HMMs generate smooth trajectories while preserving visually relevant phonetic contrasts, concatenative synthesis has the intrinsic properties of capturing inter-articulatory phasing and idiosyncratic articulation. Concatenative synthesis also intrinsically preserves the variability of natural speech.

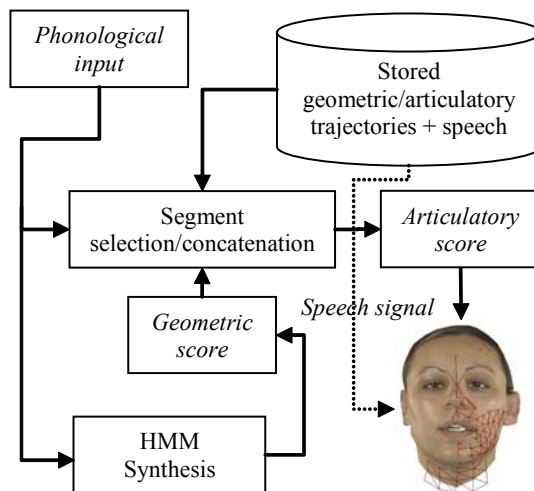


Figure 4: The proposed trajectory formation system TDA. A geometric score is thus computed by HMM-based synthesis. Segments are then retrieved that best match this planned articulation. Articulatory trajectories also stored in the segment dictionary are then warped, concatenated and smoothed and drive the talking head. Since the speech signal is generated using the same warping functions, audiovisual coherence of synthetic animation is preserved.

Performance analysis

Table 3 summarizes the comparative performance of the different systems implemented so far. Performance of the concatenation system is substantially increased when considering a selection cost using target parameters computed HMM trajectory planner. This is true whenever considering geometry or articulatory planning space. The performance of the current implementation of the TDA is however deceptive: the articulatory generation

often degrades the quality of the planned geometric characteristics. If the TDA compensates well for the bad planning of movement during syntactic pauses, it often degrades the timing (see Figure 5). We are currently reconsidering the procedure that warps stored articulatory segments to planned gestures.

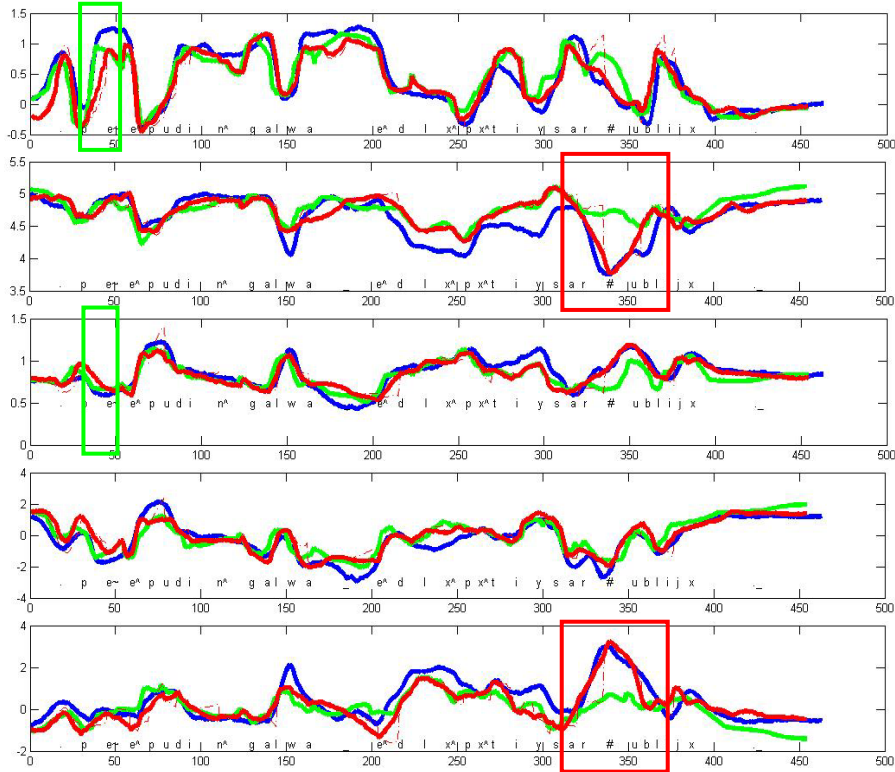


Figure 5. Comparing trajectory formation systems (blue: natural reference; red: concatenation/selection TDA; green: contextual phoneme-HMM) with a natural test stimulus (blue). From top to bottom: geometric parameters: lip aperture, width and protrusion; articulatory parameters: jaw aperture, lips rounding/spreading. Major discrepancies between TDA and contextual phoneme-HMM are enlighten.

Conclusions and perspectives

The TDA system is a trajectory formation system for generating speech-related facial movement. It combines a HMM-based trajectory formation system responsible for planning long-term coarticulation in a geometric space with a trajectory formation system that selects and concatenates segments that are best capable of realizing this gestural score. Contrary to most pro-

posals, this system builds on motor control theory – that identifies distinct modules for planning and execution of movements – and implements a theory of control of speech movements that considers characteristics of vocal tract geometry as primary cues of speech planning.

In the near future we will exploit in a more efficient way the information delivered by the HMM-based synthesis e.g. adding timing and spatial gauges to the gestural score in order to guide more precisely the segment selection.

References

- Badin, P., G. Bailly, L. Revéret, M. Baciú, C. Segebarth and C. Savariaux (2002). “Three-dimensional linear articulatory modelling of tongue, lips and face based on MRI and video images.” *Journal of Phonetics* **30** (3): 533-553.
- Bailly, G. (1998). “Learning to speak. Sensori-motor control of speech movements.” *Speech Communication* **22** (2-3): 251-267.
- Bailly, G., G. Gibert and M. Odisio (2002). Evaluation of movement generation systems using the point-light technique. *IEEE Workshop on Speech Synthesis*, Santa Monica, CA: 27-30.
- Donovan, R. (1996). Trainable speech synthesis. PhD thesis. Univ. Eng. Dept. Cambridge, UK, University of Cambridge: 164 p.
- Gibert, G., G. Bailly, D. Beautemps, F. Elisei and R. Brun (2005). “Analysis and synthesis of the 3D movements of the head, face and hand of a speaker using cued speech.” *Journal of Acoustical Society of America* **118** (2): 1144-1153.
- Govokhina, O., G. Bailly, G. Breton and P. Bagshaw (2006). Evaluation de systèmes de génération de mouvements faciaux. *Journées d'Etudes sur la Parole*, Rennes - France: accepted.
- Hardcastle, W. J. and N. Hewlett (1999). *Coarticulation: Theory, Data, and Techniques*. Cambridge, UK, Press Syndicate of the University of Cambridge.
- Munhall, K. G. and Y. Tohkura (1998). “Audiovisual gating and the time course of speech perception.” *Journal of the Acoustical Society of America* **104**: 530-539.
- Odisio, M. and G. Bailly (2004). “Tracking talking faces with shape and appearance models.” *Speech Communication* **44** (1-4): 63-82.
- Öhman, S. E. G. (1967). “Numerical model of coarticulation.” *Journal of the Acoustical Society of America* **41**: 310-320.
- Revéret, L., G. Bailly and P. Badin (2000). MOTHER: a new generation of talking heads providing a flexible articulatory control for video-realistic speech animation. *International Conference on Speech and Language Processing*, Beijing - China: 755-758.
- Saltzman, E. L. and K. G. Munhall (1989). “A dynamical approach to gestural patterning in speech production.” *Ecological Psychology* **1** (4): 1615-1623.
- Tamura, M., S. Kondo, T. Masuko and T. Kobayashi (1999). Text-to-audio-visual speech synthesis based on parameter generation from HMM. *EUROSPEECH*, Budapest, Hungary: 959-962.
- Tokuda, K., T. Yoshimura, T. Masuko, T. Kobayashi and T. Kitamura (2000). Speech parameter generation algorithms for HMM-based speech synthesis. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Istanbul, Turkey: 1315-1318.
- Whalen, D. H. (1990). “Coarticulation is largely planned.” *Journal of Phonetics* **18** (1): 3-35.
- Zen, H., K. Tokuda and T. Kitamura (2004). An introduction of trajectory model into HMM-based speech synthesis. *ISCA Speech Synthesis Workshop*, Pittsburgh, PE: 191-196.

Topics in speech perception

Diane Kewley-Port
Department of Speech and Hearing Sciences, Indiana University,
Bloomington, USA

Abstract

The study of speech perception over the past 60 years has tried to determine the human processes that underlie the rapid understanding of fluent speech. A first step was to determine the units of speech that could be experimentally manipulated. Years of examining the acoustic properties associated with phonemes led to theories such as the Motor Theory which postulate larger units that integrate vowel and consonant information. Current approaches find better support for the syllable as the most robust and coherent unit of speech. A complete theory of speech perception should systematically map how speech acoustic information is processed bottom-up through the peripheral and central auditory system, as well as how linguistic knowledge interacts top-down with the acoustic-phonetic information to extract meaning.

Introduction

The goal of the study of speech perception is to understand how fluent speech in typical environments is processed by humans to extract the talker's intended message. Towards this goal, the present overview will address three topics: (1) What are the units of speech perception? (2) How is speech processed from the peripheral to central auditory system? (3) What are the effects of the enormous variability observed in speech on speech perception?

Units of speech

Consider a fluent spoken sentence, such as "But that explanation is only partly true" (from TIMIT, Garfalo et al., 1993) recorded in the quiet (Fig. 1). The observed rapidly changing spectrotemporal properties in the spectrogram are typical of normal speech and permit a high transmission rate of information between human beings. What is even more remarkable is that communication does not usually take place in quiet, but rather in listening environments that are noisy or reverberant or have competing talkers, or in all three degrading circumstances, and yet speech understanding remains high. What do we know about how humans perceive speech?

A primary theoretical issue in speech perception is to determine the units of speech that are essential to describe human communication. Given a particular unit, various experiments can be conducted to manipulate speech and

examine the resulting perceptual consequences. Writing systems have relatively clear units such as alphabets (phonemes, Roman alphabet), syllables (Japanese hiragana syllabary) and words (Mayan) to represent graphically some of the information found in spoken language. Linguists have additionally postulated feature systems (Jacobson, Fant and Halle, 1952; Chomsky and Halle, 1968) as the basic units of speech. Thus although speech generally consists of multiword sequences (phrases and sentences), the largest unit typically used to represent speech is the word. For example, a variety of computer-based speech recognition systems have been designed to identify whole words, and those that identify words in isolation are considerably more successful than continuous word recognition in sentences such as those in Figure 1 (Lippmann, 1997).

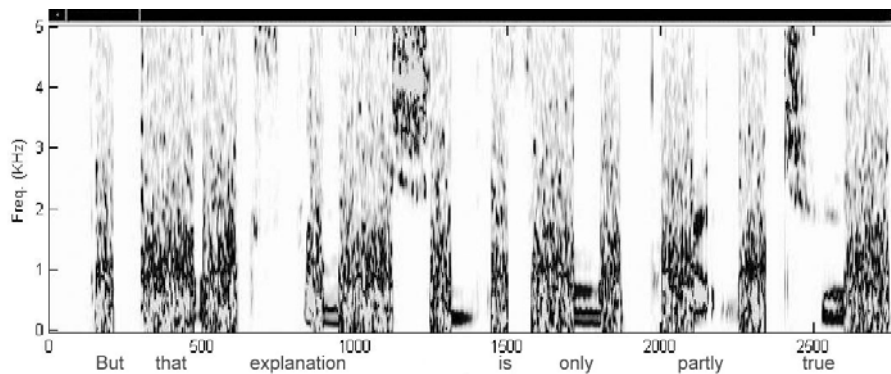


Figure 1. Spectrogram of a sentence with text roughly aligned in time.

The point of view taken in this overview of speech perception is that good experimental support must be demonstrated for postulated units of speech. Consider linguistic features. Stevens and his colleagues (1998) have long studied the acoustic properties of features such as place of articulation (Stevens and Blumstein, 1978) or sonorant/nonsonorant (/n/ versus /d/) to demonstrate that some of these properties are invariant across considerable talker variability. Recently Stevens (2002) has proposed a model that specifically states that lexical access from speech is based on the processing of feature bundles that are structured into phoneme segments. The primary evidence against this approach using discrete units is found in the substantial acoustic effects of coarticulation between segments in fluent speech (Liberman et al., 1967; Diehl et al., 2004), as well as the influence of the temporal properties of speech on linguistic categories (Port and Leary, 2005). Moreover, the details of speech acoustics required for a feature-based model such as Stevens (2002) are generally only available in quiet conditions. As noted above, human speech perception is robust under substantial

amounts of noise. This is because the speech signal is highly redundant and therefore speech perception only requires partial information to successfully extract the intended message. In the past ten years a great deal of research on speech processed through simulated cochlear implants has demonstrated that only a small number of frequency channels, four to seven, are needed to recognize sentences (Shannon et al., 1996; Dorman et al., 1997). In fact, in the extreme Remez and his colleagues (Remez et al., 1981, 1994) have demonstrated that sentences can be recognized from only three frequency modulated tones that approximate the first three formants of speech (sinewave speech), even though each individual tone sounds nothing like speech.

Given the strong evidence against discrete features or segments as being the units of speech perception, what is the alternative? There is a long history of proposing that the primary unit of speech is the gesture, starting with Liberman and his colleagues who postulated the Motor Theory of Speech Perception (Liberman et al., 1967) as based on CV units. This was followed at Haskins by Fowler (1984) whose Direct Realist Theory also referenced the speech gesture as the basic unit, but one described as having the vowel and consonant information coproduced and perceived. Additional support also from Haskins has been given by the speech production research of Browman and Goldstein (1992) whose theory of Articulatory Phonology provides details about the organization of consonants and vowels into coordinated gestures. How do these models of articulation and speech production relate to speech perception? As Browman and Goldstein (1988) initially argued, and as Studdert-Kennedy (1998) clearly states, the central unit of speech is the syllable. The syllable is the smallest unit in which the acoustic properties and temporal structure of speech are integrated into a coherent structural unit with the underlying articulatory gesture. This syllabic unit has properties related directly to other units larger (words) and smaller (features, phonemes) than the syllable. However, the syllable is the central unit that has structural descriptions that correspond across speech production, speech perception, the motor control of articulation, stored cognitive categories of speech and even language acquisition by infants (Studdert-Kennedy, 1998). Studdert-Kennedy (1998; Studdert-Kennedy and Goldstein, 2003) has argued that the relation between the syllable and other units in speech can be described in terms of the *particulate principle* in which the combination of smaller structures creates a functionally different set of objects at the next higher level. Of special importance to the view of speech perception described here is the fact that the strong coherence of acoustic information across frequency and time in syllables means that syllables can still be perceived when only fragmentary information is available, for example due to strange processing schemes (cochlear implant simulators) or noisy conditions.

Peripheral and central mechanisms in speech processing

Whatever linguistic units are the basis of speech perception, processing of the acoustic signal starts at the auditory periphery. Kewley-Port and her colleagues have attempted to describe processing of vowels using psychophysical methods to understand how the acoustic signal is represented at the most peripheral levels of the auditory system (Kewley-Port, 1991; Kewley-Port and Watson, 1994), and then describe how more central levels of linguistic processing interact with that information (Kewley-Port, 2001; Liu and Kewley-Port, 2004a). This research program began by establishing the smallest detectable difference (threshold) in a vowel formant between a standard vowel and a test vowel that can be discriminated under optimal listening conditions (after extensive training while listening to only one formant per block in the quiet). Results demonstrated that fine detail in the vowels is represented in the peripheral auditory system. Threshold differences across vowels and talkers (Kewley-Port and Zheng, 1999), and in quiet and noise (Liu and Kewley-Port, 2004b) can be modelled using loudness patterns derived from computational models for simple non-speech stimuli developed by Moore and Glasberg (1997). The conclusion of this research is that the first stages of processing of vowels yield considerable more detail about these complex, harmonic spectra than is needed to categorize vowels.

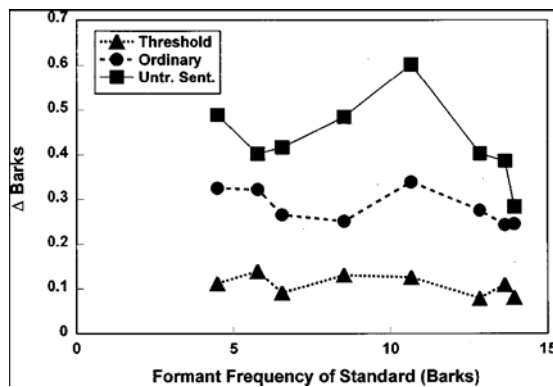


Figure 2. Thresholds for formant discrimination in Δ barks are displayed as a function of the center frequency of F1 and F2 for eight vowels. The function labelled "Threshold" is for discrimination of isolated vowels under optimum listening conditions. The function labelled "Ordinary" is for vowels embedded in phrases and sentences. The function labelled "Untr. Sent." is for the same listeners as for Ordinary, but for the Δ bark values obtained in the first half hour of testing, before listeners were trained (after Kewley-Port and Zheng, 1999).

ded in phrases and sentences. The function labelled "Untr. Sent." is for the same listeners as for Ordinary, but for the Δ bark values obtained in the first half hour of testing, before listeners were trained (after Kewley-Port and Zheng, 1999).

As more variability is included in the vowel stimuli or the task complexity increases, more central levels of processing are required to perform the vowel discrimination task. In Fig. 2 the baseline vowel thresholds (in barks) under optimal conditions are shown to be relatively constant at about 0.11 barks. In a study exploring vowels in more ordinary listening conditions (Kewley-Port and Zheng, 1999) where many vowel formants were tested in phrases and sentences, higher levels of linguistic context elevated thresholds (labelled Ordinary) by a factor of 3 to 0.33 barks. However, formant discrimination (labelled Untr. Sent.) that was measured in sentences before the subjects were trained was elevated by a factor of 5.

Our results from vowel discrimination and identification tasks demonstrate that the information available about vowels in the periphery becomes degraded as more central processes are needed to process sentences or to learn new tasks. However, when the well learned task of identifying the words in sentences was added to the discrimination task in sentences (Liu and Kewley-Port, 2004a), vowel thresholds for discrimination remained similar, just above the 0.33 bark threshold measured under more ordinary listening conditions. The implication is that when adults listen to their native language, auditory processing of vowels has a threshold norm of one-third of a bark that represents the human ability to extract critical vowel spectral information in fluent sentences. This norm limits the bottom-up processing capabilities for vowel spectra. However, predictive information from top-down processing may enhance listeners' abilities to categorize vowels, as well as visual information from the face. A complete picture of speech perception needs to establish a systematic relation between peripheral and central mechanisms for processing consonants and vowels both in syllables and in sentences.

Variability in speech

A hallmark of spoken language is the large amount of variability observed in speech. For example, if different talkers in different environments all spoke the same sentence, normal native listeners would all write down the same sentence in spite of the high variability in the acoustic signal. This “nonlinguistic” variability includes considerable information about the speaker, referred to as the indexical properties of speech (gender, emotion, Nygaard and Pisoni, 1998), as well as speaking style variation (rate, “clear speech”, Ferguson, 2004), and the cross-linguistic interference of accented speech (Flege, 1988). In the scenario above, little of this variability is preserved in the written transcription of the sentence, i.e. this nonlinguistic information is stripped off. But is it correct to treat this variability as random noise? The

clear answer is no for at least two reasons. First, listeners clearly use the indexical properties of speech in the give-and-take of every day conversation. Moreover, evidence has accumulated that this information can be stored as part of the representation of words in memory (episodic memory, Goldinger, 1998). Perhaps more important in normal discourse is that different speaking styles and rates affect the successful transmission of the intended message to the listener. Thus what has been considered nonlinguistic variability in speech can be manipulated for the purposes of improving speech intelligibility (Ferguson, 2004), and therefore represents structured information, and not random noise, in speech.

And finally, after this brief overview of many factors found to be important to understanding speech perception over the past 60 years, let's consider whether or not a comprehensive theory of speech perception is possible, at least in the near future. Stevens (2002) clearly believes that his theory is close to describing perceptual processes that span cognitive mechanisms representing the fine detail of speech in features through the retrieval of the associated words in the lexicon. However, the arguments proposed here suggest that this type of discrete unit model is not an adequate approach to understanding the mechanisms of speech perception. Rather, the approach taken by Studdert-Kennedy (1998) that uses the particulate principle for describing the structures of human behaviour is more likely to succeed. That is, we should agree that fine detail in speech may be captured by acoustic features as shown by Stevens (2002), but also acknowledge that this detail is restructured into higher level objects that have inherently different properties than feature bundles have by themselves. The particulate principle approach suggests that the syllable is the central unit that provides the most coherent relations between the structures of other units, both smaller and larger than the syllable. Whether or not this is true, our knowledge is incomplete for describing the relation between these units in the quiet, and research on the robustness of speech in noise (the typical environmental condition) is in its infancy. In fact, mechanisms for processing speech under the variety of adverse circumstances that humans encounter may differ substantially from one another (e.g. is listening in noise the same as trying to understand accented speech?). Building more comprehensive models of speech perception will require much more research.

Acknowledgements

Preparation of this manuscript supported by NIHDCD-02229.

References

- Browman, C.P. and Goldstein, L. 1988. Some Notes on Syllable Structure in Articulatory Phonology. *Phonetica* 45, 140-155.
- Browman, C. and Goldstein, L. 1992. Articulatory phonology: an overview. *Phonetica*. 49, 155-80
- Chomsky, N. and Halle, M. 1968. *The sound pattern of English*. New York:Harper and Row.
- Diehl, R., Lotto, A. and Holt, L. 2004. Speech Perception. *Annu. Rev. Psychol.* 55, 149-179.
- Dorman, M.F., Loizou, P.C. and Rainey, D. 1997. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *J. Acous. Soc. Am.* 102, 2403-2410.
- Fowler, C.A. 1984. Segmentation of coarticulated speech in perception. *Percept. & Psychophy.* 36, 359-368.
- Garofolo, J., Lamel, L., Fisher, W, Fiscus, J., Pallett, D., and Dahlgren, N. 1993. DARPA TIMIT: Acoustic-Phonetic Continuous Speech Corpus.
- Goldinger, S. 1998. Echoes of echoes? An episodic theory of lexical access. *Psych. Rev.*, 105, 251-279.
- Goldstein, L. and Fowler, C.A. 2003. Articulatory phonology: A phonology for public language use. In Schiller, N.O. and Meyer, A.S. (eds.), *Phonetics and Phonology in Language Comprehension and Production*, 159-207. Mouton de Gruyter.
- Jakobson, R., Fant, G., and Halle, M. 1952. *Preliminaries to speech analysis: The distinctive features*. Cambridge, MA: MIT Press.
- Kewley-Port, D. 1991. Detection thresholds for isolated vowels. *J. Acoust. Soc. Am.* 89, 820-829.
- Kewley-Port, D. 2001. Vowel formant discrimination II: Effects of stimulus uncertainty, consonantal context and training. *J. Acoust. Soc. Am.* 110, 2141-2155.
- Kewley-Port, D. and Watson, C.S. 1994. Formant-frequency discrimination for isolated English vowels, *J. Acoust. Soc. Am.* 95, 485-496.
- Kewley-Port, D. and Zheng, Y. 1999. Vowel formant discrimination: Towards more ordinary listening conditions. *J. Acoust. Soc. Am.* 106, 2945-2958.
- Liberman, A., Cooper, F., Shankweiler, D. and Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychol. Rev.* 74, 431-461.
- Lippmann, R. 1997. Speech recognition by machines and humans. *Speech Com.* 22, 1-15.
- Liu, C. and Kewley-Port, D. 2004a. Vowel formant discrimination in high-fidelity speech. *J. Acoust. Soc. Am.* 116, 1224-1233.
- Liu, C. and Kewley-Port, D. 2004b. Formant discrimination in noise for isolated vowels. *J. Acoust. Soc. Am.* 116, 3119-3129.
- Moore, B. C. J., and Glasberg, B. R. 1997. A model of loudness perception applied to cochlear hearing loss. *Auditory Neurosci.* 3, 289-311.
- Nygaard, L. and Pisoni. D. 1998. Talker-specific perceptual learning in speech perception. *Percept. & Psychophy.* 60, 355-376.

- Port, R. and Leary, A. 2005. Against formal phonology. *Language* 72, 927–964.
- Remez, R.E., Rubin, P.E., Pisoni, D.B., and Carrell, T.D. 1981. Speech perception without traditional speech cues. *Science* 212, 947-950.
- Remez, R.E., Rubin, P.E., Berns, S.M., Pardo, J.S., and Lang, J.M. 1994. On the Perceptual Organization of Speech. *Psych. Rev.* 101, 129-156.
- Shannon, R., Zeng, F-G, and Wygonski, J. 1996. Altered temporal and spectral patterns produced by cochlear implants: Implications for psychophysics and speech recognition. *J. Acoust. Soc. Am.* 96, 2470-2500.
- Stevens, K.N. 1998. *Acoustic Phonetics*. Cambridge, MA, MIT Press.
- Stevens, K.N. 2002. Toward a model for lexical access based on acoustic landmarks and distinctive features. *J. Acoust. Soc. Am* 111, 1872-1891.
- Stevens, K.N. and Blumstein, S. 1978. Invariant cues for place of articulation in stop consonants. *J. Acoust. Soc. Am.* 64, 1358-1368.
- Studdert-Kennedy, M. 1998. The particulate origins of language generativity: from syllable to gesture. In: Hurford, J., Studdert-Kennedy, M., and Knight, C. (eds.), *Approaches to the evolution of language*, Cambridge University Press, Cambridge, U.K.
- Studdert-Kennedy, M. and Goldstein, L. 2003. Launching language: The gestural origin of discrete infinity. In Morten Christiansen and Simon Kirby (eds.), *Language Evolution*, 235-254 Oxford University Press, Oxford, U.K.

Spatial representations in language and thought

Anna Papafragou

Department of Psychology, University of Delaware, USA

Abstract

The linguistic expression of space draws from and is constrained by basic, probably universal, elements of perceptual/cognitive structure. Nevertheless, there are considerable cross-linguistic differences in how these fundamental space concepts are segmented and packaged into sentences. This cross-linguistic variation has led to the question whether the language one speaks could affect the way one thinks about space – hence whether speakers of different languages differ in the way they see the world. This chapter addresses this question through a series of cross-linguistic experiments comparing the linguistic and non-linguistic representation of motion and space in both adults and children. Taken together, the experiments reveal remarkable similarities in the way space is perceived, remembered and categorized despite differences in how spatial scenes are encoded cross-linguistically.

Introduction

The linguistic expression of space draws from and is constrained by basic, probably universal, elements of perceptual/cognitive spatial structure. As is well known, the representation of space is a fundamental human cognitive ability (Pick & Acredolo 1983; Stiles-Davis, Kritchevsky & Bellugi 1988; Emmorey & Reilly 1995; Hayward & Tarr 1995; Newcombe & Huttenlocher 2003; Carlson & van der Zee 2005), and appears early in life (Pulverman, Sootsman, Golinkoff & Hirsh-Pasek 2003; Casasola, Cohen & Chiarello 2003; Casasola & Cohen 2002; Quinn 1994; Pruden, Hirsh-Pasek, Maguire, Meyers & Golinkoff 2004).

Nevertheless, there are considerable cross-linguistic differences in how these fundamental space components are segmented and packaged into sentences. This cross-linguistic variation has led to the question whether the way space is encoded cross-linguistically affects the way space is perceived, categorized and remembered by people who speak different languages (Bowerman & Levinson, 2001; cf. Whorf, 1956). The goal of this paper is to address this question focusing on two strands of empirical work.

Motion events

The first set of studies we will review focus on a comparison of the linguistic and nonlinguistic representation of motion in speakers of English and Greek. These two languages differ in the way they encode the trajectory, or path,

and the manner of motion (cf. Talmy, 1985): English includes a large class of manner of motion verbs (*strut, stroll, sashay*, etc.) which can be freely combined with adverbs, particles or prepositional phrases encoding trajectory information (*away, into the forest, upwards*, etc.). By contrast, Greek mostly expresses motion information in path verbs (*beno* ‘enter’, *vjeno* ‘exit’, *perno* ‘cross’, *pao* ‘go’, etc.) combined with prepositional phrases or adverbials which further specify path (*sto spiti* ‘into the house’, *makria* ‘away’, etc.). Greek does have a substantial inventory of manner verbs (*xorevo* ‘dance’, *trexo* ‘run’, *pleo* ‘float’, etc) but their distribution is constrained by what we will call a ‘boundedness constraint’: most manner verbs cannot combine with a modifier which denotes a bounded, completed path (**To puli petakse sto kluvi*) unlike their English counterparts (*The bird flew into the cage*). This constraint leads to higher use of path verbs in Greek compared to English. A similar constraint is found in several languages (Aske 1989; Jackendoff 1990; Slobin & Hoiting 1994; Levin & Rapoport 1988) and has led commentators to conclude that manner of motion is less salient as a verb grammaticalization feature in languages such as Greek.

In our own work (Papafragou, Massey & Gleitman 2002, 2006), we have confirmed the Manner/ Path asymmetry in the description of motion scenes by Greek- versus English-speaking children and, much more strongly, for Greek versus English-speaking adults. The very same studies, however, revealed no differences in the English- and Greek- speaking subjects’ memory of path or manner details of motion scenes. Further experiments showed that, despite the asymmetry in verbally encoding motion events, English and Greek speakers did not differ from each other in terms of motion event categorization. More recent studies compared on-line inspection of motion events by Greek- and English-speaking adults using eye-tracking methodology (Papafragou, Hulbert & Trueswell, 2006). Taken together, the experiments reveal remarkable similarities in the way motion is perceived, remembered and categorized despite differences in how motion scenes are encoded cross-linguistically.

Spatial frames of reference

The second set of experiments focuses on the linguistic description of location and orientation (Li, Abarbanell & Papafragou, 2006). We study the spatial abilities of speakers of Tseltal, a Mayan language which lacks projective terms for *left* and *right*. Unlike English or other familiar languages, Tseltal speakers use absolute terms equivalent to north/south/east/west to locate objects in small-scale space (Levinson, 1996). As a result of this gap in linguistic resources, Tseltal speakers have been claimed not to use left-right distinctions in their habitual reasoning about

space (Pederson, Danziger, Wilkins, Levinson, Kita, & Senft, 1998; but see Li & Gleitman, 2002 for critical discussion).

Our experiments test the use of left/right concepts in Tseltal speakers and compare them to absolute systems of spatial location and orientation (Li et al., 2006). We find that Tseltal speakers, when given implicit cues that body-centered (left-right) distinctions are needed to solve a spatial task, use these distinctions without problems. On certain tasks, performance with such body-centered distinctions is better than performance with absolute systems of orientation which correspond more closely to the preferred linguistic systems of encoding space in Tseltal. These results argue against the claim that left-right distinctions are dispreferred or less salient in Tseltal spatial cognition. We take this as another demonstration of the independence of spatial reasoning from linguistic (encoding) preferences. We conclude that the linguistic and non-linguistic representations of space, even though correlated, are distinct and dissociable.

References

- Aske, J. 1989. Path predicates in English and Spanish: A closer look. Proceedings of the 15th Annual Meeting of the Berkeley Linguistics Society, 1-14. Berkeley, CA: BLS.
- Bowerman, M. and Levinson, S., eds. 2001. Language acquisition and conceptual development. Cambridge: Cambridge University Press.
- Carlson, L. and van der Zee, E., eds. 2005. Functional features in language and space: Insights from perception, categorization and development. Oxford: Oxford University Press.
- Casasola, M. and Cohen, L. 2002. Infant spatial categorization of containment, support or tight fit spatial relations. *Developmental Science*, 5, 247-264.
- Casasola, M., Cohen, L.B. and Chiarello, E. 2003. Six-month-old infants' categorization of containment spatial relations. *Child Development*, 74, 679-693.
- Choi, S., and Bowerman, M. 1991. Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, 41, 83-122.
- Emmorey, K., and Reilly, J., eds. 1995. Language, gesture and space. Hillsdale, NJ: Erlbaum.
- Hayward, W.G. and Tarr, M.J. 1995. Spatial language and spatial representation. *Cognition*, 55, 39-84.
- Jackendoff, R. 1990. Semantic structures. Cambridge, MA: MIT Press.
- Levin, B., and Rapoport, T. 1988. Lexical subordination. Papers from the 24th Regional Meeting of the Chicago Linguistics Society, 275-289. Chicago, IL: University of Chicago.
- Levinson, S. 1996. Frames of reference and Molyneux's question: Crosslinguistic evidence. In P. Bloom, M. Peterson, L. Nadel and M. Garrett eds., *Language and space*, 109-170. Cambridge, MA: MIT Press.

- Li, P., Abarbanell, L., and Papafragou, A. 2005. Language and spatial reasoning in Tenejapan Mayans. Proceedings from the Annual Meeting of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.
- Li, P., and Gleitman, L. 2002. Turning the tables: Spatial language and spatial cognition. *Cognition*, 83, 265-294.
- Newcombe, N. and Huttenlocher, J. 2003. Making space: The development of spatial representation and reasoning. Cambridge, MA: MIT Press.
- Papafragou, A., Hulbert, J., and Trueswell, J. 2006. Does language guide event perception: Evidence from eye movements. Talk delivered at the Annual Meeting of the Linguistic Society of America, Albuquerque, 5-8 January.
- Papafragou, A., Massey, C., and Gleitman, L. 2002. Shake, rattle, 'n' roll: the representation of motion in language and cognition. *Cognition*, 84, 189-219.
- Papafragou, A., Massey, C., and Gleitman, L. 2006. When English proposes what Greek presupposes: The cross-linguistic encoding of motion events. *Cognition*, 98, B75-87.
- Pederson, E., Danziger, E., Wilkins, D., Levinson, S., Kita, S. & Senft, G. 1998. Semantic typology and spatial conceptualization. *Language*, 74, 557-589.
- Pick, H., and Acredolo, L., eds. 1983. *Spatial orientation: theory, research and application*. New York: Plenum Press.
- Pruden, S.M., Hirsh-Pasek, K., Maguire, M., Meyers, M., and Golinkoff, R. M., 2004. Foundations of verb learning: Infants form categories of path and manner in motion events. *BUCLD 28*, 461-472. Somerville, MA: Cascadilla Press.
- Pulverman, R., Sootsman, J., Golinkoff, R.M., and Hirsh-Pasek, K. 2003. Infants' non-linguistic processing of motion events: One-year-old English speakers are interested in manner and path. In E. Clark ed., *Proceedings of the Stanford Child Language Research Forum*. Stanford: Center for the Study of Language and Information.
- Quinn, P.C. 1994. The categorization of above and below spatial relations by young infants. *Child Development*, 65, 58-69.
- Slobin, D., and Hoiting, N. 1994. Reference to movement in spoken and signed languages: Typological considerations. *Proceedings of the 20th Annual Meeting of the Berkeley Linguistics Society*, 487-505. Berkeley: BLS.
- Stiles-Davis J., Kritchevsky, and Bellugi, U. eds. 1988. *Spatial cognition: brain bases and development*. Hillsdale, NJ: Erlbaum.
- Talmy, L. 1985. Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen ed., *Language typology and syntactic description*, 57-149. New York: Cambridge University Press.

Sensorimotor control of speech production: models and data

Joseph S. Perkell

Speech Communication Group, Research Laboratory of Electronics,
Massachusetts Institute of Technology, USA

Abstract

A theoretical, model-based overview of speech production is described in which the goals of phonemic speech movements are in auditory and somatosensory domains and the movements are controlled by a combination of feedback and feedforward mechanisms. Examples of experimental results are presented that provide support for the overview.

Introduction

Speech production is an extraordinarily complex feat of motor coordination that conveys linguistic information, primarily via an acoustic signal, from a speaker to a listener. For information transmission at the phonemic level, speech sounds are differentiated from one another by the type of sound source and also by the wide variety of vocal-tract configurations that are produced by movements of the mandible, pharynx, tongue, soft palate (velum) and lips. These structures are slowly moving; however, because their movements are coarticulated, speech sounds can be produced in intelligible sequences at rates as high as about 15 per second.

The current paper focuses on the control of phonemic movements of the vocal-tract articulators, which are generated by the coordinated contractions of over 50 different muscle pairs. Clearly, the control mechanism is faced with a large number of degrees of freedom and the control problem is immensely complicated.

Models and Data

Phonemic goals

It is widely acknowledged that properties of the speech production mechanism have had major influences on the inventories of sounds or phonemes that languages employ, and also on some of the strategies that languages adopt for concatenating phonemes into meaningful sequences. A great deal of research on speech motor control and the mechanisms that underlie sound

categories has been directed at identifying the *controlled variables*, that is, the basic units of speech motor programming. To address this issue, investigators have asked, “What is the *task space*, or the domain of the fundamental control parameters?”

Our approach to such questions is motivated by observing that the objective of the speaker is to produce sound strings with acoustic cues that can be transformed into intelligible patterns of auditory sensations in the listener. These acoustic cues consist mainly of time-varying patterns of formant frequencies for vowels and glides, and noise bursts, silent intervals, aspiration and frication noises and rapid formant transitions for consonants. The properties of such cues are determined by parameters that can be observed in several domains, including: levels of muscle tension, movements of articulators, changes in the vocal-tract area function and aerodynamic events. Hypothetically, motor control variables could consist of any combination of these parameters.

In order to make the approach to this issue as tractable as possible, we formulate research hypotheses in terms of the function of the DIVA model of speech motor planning (cf. Guenther et al, 2006). DIVA is a neurocomputational model of relations among cortical activity, its motor output and the resulting sensory consequences of producing speech sounds. In the model, phonemic goals are encoded in neural projections (mappings) from premotor cortex to sensory cortex that describe *regions in multidimensional auditory-temporal and somatosensory-temporal spaces*. The model has two control subsystems, a feedback subsystem and a feedforward subsystem. Feedback control employs error detection and correction to teach, refine and update the feedforward control mechanisms. As speech is acquired and becomes fluent, speech sounds, syllables and words become encoded as strings of feedforward commands.

How are phonemic goal regions determined? One factor is based on properties of speakers’ production mechanisms that are characterized by quantal relations between articulation and acoustics (Stevens, 1989). There are a number of examples in which a continuous change in an articulatory parameter produces discontinuous changes in a salient acoustic parameter, resulting in regions of relative acoustic stability and regions of rapid change. Modelling and experimental results support the idea that such regions of stability help to define phonemic goals and sound categories (cf. Stevens, 1989; 1998; Perkell and Nelson, 1985; Perkell et al., 2000).

There are also quantal relations between articulatory movements and the area function, which are expressed when two articulators come into contact with one another. Fujimura and Kakita (1979) have modelled such a “saturation effect” for the vowel /i/ by showing how the vocal-tract cross-sectional area at the acousti-

cally sensitive place of maximum constriction can be stabilized by pressing the lateral edges of a stiffened tongue blade against the sides of the hard palate.

Another general, model-based principle that likely influences phoneme categories is a balance between sufficient perceptual contrast and ease of articulation (called “economy of effort” – Lindblom, 1990). Other important influences on sound systems are not amenable to being modelled, and it is not claimed that quantal effects and a balance between contrast and economy of effort can themselves account for the wide variety of sounds that are found in different languages. Nevertheless, quantifiable principles can provide a general framework for the formation of sound patterns, and more specific implementations of these particular principles can be utilized by individual speakers. An example of one such implementation is given below for a saturation effect. Other examples below provide support for some of the features of the DIVA model, including the use of sensory goals regions and feedback and feedforward control.

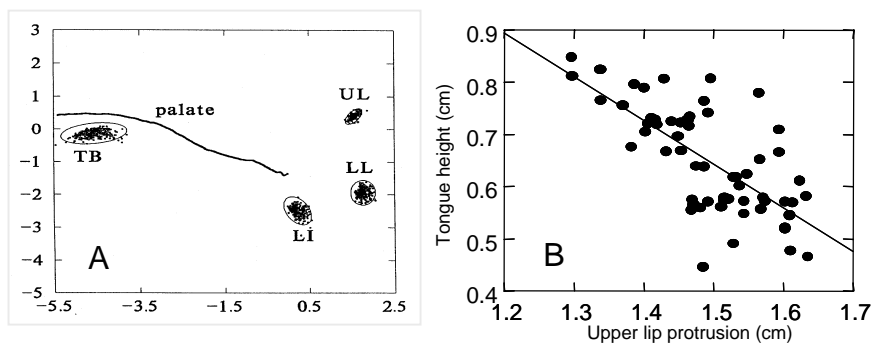


Figure 1. A: Example of points on the tongue body (TB), Upper lip (UL), lower lip (LL) and lower incisors (LI) for many repetitions of the vowel /u/ by a single speaker in a context phrase. B: Tongue height versus lip protrusion for many repetitions of the vowel /u/ by a single speaker.

Auditory goals for /u/ and /r/: Motor equivalence

The vowel /u/ in American English is produced by forming a narrow constriction with tongue raising in the velo-palatal region and by rounding the lips. Because of many-to-one relations between vocal-tract shapes and acoustics, approximately the same acoustic output can be produced with more tongue raising and less lip rounding and vice-versa. Figure 1B shows an example of tongue height versus lip protrusion for many repetitions of the vowel /u/ by a single speaker. The negative correlation reflects a motor-equivalent trading relation between the two articulations. Such reciprocal variation of two independently controllable articulations provides evidence

that the goal for the vowel /u/ is in an acoustic/auditory frame of reference, rather than a spatial or gestural one (Perkell et al., 1993). Evidence of an acoustic/auditory goal for /r/ in American English was obtained in a similar motor-equivalence study by Guenther et al. (1999).

Auditory goals: Relations between speech production and perception

Further insight about auditory goals can be gained by examining relations between speech production and perception. It is well known that if an individual is born without hearing, that person has a very difficult time learning how to speak intelligibly. On the other hand, if someone acquires speech normally and then becomes profoundly deaf postlingually, that person's speech can remain intelligible for decades without any useful hearing. However, the speech of such individuals does gradually develop some anomalies following hearing loss. A number of studies have been conducted on speakers who became deaf in adulthood, went without hearing for a number of years and then received a cochlear implant. Results have shown that phonemic goals are stable, but contrasts can gradually diminish without hearing. Restoration of some hearing with an implant usually results in parallel improvements in perception, measures of contrast in production and speech intelligibility (cf. Perkell et al., 2000; Vick et al, 2001).

In another approach, we have conducted studies of vowel and sibilant production and perception with 19 normal-hearing young adult speakers of American English. For two vowel contrasts and the sibilant (/s/-/ʃ/) contrast, we measured each speaker's degree of produced contrast and the speaker's auditory acuity for the contrast. Produced vowel contrast distances were measured in articulatory and formant (F1, F2) spaces and the produced sibilant contrast was measured as the difference in spectral means between /s/ and /ʃ/. Auditory acuity was measured as the subjects' ability to discriminate between pairs of natural-sounding synthetic stimuli along continua between each of the contrasting sounds. Both studies found that speakers with greater acuity produced the sounds with greater contrast. To interpret these results, we assume that spoken-language learners find it advantageous to be as intelligible as possible and therefore acquire auditory goal regions that are as distinct as possible. We reason that speakers who can perceive fine acoustic details will learn auditory goal regions that are smaller and spaced further apart than speakers with less acute perception, because the speakers with more acute perception are more likely to reject poorly produced tokens when acquiring the goals (Perkell et al., 2004a, 2004b).

A somatosensory goal and a saturation effect: The sibilant contrast

We have hypothesized that the sibilant sound /s/ has a somatosensory goal as well as an auditory one. The somatosensory goal is characterized by a saturation effect, which enhances the contrast of /s/ with its homologue, /ʃ/. As schematized in Fig. 2, /ʃ/ is produced by positioning the tongue blade so that there is a sublingual cavity. This cavity adds volume and complexity to the resonant cavity anterior to the constriction and thereby contributes to the lower spectral center of gravity of the frication noise. On the other hand, /s/ is produced by pressing the under-side of the tongue blade against the lower alveolar ridge and incisors, which eliminates the sublingual cavity and results in a smaller anterior resonator that contributes to a higher spectral center of gravity. When the tongue blade is moved forward to produce an /s/, once the sublingual cavity is eliminated, further contraction of the muscles that produce the forward movement will increase the contact pressure but will have a negligible effect on the size of the resonant cavity. Thus making this contact, which can be considered an somatosensory goal for the sound /s/, is characterized as a saturation effect. We also made measurements of the consistency of sublingual contact during /s/ production in the above-described perception/production study. The most distinct sibilant productions were made by subjects who used contact in producing /s/ but not /ʃ/, and had higher acuity. Subjects who did not use contact differentially and had lower acuity produced the least distinct contrasts. Intermediate degrees of contrast were found with subjects who used contact differentially or had higher acuity (Perkell et al, 2004b).

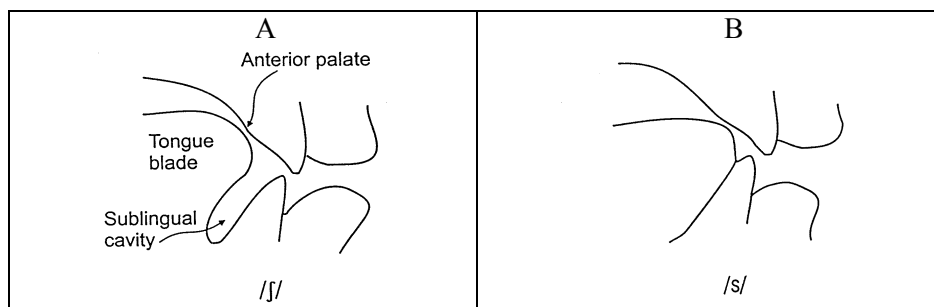


Figure 2. Schematic of tongue blade configurations for producing an /ʃ/ (A) and an /s/ (B). /ʃ/ is produced with a sublingual cavity, which contributes to the lower mean frequency of its acoustic spectrum; /s/ is produced with contact between the under side of the tongue blade and the lower incisors.

Feedback and feedforward control

To learn more about feedback and feedforward control mechanisms in speech, investigators have conducted a large number of studies in which auditory or somatosensory feedback (or both) have been perturbed and subjects' compensatory responses have been measured. Some of these studies have used steady-state perturbations, such as inserting a bite block between the teeth or blocking hearing with masking noise; others have used intermittent articulatory or auditory perturbations that the subjects cannot anticipate. Unanticipated perturbations of jaw movements, palatal shape, or auditory feedback have revealed that mechanisms are available that can detect and correct production errors within about 100 to 150 ms from the onset of the perturbation. Therefore, if a movement lasts long enough, somatosensory and auditory errors can be corrected during the movement itself by closed-loop feedback mechanisms. However, many articulatory movements in fluent speech do not last long enough to be corrected by feedback mechanisms. It follows that fluent adult speech production is controlled almost entirely by feedforward mechanisms, as in the DIVA model.

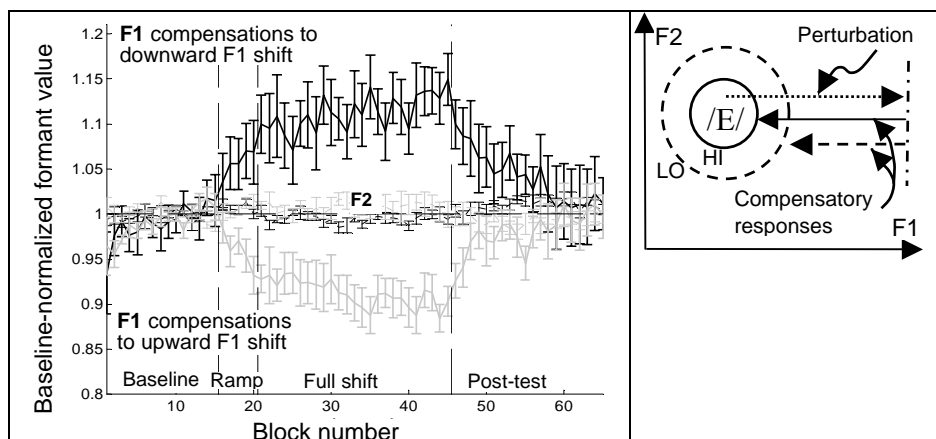


Figure 3. A: Compensatory responses to F1 shifts in normal-Baseline-normalized formant value of subjects' baseline-normalized F1 and F2 vs. block number. Each block contains one repetition of each of 18 different words in the corpus. The curves above baseline show the average of 10 subjects' productions in response to a downward shift of F1; the curves below baseline, the average of 10 subjects' responses to an upward F1 shift. B: Schematic of goal regions and compensatory responses for / ϵ / for a high-acuity speaker (solid circle) and a low-acuity speaker (dashed circle). F1 perturbation is indicated by the dotted arrow, and compensatory responses, by the solid and dashed arrows.

Figure 3A shows the results of an experiment that investigated feedforward control in 20 normal-hearing speakers. The subjects pronounced /CεC/ words while the first formant frequency (F1) of the vowel in their auditory feedback being was shifted in nearly real time (18 ms. delay), without their being aware of the shift (Villacorta et al., 2005). Ten of the subjects received upward shifts and the other 10, downward shifts. The plots show that the subjects partially compensated for the shifts over many trials by modifying their productions so that F1 moved in the direction opposite to the shift. The subjects' auditory acuity was also measured. There was a significant correlation between subjects' acuity and amount of compensation to F1 shift: speakers with better acuity tended to compensate more.

What underlies this correlation between acuity and compensation? Figure 3B schematizes how two speakers differing in acuity, and therefore in the sizes of their auditory goal regions for the vowel /ε/, might respond to a perturbation of F1. The high-acuity speaker has a smaller goal region. The perturbation of F1 is indicated by a dotted arrow pointing to the right, and the shifted value of F1, by a vertical broken line. This high-acuity speaker, in response to the shift in F1, will produce a greater compensatory response (middle arrow) than the one with lesser acuity. This is because the speaker continues to compensate until the F1 of his or her auditory feedback (which includes the shift) moves into the goal region. The distance between the shifted value of F1 (vertical line) and the edge of the goal region is greater for the high acuity speaker. In the DIVA model, auditory feedback provides closed-loop corrections of current motor commands and then modifications of feedforward commands for subsequent movements.

Summary

According to our theoretical overview and experimental results, the control variables for phonemic movements consist of auditory-temporal and somatosensory-temporal goal regions, which correspond to expected sensory consequences of producing speech sounds. Goal regions for languages in general and for individual speakers are determined by a number of factors, including quantal and saturation effects. Findings of motor-equivalent trading relations for the sounds /u/ and /r/ provide evidence that their goals are at least partly auditory. Auditory feedback is crucial for acquisition of phonemic goals, and it is needed to maintain appropriate motor commands with vocal-tract growth and perturbations. The goals are usually stable; however degree of contrast can diminish with prolonged postlingual hearing loss and increase with hearing restored by a cochlear implant. Findings that speakers with better acuity produce more distinct sound contrasts indicate that more acute speakers may learn smaller, more distinct goal regions.

Feedback and feedforward control operate simultaneously; however feedforward control predominates in fluent speech. Frequently used sounds (syllables and words) are encoded as feedforward commands. Feedback control intervenes when a perturbation produces a large enough mismatch between expected and produced sensory

consequences. In such cases if the movement lasts long enough, a correction is expressed during the movement itself, i.e., closed loop). Otherwise, the correction is incorporated into feedforward control of subsequent movements.

Since the DIVA model is formulated in terms of patterns of cortical connectivity and activity, it can be tested with brain imaging experiments. And, as reflected in the examples described above, it provides a valuable means of quantifying relations among phonemic specifications, brain activity, articulatory movements and the speech sound output.

Acknowledgements

The work from our laboratory that is described in this chapter was done in collaboration with a number of people, including Frank Guenther, Harlan Lane, Melanie Matthies, Mark Tiede, Majid Zandipour, Margaret Denny, Jennell Vick and Virgilio Villacorta. Support was from grants R01-DC001925 and R01-DC003007 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

References

- Fujimura, O. and Kakita, Y. 1979. Remarks on quantitative description of lingual articulation. In B. Lindblom and S. Öhman (eds.) *Frontiers of Speech Communication Research*, Academic Press.
- Guenther, F.H., Espy-Wilson, C., Boyce, S.E., Matthies, M.L., Zandipour, M. and Perkell, J.S. 1999. Articulatory tradeoffs reduce acoustic variability during American English /t/ production. *J. Acoust. Soc. Am.*, 105, 2854-2865.
- Guenther, F.H., Ghosh, S.S., and Tourville, J.A. 2006. Neural modelling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96, 280-301.
- Lindblom, B.E.F. 1990. Explaining phonetic variation: A sketch of the H&H theory. In W.J. Hardcastle and A. Marchal (Eds.), *Speech Production and Speech Modelling*. (pp. 403-439). Netherlands: Kluwer Academic Publishers.
- Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Perrier, P., Vick, J., Wilhelms-Tricarico, R. and Zandipour, M. 2000. A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *J. Phonetics* 28, 233-372.
- Perkell J.S., Guenther F.H., Lane, H., Matthies, M.L., Stockmann, E., Tiede, M. and Zandipour, M. 2004a. The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts, *J. Acoust. Soc. Am.* 116, 2338-44.
- Perkell, J.S., Matthies, M.L., Svirsky, M.A. and Jordan, M.I. 1993. Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot motor equivalence study, *J. Acoust. Soc. Am.* 93, 2948-2961.
- Perkell J.S., Matthies, M.L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E. and Guenther, F.H. 2004b. The Distinctness of Speakers' /s/-/ʃ/ Contrast is related to their auditory discrimination and use of an articulatory saturation effect, *JSLR* 47, 1259-69.
- Perkell, J.S. and Nelson, W.L. 1985. Variability in production of the vowels /i/ and /a/, *J. Acoust. Soc. Am.* 77, 1889-1895.
- Stevens, K. N. 1989. On the quantal nature of speech. *J. Phonetics* 17, 3-46.
- Stevens, K.N. 1998. *Acoustic Phonetics*, MIT Press, Cambridge, MA.
- Vick, J., Lane, H., Perkell, J.S., Matthies, M.L., Gould, J., and Zandipour, M. 2001. Speech perception, production and intelligibility improvements in vowel-pair contrasts in adults who receive cochlear implants. *J. Speech, Language and Hearing Res.* 44, 1257-68.
- Villacorta, V., Perkell, J.S., and Guenther, F.H. 2005. Relations between speech sensorimotor adaptation and perceptual acuity. *J. Acoust. Soc. Am.* 117, 2618-19 (A).

Phonological encoding in speech production

Niels O. Schiller

Department of Cognitive Neuroscience, Maastricht University, The Netherlands

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Leiden Institute for Brain and Cognition (LIBC), The Netherlands

Abstract

Language production comprises conceptual, grammatical, and word form encoding as well as articulation. This paper focuses on word form or phonological encoding. Phonological encoding in speech production can be subdivided into a number of sub-processes such as segmental, metrical, and syllabic encoding. Each of these processes is briefly described and illustrated with examples from my own research. Special attention is paid to time course issues introducing behavioural and electrophysiological research methods such as LRPs and ERPs. It is concluded that phonological encoding is an incremental planning process taking into account segmental, metrical, and syllabic encoding.

Models of spoken language production

Models of speech production (e.g., Caramazza, 1997; Dell, 1986, 1988; Fromkin, 1971; Garrett, 1975; Levelt, 1989; Levelt, Roelofs, and Meyer, 1999) assume that the generation of a spoken utterance involves several processes, such as conceptual preparation, lexical access, word form encoding, and articulation. Word form encoding or phonological encoding can be further divided into a number of processes (see Figure 1). Levelt et al. (1999) presented one of the most fine-grained models of phonological encoding to date (see also Dell, 1986, 1988). According to this model, phonological encoding can start after the word form (e.g., *table* /tEɪb↔l/) of a lexical item has been accessed in the mental lexicon. First, the phonological encoding system must retrieve the corresponding segments and the metrical frame of a word form. According to Levelt et al. (1999), segmental and metrical retrieval are assumed to run in parallel. During segmental retrieval the ordered set of segments (phonemes) of a word form are retrieved (e.g., /t/, /Eɪ/, /b/, /↔/, /l/), while during metrical retrieval the metrical frame of a word is retrieved, which consists at least of the number of syllables and the location of the lexical stress (e.g., for *Table* – capital letters mark stressed syllables – this would be a frame consisting of two syllables the first of which is stressed).

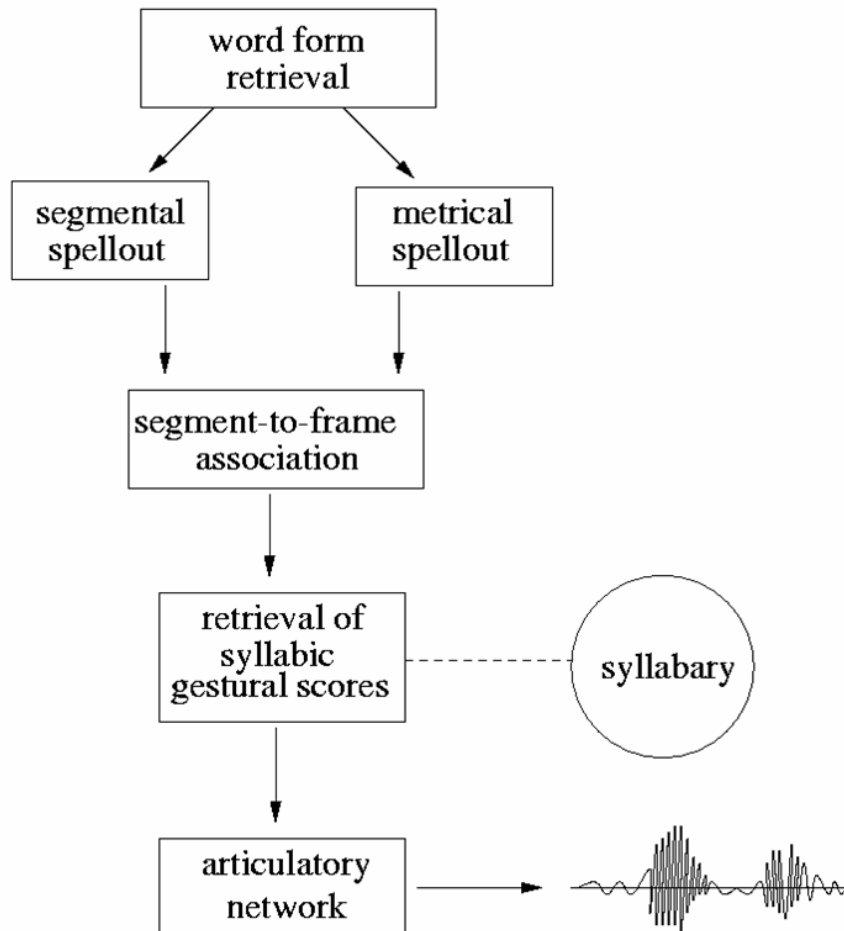


Figure 1. A model of phonological encoding in speech production (slightly adapted from Levelt and Wheeldon, 1994).

Then, during segment-to-frame association previously retrieved segments are combined with their metrical frame. The retrieved ordering of segments prevents them from being scrambled ($/t/_1$, $/EI/_2$, $/b/_3$, $/\leftrightarrow/_4$, $/l/_5$). They are inserted incrementally into slots made available by the metrical frame to build a so-called phonological word. This incremental syllabification process respects universal and language-specific syllabification rules, e.g. *TA.ble* (dots mark syllable boundaries). A phonological word is not necessarily identical to the syntactic word because some syntactic words such as pronouns or prepositions, which cannot bear stress themselves, cliticize onto

other words forming one phonological word together, e.g. *gave + it* /gEI.vIt/. Roelofs (1997) provided a computational model of this theory including a suspense/resume mechanism making initiation of encoding in the absence of complete information possible. For instance, segment-to-frame association can start before all segments have been selected, then be suspended until the remaining segments become available, and then the process can be resumed. Evidence for the incremental ordering during segmental encoding comes from a number of studies using different experimental paradigms (e.g., Meyer, 1990, 1991; Van Turenout, Hagoort, & Brown, 1997; Wheeldon & Levelt, 1995; Wheeldon & Morgan, 2002). Segment-to-frame association is the process that lends the necessary flexibility to the system depending on the speech context (Levelt et al., 1999). After the segments have been associated with the metrical frame, the resulting phonological syllables may be used to activate the corresponding phonetic syllables in a mental syllabary (Cholin, Levelt, & Schiller, 2006; Cholin, Schiller, & Levelt, 2004; Levelt & Wheeldon, 1994; Schiller, Meyer, Baayen, & Levelt, 1996; Schiller, Meyer, & Levelt, 1997). Once the syllabic gestural scores are made available, they can be translated into neuro-motor programs, which are used to control the movements of the articulators, and then be executed resulting in overt speech (Goldstein & Fowler, 2003; Guenther, 2003).

Segmental encoding

Word forms activate their segments and the rank order in which these segments have to be inserted into a phonological frame with slots for each segment (slot-filler-theory; Shattuck-Hufnagel, 1979 for an overview). Evidence for this hypothesis comes, for instance, from speech errors such as “**q**ueer old **d**ean” instead of “**d**ear old **q**ueen”, a spoonerism. These errors show that word forms are not retrieved as a whole, but rather they are computed segment by segment. Retrieving all segments separately and putting them together into word frames afterwards may seem more complicated than retrieving word forms as a whole. However, this mechanism has an important function when it comes to the production of more than one word. Usually, we do not speak in single, isolated words, but produce more than one word in a row. Let us take the above example *gave it* /gEI.vIt/. Whereas *gave* is a monosyllabic CVC word, the phrase *gave it* consists of a CV and a CVC syllable. That is, the syllable boundaries straddle word or lexical boundaries. In other words, the syllabification process does not respect lexical boundaries because the linguistic domain of syllabification is not the lexical word, but the phonological word (Booij, 1995). Depending on the phonological context in the phonological word, the syllabification of words may also change. Therefore, it does not make a lot of sense to store syllable boundaries with the word forms in the mental lexicon since syllable boundaries may

change during the speech production process as a function of the phonological context (Levelt & Schiller, 1998). Syllable boundaries will be generated on-line during the construction of phonological words to yield maximally pronounceable syllables. This architecture lends maximal flexibility to the speech production system in different phonological contexts.

Time course of segmental processing

One important question in word form encoding is the time course of the processes involved. For instance, are the segments of a word encoded one after the other or are they encoded in parallel? It was argued above on the basis of empirical evidence (e.g., sound errors) as well as on theoretical grounds that word forms are planned in terms of abstract units called segments or phonemes. Behavioural evidence for these claims has been provided in priming studies by Meyer (1990, 1991) and in self-monitoring studies by Wheeldon and Levelt (1995), Wheeldon and Morgan (2002), and Schiller (2005). For a summary of these studies see Schiller (2006).

However, there are also electrophysiological studies investigating the time course of segmental encoding. Van Turenout et al. (1997), for instance, investigated the time course of segmental encoding using lateralized readiness potentials (LRPs). The LRP is a derivative of the electroencephalogram (EEG) which can be measured by using scalp electrodes. Participants in Van Turenout et al.'s experiment named pictures on a computer screen, one at a time. Whenever a visual cue was presented, participants were requested to carry out a dual task (retrieve certain properties about the to-be-named word) and afterwards name the picture. For instance, participants were asked to make a decision about the animateness of the target concept and about the identity of the initial and final segment of the word. Interestingly, the onset of the nogo-LRP started to develop about 80 ms earlier when the segment was at the onset of words than when it was at the offset of words. This has been interpreted as reflecting the time course of the availability of phonological segments during phonological encoding in speech production planning.

The targets in the Van Turenout et al. (1997) study were 1.5 syllables long on average. Dividing 80 ms by 1.5 corresponds well to the 55 ms difference reported by Wheeldon and Levelt (1995) for the monitoring of syllable onset vs. offset phonemes. One may assume that phonological encoding of a whole syllable takes approximately 50 to 60 ms.

Metrical encoding

The above-mentioned studies investigating the time course of segmental encoding all have in common that they assume the measured effects to take

place at the level of the phonological word. This holds both for the priming studies by Meyer (1990, 1991) and for the monitoring studies by Wheeldon and Levelt (1995) as well as Van Turenout et al. (1997). However, it is unclear how the metrical stress of words is retrieved and encoded. Roelofs and Meyer (1998) found evidence that metrical stress of words is retrieved from the lexicon when it is in non-default position. However, Schiller, Fikkert, and Levelt (2004) could not find any evidence for stress priming in a series of picture-word interference experiments. Schiller et al. (2004) suggested that lexical stress may be computed according to language-specific linguistic rules (see also Fikkert, Levelt, & Schiller, 2005). Furthermore, lexical stress may be encoded incrementally – just like segments – or it may become available in parallel.

Time course of metrical processing

To investigate the time course of metrical processing, Schiller and colleagues employed a tacit naming task and asked participants to decide whether the bisyllabic name of a visually presented picture had initial or final stress. Their hypothesis was that if metrical encoding is a parallel process, then there should not be any differences between the decision latencies for initial and final stress. If, however, metrical encoding is also a rightward incremental process – just like segmental encoding –, then decisions to picture names with initial stress should be faster than decision latencies to picture names with final stress. The latter turned out to be the case (Schiller, Jansma, Peters, & Levelt, 2006). However, Dutch – like other Germanic languages – has a strong preference for initial stress. More than 90% of the words occurring in Dutch have stress on the first syllable. Therefore, this effect might have been due to a default strategy. However, when pictures with trisyllabic names were tested, participants were still faster to decide that a picture name had penultimate stress (e.g., *asPERge* 'asparagus') than that it had ultimate stress (e.g., *artiSJOK* 'artichoke'). This result suggests that metrical encoding proceeds from the beginning to the end of words, just like segmental encoding.

Recently, Schiller (in press) extended this research into the area of electrophysiology. Event-related brain potentials have the advantage of being able to determine processes more precisely in time, whereas behavioural studies such as reaction time studies can only measure the end of processes. In his study, Schiller (in press) used N200 effects to measure the availability of lexical stress in the time course of speech planning. He replicated the behavioural effect demonstrated by Schiller et al. (2006) and showed that the N200 peak latencies were significantly earlier when stress was on the first as compared to the second syllable. Furthermore, the N200 effects occurred in a

time window (400-500 ms) previously identified by Indefrey and Levelt (2004) for phonological encoding.

Syllabic encoding

We have already mentioned above that syllables are presumably created on the fly during speech production. There is quite some linguistic and psycholinguistic evidence (see Cholin et al., 2004 for a recent review and some new data) for the existence of syllables. However, in Levelt's model syllables form the link between the phonological planning process and the articulatory-motor execution of speech in a so-called *mental syllabary* (Levelt, 1989; Levelt et al., 1999). Such a mental syllabary is part of long-term memory comprising a store of syllable-sized motor programs. Ferrand and colleagues (1996, 1997) reported on-line data confirming the hypothesis about a mental syllabary, but Schiller (1998, 2000; see also Schiller, Costa, & Colomé, 2002 and Schiller & Costa, in press) disconfirmed this finding. Rather the results of these latter studies support the idea that syllables are not retrieved, but created on-line during phonological encoding.

The existence of the mental syllabary hinges on the existence of syllable frequency effects. Levelt and Wheeldon (1994) were the first to report effects of syllable frequency effects. However, segment frequency was not controlled well enough and therefore these results are not conclusive. Recently, Cholin et al. (2006) were able to demonstrate syllable frequency effects in very controlled set of materials. Following Schiller (1997), they used quadruples of CVC syllables controlling the segments in onset and offset position (e.g., HF *kem* – LF *kes* and HF *wes* – LF *wem*; HF = high frequency, LF = low frequency). In two experiments, Cholin et al. (2006) showed that HF syllables were named significantly faster than LF syllables. So far, this study includes the best controlled materials demonstrating a syllable frequency effect and hence evidence in favour of a mental syllabary, which may be accessed during phonological encoding.

References

- Booij, G. 1995. The phonology of Dutch. Oxford, Clarendon Press.
- Caramazza, A. 1997. How many levels of processing are there in lexical access? *Cognitive Neuropsychology* 14, 177-208.
- Cholin, J., Levelt, W. J. M., and Schiller, N. O. 2006. Effects of syllable frequency in speech production. *Cognition* 99, 205-235.
- Cholin, J., Schiller, N. O., and Levelt, W. J. M. 2004. The preparation of syllables in speech production. *Journal of Memory and Language* 50, 47-61.
- Dell, G. S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93, 283-321.

- Dell, G. S. 1988. The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language* 27, 124-142.
- Ferrand, L., Segui, J., and Grainger, J. 1996. Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language* 35, 708-723.
- Ferrand, L., Segui, J., and Humphreys, G. W. 1997. The syllable's role in word naming. *Memory & Cognition* 35, 458-470.
- Fikkert, P., Levelt, C. C., and Schiller, N. O. 2005. "Can we be faithful to stress?" Poster presented at the 2nd Old World Conference in Phonology (OCP2), 20-22 January 2005 in Trømsø (Norway).
- Fromkin, V. A. 1971. The non-anomalous nature of anomalous utterances. *Language* 47, 27-52.
- Garrett, M. F. 1975. The analysis of sentence production. In G. H. Bower (ed.) 1975, *The psychology of learning and motivation*, Vol. 9., 133-177. San Diego, CA, Academic Press.
- Goldstein, L., and Fowler, C. A. 2003. Articulatory Phonology: A phonology for public language use. In N. O. Schiller and A. S. Meyer (eds.) 2003, *Phonology and phonetics in language comprehension and production: Differences and similarities*, 159-207. Berlin: Mouton de Gruyter.
- Guenther, F. 2003. Neural control of speech movements. In N. O. Schiller and A. S. Meyer (eds.) 2003, *Phonology and phonetics in language comprehension and production: Differences and similarities*, 209-239. Berlin: Mouton de Gruyter.
- Indefrey, P., and Levelt, W. J. M. 2004. The spatial and temporal signatures of word production components. *Cognition* 92, 101-144.
- Levelt, W. J. M. 1989. *Speaking. From intention to articulation*. Cambridge, MA, MIT Press.
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22, 1-75.
- Levelt, W. J. M., and Schiller, N. O. 1998. Is the syllable frame stored? [commentary] *Behavioral and Brain Sciences* 21, 520.
- Levelt, W. J. M. and Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition* 50, 239-269.
- Meyer, A. S. 1990. The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language* 29, 524-545.
- Meyer, A. S. 1991. The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language* 30, 69-89.
- Roelofs, A. 1997. The WEAVER model of word-form encoding in speech production. *Cognition* 64, 249-284.
- Roelofs, A., and Meyer, A. S. 1998. Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24, 922-939.
- Schiller, N. O., Meyer, A. S., Baayen, R. H., and Levelt, W. J. M. 1996. A comparison of lexeme and speech syllables in Dutch. *Journal of Quantitative Linguistics* 3, 8-28.

- Schiller, N. O. 1997. Does syllable frequency affect production time in a delayed naming task? In G. Kokkinakis, N. Fakotakis, and E. Dermatas (eds.), *Proceedings of Eurospeech '97. ESCA 5th European Conference on Speech Communication and Technology*, 2119-2122. University of Patras, Greece, WCL.
- Schiller, N. O., Meyer, A. S., and Levelt, W. J. M. 1997. The syllabic structure of spoken words: Evidence from the syllabification of intervocalic consonants. *Language and Speech* 40, 103-140.
- Schiller, N. O. 1998. The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language* 39, 484-507.
- Schiller, N. O. 2000. Single word production in English: The role of subsyllabic units during phonological encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26, 512-528.
- Schiller, N. O., Costa, A., and Colomé, A. 2002. Phonological encoding of single words: In search of the lost syllable. In C. Gussenhoven and N. Warner (eds.) 2002, *Laboratory phonology* 7, 35-59. Berlin: Mouton de Gruyter.
- Schiller, N. O., Fikkert, P., and Levelt, C. C. 2004. Stress priming in picture naming: An SOA study. *Brain and Language* 90, 231-240.
- Schiller, N. O. 2005. Verbal self-monitoring. In A. Cutler (ed.) 2005, *Twenty-first century psycholinguistics: Four cornerstones*, 245-261. Mahwah, NJ, Lawrence Erlbaum Associates.
- Schiller, N. O. 2006. Phonology in the production of words. In K. Brown (ed.) 2006, *Encyclopedia of language and linguistics*, 545-553. Amsterdam et al., Elsevier.
- Schiller, N. O., Jansma, B. M., Peters, J., and Levelt, W. J. M. 2006. Monitoring metrical stress in polysyllabic words. *Language and Cognitive Processes* 21, 112-140.
- Schiller, N. O. in press. Lexical stress encoding in single word production estimated by event-related brain potentials. *Brain Research*.
- Schiller, N. O., and Costa, A. in press. The role of the syllable in phonological encoding: Evidence from masked priming? *The Mental Lexicon*.
- Shattuck-Hufnagel, S. 1979. Speech errors as evidence for a serial ordering mechanism in sentence production. In W. E. Cooper and E. C. T. Walker (eds.) 1979, *Sentence processing*, 295-342. New York, Halsted Press.
- Van Turenout, M., Hagoort, P., and Brown, C. M. 1997. Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23, 787-806.
- Wheeldon, L., and Levelt, W. J. M. 1995. Monitoring the time course of phonological encoding. *Journal of Memory and Language* 34, 311-334.
- Wheeldon, L., and Morgan, J. L. 2002. Phoneme monitoring in internal and external speech. *Language and Cognitive Processes* 17, 503-535.

Experiments in investigating sound symbolism and onomatopoeia

Åsa Abelin

Department of Linguistics, University of Göteborg, Sweden

Abstract

The area of sound symbolism and onomatopoeia is an interesting area for studying the production and interpretation of neologisms in language. One question is whether neologisms are created haphazardly or governed by rules. Another question is how this can be studied. Of the approximately 60 000 words in the Swedish lexicon 1 500 have been judged to be sound symbolic (Abelin 1999). These were analyzed in terms of phonesthemes, (i. e. sound symbolic morpheme strings) which were subjected to various experiments in order to evaluate their psychological reality in production and understanding. In test 1 nonsense words were constructed according to the results of the preliminary analysis of phonesthemes and then interpreted by subjects. In test 2 subjects were instructed to create new words for some given, sense related, domains. In test 3 subjects were to interpret these neologisms. Test 4 was a lexical decision experiment on onomatopoeic, sound symbolic and arbitrary words. The results of the first three tests show that the phonesthemes are productive, to different degrees, in both production and understanding. The results of the lexical decision test do not show perceptual productivity. The methods are presented and discussed.

Background

Onomatopoeic and sound symbolic neologisms are interesting insofar as they show productivity at the lexical level. They have a relation to the issues of phylogeny and ontogeny of language. Was onomatopoeia involved in the development of language and does it help the child to acquire language? Onomatopoeia and sound symbolism often have an iconic or indexical relation between expression and meaning (just as gestures, cf Arbib 2005). Most linguists who are specifically interested in the phenomenon of sound symbolism and who view it as an integral part of language, also regard it as productive. In traditional etymology, on the other hand, the explanation of new coinages is often just by analogy with *one* other word (which implies non-productivity). Rhodes (1994) discusses onomatopoeia and he distinguishes between wild and tame words, these being the ends of a scale. "At the extreme wild end the possibilities of the human vocal tract are utilized to their fullest to imitate sounds of other than human origin. At the tame end the imitated sound is simply approximated by an acoustically close phoneme or

phoneme combination.” Bolinger (1950) did an assonance-rime analysis of English monosyllables where the initial consonants constitute the assonance and the remainder of the syllable is the rime. He argues that assonance-rime analysis (of tame words) is morphology because assonances and rimes do not combine productively. That, however, does not mean that a construction is frozen. He introduces the term “active” for constructions that produce monosyllables continuously, at a slow rate.

The questions that were tested in the present experiments are 1) whether phonesthemes are productive, (also in the interpretation of neologisms) 2) whether some phonesthemes are more productive than others. The intermittent occurrence of new forms, which fit into a pattern – in speech, prose and fiction, especially in child literature, constitutes an argument for productivity. The opposite view would mean that new coinages would be phonetically and semantically haphazard. However, with that view, the fairly wide-spread and easy comprehension of new forms would be difficult to account for. When being presented with deliberately constructed nonsense words, listeners usually have no objections to or difficulties in assigning some interpretation to them.

Tests 1–3

Test 1 is a free choice test which goes from expression to meaning in order to test the understanding of presumptive sound symbolic clusters (based on the analysis in Abelin 1999), e.g. "What would be a good meaning for the word *fnotig*?" Test 2 is a free production test, which goes from meaning to expression, to test the production of sound symbolism, e.g. "Invent a short word for somebody who is stupid". Test 3 is a matching test between the meanings and the neologisms of test 3.

Results from test 1–3

Test 1: The forms and meanings that gave the highest number of expected results (according to the previous lexical analysis) were: pj– pejorative, skr– broken and skv– wetness. There was a difference in interpretability between the clusters and subjects also interpreted differently.

Test 2: The meanings that were rendered the best (according to the previous lexical analysis) were pejorative, bad mood and wetness. The meanings were encoded mostly in initial clusters. The less frequent semantic features (like dryness) produced more forms breaking phonotactic rules.

Test 3: The matching between the six meanings and columns of neologisms gave a 100% correct results.

Test 4

Test 4 is a lexical decision test (described in Abelin, 1996). The purpose was to find out how real onomatopoeic (A), sound symbolic (B) and arbitrary words (C) and constructed onomatopoeic (D), sound symbolic (E) and arbitrary (F) words behave in a lexical decision experiment. In previous lexical decision experiments one finding is that non-words are recognized more slowly than real words. This raises the question if non-words made up from clusters which are highly onomatopoeic or sound symbolic are recognized more slowly or more quickly than nonsense words constructed from sound combinations which are normally arbitrary. Another question is: Which onomatopoeic and sound symbolic non-words (i. e. words built from onomatopoeic or sound symbolic elements) are confused for "real words"? "Real words" are (in this experiment) words that are either found in a (modern) lexicon or judged by speakers to be lexicalized, i. e. not neologisms.

The research questions concerned whether:

1. Onomatopoeic and sound symbolic words will more often be responded to incorrectly as compared with arbitrary words.
2. These words will have longer reaction times than arbitrary words.
3. Non-words constructed from consonant clusters typical for onomatopoeic and sound symbolic words will be responded to more incorrectly than nonsense words constructed from arbitrary words.
4. These words will have longer reaction times than nonsense words constructed from arbitrary words.

Results from test 4

Subjects were fastest and most free from errors with the arbitrary words. They were slower with onomatopoeic and sound symbolic words (and made many more mistakes). They were slowest on non-words, but did less mistakes with non-words, as a whole. They made most mistakes with real sound symbolic words.

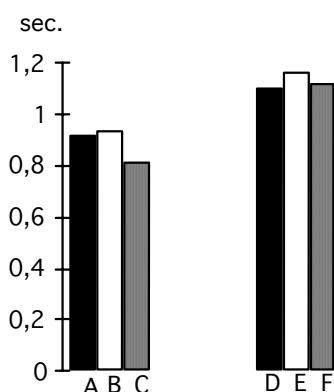


Figure 1. Mean length of reaction times for the different word groups. The differences between C and A, B are significant.

Somewhat surprising was that real onomatopoeic and sound symbolic words were judged as non-words more often than the corresponding non-words were judged as real ones. But – they were still significantly faster than the non-words. The intermediate speed gives them a status between arbitrary words and nonsense words implying an intermediate processing time.

Discussion of tests 1–4

The test 1–3 showed productivity for both production and perception, to different degrees for different phonesthemes. The results of test 4 are not in favour of perceptual productivity, since, instead of non-words modelled on sound symbolic phonesthemes being interpreted as real words, real sound symbolic word were often interpreted as non-words. These experiments are further developed for the study of neologisms.

References

- Abelin, Å. 1996. A lexical decision experiment with onomatopoeic, sound symbolic and arbitrary words. *TMH-QPSR* 2, 151–154, Department of Speech, Music and Hearing, Royal institute of Technology, Stockholm.
- Abelin, Å. 1999. *Studies in Sound Symbolism*. Göteborg, Göteborg monographs in linguistics 17
- Arbib, M. A. 2005. From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and brain sciences* 28, 105–167.
- Bolinger, D. 1950. Rime, assonance and morpheme analysis. *Word* 6. 117–136
- Rhodes, R. 1994. Aural images. In Hinton, L., Nichols, J., Ohala, J.J. (eds.) 1994, *Sound symbolism*, 276–292. Cambridge, Cambridge University Press.

Prosodic emphasis versus word order in Greek instructive texts

Christina Alexandris and Stavroula-Evita Fotinea
Institute for Language and Speech Processing (ILSP), Athens, Greece

Abstract

In Greek instructive texts, the production of clear and precise instructions is achieved by both prosodic modelling (Alexandris, Fotinea & Efthimiou, 2005) and morphosyntactic strategies (Alexandris, 2005). In the set of experiments performed in the present study, we attempt to determine whether prosodic modelling or whether morphosyntactic strategies, namely word order, constitutes the most crucial factor in the production of Greek spoken instructions most clearly and correctly understood by the receiver. In the present set of experiments, emphasis will be given in respect to utterances containing temporal expressions. The set of experiments performed in the present study involves the contrastive comparison and evaluation of 36 written and spoken utterances containing the temporal expressions, placed in different syntactic positions.

Word order in technical manuals and prosodic emphasis in task-oriented dialog systems

In the present study we attempt to determine whether the perceived prominence of elements in a sentence is primarily defined by prosody or by word order. The results of this experiment will contribute to the determination of whether prosodic modelling (1) or whether morphosyntactic strategies, namely word order (2), constitutes the most crucial factor in the production of Greek instructions most clearly and correctly understood by the receiver. The set of experiments performed will focus on utterances containing temporal expressions. Temporal and spatial expressions, as observed in recorded corpora of spoken Greek instructive texts, constitute the third largest group of elements receiving prosodic emphasis (after quantifiers and numerical expressions) and prosodic emphasis on temporal expressions is observed to produce a more precise and restrictive reading than when the same element in the same phrase is not emphasized (Alexandris, Fotinea & Efthimiou, 2005). In written Greek instructive texts such as technical manuals, usually involving a relatively high degree of Information Management (Hatim, 1997), ambiguities in respect to the semantics of spatial and temporal representation are solved with morphosyntactic strategies such as Rephrasing (Alexandris, 2005). Rephrasing is a controlled-language like approach (Lehrndorfer, 1996), where the key-word of the sentence, usually a negation,

a sublanguage-specific expression or a spatial and temporal expression, is positioned in the beginning of the sentence or phrase (Alexandris, 2005).

Word order versus prosodic emphasis in temporal expressions

Experiment

The set of experiments performed in the present study involves the evaluation of 18 written and 18 corresponding spoken utterances containing the temporal expressions, placed in different syntactic positions. The precise meaning of the spoken utterances containing elements of prosodic emphasis will be compared with the same set of corresponding written utterances. The written utterances were evaluated before the corresponding spoken utterances. The spoken utterances were evaluated in a separate session. The evaluation was performed by 30 native speakers (Standard Greek), both male (43,33%) and female (56,67%) (ages 25-45). In the first set, the respondents were to indicate the most important element of each sentence. The second set of sentences involved spoken sentences in which the respondents were requested to repeat exactly the same process. The sets of utterances were retrieved from an instructive text, namely a technical manual accompanying a coffee machine as a professional appliance. The corpus of written and spoken utterances consisted of permutations of simple and compound sentences containing temporal expressions placed in various grammatically acceptable and equally distributed positions before or after the verb, deverbal noun or entire sentence modified. Temporal expressions or phrases containing temporal expressions were placed in grammatically correct positions, and also in accordance to semantic acceptability, so as not to bias the respondents in their evaluation.

Results

In the set of spoken sentences, the results indicated that 93,7 % of the respondents selected the prosodically emphasized temporal expressions as the most important element in the sentence. In contrast to the spoken utterances, the results obtained from the set of written sentences presented very diverse and rather equally distributed results (Table 1). Specifically, in the set of written utterances, only 10% of the respondents marked all 18 phrases containing the temporal expressions that fully coincided with the prosodically emphasized temporal expressions in the set of spoken sentences. However, it should be noted that all respondents selected at least one written utterance with a temporal expression.

Table 1. Results for the Spoken and Written Utterances.

Percentages of respondents		
Number of utterances with Temporal Expression selected	Spoken Utterances	Written Utterances
18 out of 18	93,7%	0,0%
17 – 18 out of 18	0,0%	10,0%
14 -16 out of 18	0,0%	26,6%
11- 13 out of 18	0,0%	20,0%
5 – 10 out of 18	0,0%	13,3%
1-4 out of 18	6,3 %	20,0%

The second most commonly marked element in the set of written utterances is observed to be the subjunctive na-clause (Table 2), a complement clause with infinitival behaviour (Giannakidou, 2005). The respondents' selection of the na-clauses, occurring in 16 out of the 18 utterances, was equally distributed among the different syntactic structures of the utterances. Thus, we observed no evident relation between na-clause selection and the position of the na-clause in respect to the temporal expression or the phrase containing the temporal expression. A third group comprised elements containing nouns with qualitative modifiers such as “golden filter” and “medium grinded powder”. These sentences often occurred within the marked na-clauses, but were also often individually marked by some respondents. Further investigation is necessary to determine whether the criteria for the marking of the na-clauses are primarily syntactically or semantically based.

Table 2. An utterance with a na-clause (underlined).

Τώρα μπορείτε <u>να αφαιρέσετε</u>	την κανάτα από τη συσκευή
'Tora bo'rite <u>na afe'resete</u>	tin ka'nata a'po ti siske'vi
Now you-can na.-particle remove	the coffeepot from the appliance

Table 3. Results for the most commonly marked element.

Percentages of Written Utterances selected by Respondents		
Number of utterances	temporal expressions	na-clauses
All (18)	10,0%	0,0%
10 and over (10-17)	46,6,%	13,3%
5 and over (5 – 9)	33,3%	20,0%
Less than 5 (4- 1)	26,6%	30,0%
None (0)	0,0%	20,0%

The obtained data indicates that in written utterances, the class of elements considered to require special attention by the reader is perceived dif-

ferently among the respondents. The present data shows that temporal expressions do tend to be of semantic prominence to the receiver (Table 3), however, the data also shows that this phenomenon involving temporal expressions is not consistent and varies from respondent to respondent and other elements of the utterance may be considered semantically more important.

Conclusion

The present data indicates that prominence of elements in written utterances may be perceived differently among the respondents and further studies are necessary to investigate whether perceived prominence is syntactically or semantically determined. On the other hand, prosodic prominence is equally perceivable to most respondents since, in the spoken utterances, the temporal expressions, constituting the prosodically emphasized elements, were considered by almost all respondents to constitute the semantically most important element of the sentence. Therefore, according to the data obtained from the present study in Greek, prosodic modelling (1) constitutes the most crucial factor in the production of instructions most clearly and correctly understood by the receiver. In contrast, morphosyntactic strategies, namely word order (2) play a secondary role. The relation between syntax and semantics in respect to the determination of the most important element in written utterances constitutes an area of further research.

References

- Alexandris, C. 2005. English as an intervening language in manuals of Asian industrial products: Linguistic Strategies in technical translation for less-used European languages. In Proceedings of the Japanese Society for Language Sciences - JSLs 2005, Tokyo, Japan, 91-94.
- Alexandris, C., Fotinea, S.-E., Efthimiou, E. 2005. Emphasis as an Extra-Linguistic Marker for Resolving Spatial and Temporal Ambiguities in Machine Translation for a Speech-to-Speech System involving Greek. In Proceedings of HCII 2005, Las Vegas USA.
- Giannakidou, A. 2005. N-words and the Negative Concord. In M. Everaert, H. Van Riemsdijk, R. Goedemans and B. Hollebrandse (eds), *The Blackwell Companion to Syntax*, Vol III, Oxford: Blackwell.
- Lehrndorfer A. 1996. *Kontrolliertes Deutsch: Linguistische und Sprachpsychologische Leitlinien fuer eine (maschiell) kontrollierte Sprache in der technischen Dokumentation*, Tuebingen : Narr.
- Hatim, B. 1997. *Communication Across Cultures: Translation Theory and Contrastive Text Linguistics*, University of Exeter Press.

Gradience and parametric variation

Theodora Alexopoulou and Frank Keller
Savoirs, Textes, Langage, Lille III/RCEAL Cambridge
Informatics, University of Edinburgh

Abstract

The paper assesses the consequences of gradience for approaches to variation based on the Principles and Parameters model. In particular, the discussion focuses on recent crosslinguistic results obtained through magnitude estimation, a methodology particularly suited to the study of gradient acceptability/grammaticality. Results on superiority and relativised minimality effects in questions are discussed in the light of current theoretical assumptions regarding the locus of crosslinguistic variation.

Introduction

Gradient grammaticality has received attention in recent years mainly due to a recent experimental methodology, *magnitude estimation* (Bard, Robertson and Sorace 1996, Cowart 1997) that allows the elicitation of reliable gradient acceptability judgements. The application of this methodology to crosslinguistic variation has revealed very interesting results, but also an important challenge for a parametric approach to variation, namely that, often, variation is confined to quantitative differences in the magnitude of otherwise identical principles. Here we approach the issue with particular reference to crosslinguistic studies focusing on superiority and locality violations involved in *whether*-islands.

Basic findings on superiority and relativised minimality

Below we summarise the results of Featherston (2004) and Meyer (2003); the former is a comparative study of superiority effects and d-linking in English and German; the latter is a comparative study between Russian, Polish and Czech. The main findings of these studies indicate:

- i) A clear (statistically significant) dispreference for in-situ subjects (English, German, Russian, Polish, Czech, modulo a “reverse animacy” effect in Polish).
- ii) A clear crosslinguistic effect of discourse-linking, where in-situ d-linked subjects are essentially as acceptable as other in-situ phrases.
- iii) Crosslinguistically, the d-linking status of the object is irrelevant (English, German, Polish, Czech).

- iv) No clear interactions between arguments and adjuncts are detected (English, German, Polish, Czech, Russian).
- v) Not only in-situ subjects are dispreferred, but initial subjects are also preferred (marginal effect in German, quite significant in English).

Crosslinguistic variation is confined to quantitative differences in otherwise crosslinguistically stable preferences. For example, while initial subjects are clearly preferred in English, only a marginal preference was detected in German.

A similar picture emerges from the studies of Alexopoulou and Keller (2002, 2003, to appear); these studies investigate the effect of embedding (*that*-clauses), weak islands (*whether*-clauses) and strong islands (relative clauses) in questions and its interaction with the acceptability of resumption, in Greek, English and German. The main findings of these studies are summarised below.

- i) A clear crosslinguistic effect of embedding under a *that*-clause; that is, questions extracted from a *that*-clause were less acceptable than non-embedded questions.
- ii) A clear crosslinguistic effect of weak island violations, i.e. questions extracted from *whether*-clauses were significantly less acceptable than non-embedded questions. This effect, though stronger in magnitude, was similar in nature to the effect of embedding induced by *that*-clauses.
- iii) A strong contrast between weak and strong island violations, in that, questions extracted from relative clauses induced a severe drop in acceptability; in all three languages questions violating strong islands were much worse than questions violating weak islands.
- iv) Resumption is unacceptable in unembedded questions in all three languages.
- v) The acceptability of pronominals improves when they are embedded in a *that*-clause and a *whether*-clause, but not in a relative clause. Thus, there is no interaction between resumption and strong islands.

Again, crosslinguistic variation is confined to quantitative variation in the magnitude of the effects in question. For example, resumption in questions is more acceptable in Greek than in German and English; for instance, Greek unembedded questions with pronominals, though significantly worse than corresponding questions with gaps, are more acceptable than questions extracted from relative clauses (with or without pronominals). By contrast, in English and German, unembedded questions with pronominals are as bad as questions extracted from relative clauses. Further, in German, questions embedded under *dass* are almost as bad as questions embedded under a weak island (*whether*-clause), while in English and Greek questions embed-

ded in *that*-clauses are significantly better than questions embedded in weak islands.

Quantitative variation and parameters

The most important aspect of these studies is that they indicate that effects relating to superiority and relativised minimality are present crosslinguistically. In this way, these studies confirm the existence of some universal constraints where their status has either been disputed (e.g. superiority in German, Polish, Czech and Russian, whether-islands in Greek and German) or where their existence was not properly acknowledged (e.g. the fact that resumption improves weak islands in English even though such resumptives are less acceptable than gaps).

The main question we address is whether this quantitative variation should be taken at face value and modelled as such or reduced to structural differences between the languages in question, attributable to parametric variation. The former approach is advocated by Stochastic OT analyses, while, the latter, is consistent with a modular view of grammar as conceived by standard generative grammar.

We argue that, rather than taken at face value, quantitative variation is an epiphenomenon of structural variation in the languages in question. However, at the same time, we argue that quantitative variation cannot be discarded as surface "noise", of no theoretical importance, since some differences between languages are shown to be a consequence of such quantitative differences. For instance, we argue that the higher acceptability of pronominals in Greek questions is related to the availability of Clitic Left Dislocation in Greek and its absence from Greek and German. Thus, unembedded questions with pronominals are instances of CLLD where the requirement that the dislocated DP is referential/specific is violated. Such violations though, are of a semantic nature involving soft constraints (see Sorace and Keller 2004, Keller 2000) and induce milder unacceptability. By contrast, in English and German, in absence of CLLD, questions with pronominals involve violation of a hard, syntactic constraint (blocking pronominals in questions) that gives rise to strong unacceptability. This structural difference between Greek and English is indirectly responsible for the surface fact that embedded pronominals in Greek are as acceptable as gaps but less acceptable than gaps in English. Though in both languages the acceptability of embedded pronominals improves, in Greek pronominals in questions are generally more acceptable than in English, and, thus, "closer" to the acceptability of gaps.

We further argue that locality principles underlying superiority and weak-island effects are also related to soft constraints. They only induce

mild ungrammaticality which may be further improved by interaction with d-linking (Featherston 2004 has demonstrated experimentally the effect of d-linking on superiority violations for English and German). Further, such constraints have been argued to operate at the interface between grammar and the human sentence processor (Alexopoulou and Keller, to appear).

The consequences of this approach is the hypothesis that universal principles are generally subject to quantitative variation across languages (indirectly reducible to parametric variation) and involve interface principles, while categorical judgements are associated with parameter settings involving core grammatical phenomena. We will discuss this hypothesis with reference to further evidence from magnitude estimation studies from the domain of information structure and lexical semantics.

Acknowledgements

We are grateful to David Adger, Kook-hee Gill, John Hawkins, Napoleon Katsos, Dimitra Kolliakou, Ian Roberts, Christina Sevdali and George Tsoulas.

References

- Alexopoulou Th. And Keller, F. to appear. Locality, Cyclicity and Resumption: at the interface between grammar and the human sentence processor, *Language*.
- Bard E.G., Robertson D and Sorace, A, Magnitude Estimation for linguistic acceptability, *Language* 72(1).32-68.
- Cowart W, 1997. *Experimental Syntax, Applying objective methods to sentence judgements*, Thousand Oaks, CA: Sage Publications.
- Featherston F, 2004. Magnitude Estimation and what it can do for your syntax: Some wh-constraints in German. *Lingua*.
- Keller, F. 2000. *Gradience in grammar: experimental and computational aspects of degrees of grammaticality*, Ph.D thesis, University of Edinburgh.
- Meyer, R. 2003, Superiority effects in Russian, Polish and Czech: comparative evidence from studies on linguistic acceptability. In *Proceedings of the 12th Conference on Formal Approaches to Slavic Linguistics*, Ottawa, Canada.
- Sorace A and Keller, F. 2004, Gradience in Linguistic Data, *Lingua*.

Stress and accent: acoustic correlates of metrical prominence in Catalan

Lluïsa Astruc¹ and Pilar Prieto²

¹ Associate Lecturer, Faculty of Education and Languages,
The Open University, UK

² ICREA-Universitat Autònoma de Barcelona, Spain

Abstract

This study examines the phonetic correlates of stress and accent in Catalan, analyzing syllable duration, spectral balance, vowel quality, and overall intensity in two stress [stressed, unstressed] and in two accent conditions [accented, unaccented]. Catalan reveals systematic phonetic differences between accent and stress, consistent with previous work on Dutch, English, and Spanish (Slujter & van Heuven 1996a, 1996b; Campbell & Beckman 1997, Ortega-Llebaria & Prieto 2006). Duration, spectral balance, and vowel quality are reliable acoustic correlates of stress, while accent is acoustically marked by overall intensity and pitch. Duration, at least in Catalan, is not a reliable indicator of accent since accentual lengthening was found only in speakers who produced some accents with a wider pitch range.

Introduction

The search for consistent acoustic correlates of metrical prominence is complicated by the fact that stress and accent interact, since only stressed syllables can be accented. Some studies claim that stress does not have any phonetic reality and that only knowledge of the language allows listeners to distinguish minimal pairs such as 'pérmit' and 'permit'. According to this view (Bolinger 1958, Fry 1958), the main correlate of stress is pitch movement and, in the absence of pitch, nothing in the speech signal indicates where stress is. According to the alternative view (Halliday 1967, Vanderslice & Ladefoged 1972), metrical prominence consists of two categories with two conditions each, which ranked from lower to higher yield the following hierarchy: [-stressed, -accented] > [+stressed, -accented] > [+stressed, +accented]. Stress would then have separate phonetic correlates, although they strongly interact with those of accent. Recent experimental work on stress and accent has had contradictory results. Slujter & van Heuven (1996a, 1996b) modelled metrical prominence as a two-dimensional scale with two categories in each dimension (accent and stress). They found that differences in duration (stressed syllables are longer) and in spectral balance (stressed syllables show an increase in intensity that affects the higher regions of the spectrum), were strong correlates of stress, while overall

intensity was a cue of accent rather than of stress. Their results were confirmed in American and British English (Turk & Sawusch 1997, Turk & White 1999), and in Spanish (Ortega-Llebaria & Prieto 2006). However, Turk and collaborators (1997, 1999) also found that duration interacted strongly with accent. On the other hand, Beckman & Campbell (1997) modelled prominence as a one-dimensional scale with three categories: stressed-accented, stressed, and unstressed. They did not find consistent phonetic correlates of stress in American English. They concluded that the apparent phonetic correlates of stress were only a side-effect of vowel reduction and when full vowels are examined, no correlates of stress are found. Our research question is whether different levels of prominence are indeed cued by a separate set of phonetic correlates in Catalan, a weakly stressed-timed language with lexical stress and phonemic vowel reduction as Dutch and English.

Methodology

The corpus is formed by 576 target sentences, read by six female native speakers of Central Catalan. The experimental design has four experimental conditions: [+accent, +stress], [+accent, -stress], [-accent, +stress], and [-accent, -stress]. We have three vowels, two unreduced vowels, [u] and [i], and [a], reduced in unstressed position. Eight minimal pairs with CVCV structure and with ultimate and penultimate stress (Mimi-Mimí, Lulu-Lulú, mama-mamà, Mila-Milà, Milu-Milú, Vila-Vilà, mula-muler, Mula-Mulà) provide the stress conditions. The accent conditions are provided by minimal pairs of appositive and right-dislocated noun phrases (described respectively as accented and deaccented. See Astruc 2005, for a review). The intended interpretation (apposition or right-dislocation) is elicited with a question. Target syllables are word-initial in segmentally identical words in postfocal contexts, which allow us to control for position effects, for polysyllabic shortening, and for focal lengthening. Table 1 shows the four experimental conditions.

Table 1. Target syllable *mi* (in bold) in four accent and stress conditions

	[+ accent] apposition	[-accent] right-dislocation
[+stress]	M'agrada la protagonista, la Mimi 'I like the protagonist, Mimi'	Vol ser la protagonista, la Mimi 'She wants to be the protagonist, Mimi'
[-stress]	M'agrada la protagonista, la Mimí	Vol ser la protagonista, la Mimí

Procedure

Six female native speakers of Central Catalan were recorded at 44.1 kHz directly onto a computer in a studio. They were instructed to read the target sentences naturally at an average voice level using a Shure SM10A head-worn microphone to keep constant mouth-microphone distance. Some target utterances did not receive the intended interpretation and they had to be repeated. Some speakers produced some pitch accents in a wide pitch range. Acoustic and instrumental analysis were performed using Praat (4.3.09). Segmentation and labelling were done by hand, marking CV boundaries and the highest and lowest F0 point in both vowels. Measurements of duration (ms), pitch (Hz), frequency of the formants (F1, F2, F3, in Hz), spectral balance (in four bands: B1: 0-500Hz, B2: 500-1000Hz, B3: 1000-2000Hz, B4: 2000-4000Hz), and intensity (dB) were taken automatically at the peak of intensity of both syllables.

Results

The experimental paradigm worked well: appositions were consistently accented and right-dislocations were consistently deaccented. A one-way ANOVA ($F(1)=147.534$; $p<.05$) confirmed significant effects of accent on pitch range. Figure 1 shows mean results for pitch, duration, intensity, vowel quality, and spectral balance of the target syllable.

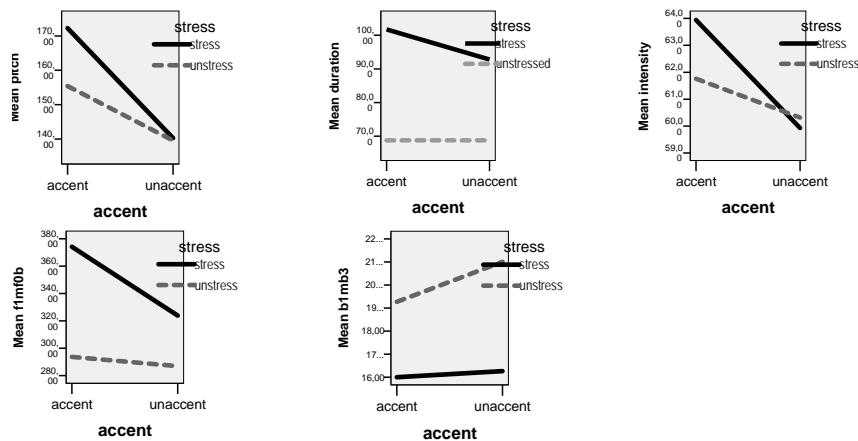


Figure 1. First row: pitch, duration, and intensity of V1 for all speakers. Second row: vowel quality and spectral balance of [a] for all speakers.

Repeated measures ANOVAs show significant effects of stress on the vowel quality of [a] ($F(1,5)=12.003$, $p<.05$, partial Eta squared=.706), and on the spectral balance (B3-B1) of [a] ($F(1,5)=7.756$,

$p < .05$, partial Eta squared = .608) and [u] ($F(1,5) = 23.039$, $p < .05$, partial Eta squared = .882), and on the duration (multivariate $F(1,5) = 265.77$, $p < .05$, partial Eta squared = .982) of all three vowels. Accent also has a strong effect on duration, but less stronger than that of stress (multivariate $F(1,5) = 11.320$, $p < .05$, partial Eta squared = .694). A speaker-by-speaker analysis reveals that only half of the speakers showed accentual lengthening, and these were the speakers who also used pitch accents with wider pitch excursions. Syllables were plausibly lengthened to accommodate these wider pitch movements. In conclusion, we controlled for focal and positional effects and for vowel reduction and found that stress differences are cued by systematic acoustic correlates, thus supporting the hypothesis that different levels of prominence are signalled by separate sets of acoustic cues.

Acknowledgements

Thanks to our informants (A. Abella, E. Bonet, T. Cabré, A. Gavarró, M. Mata, M. Llinàs, P. Prieto) and to E. Ferragne and M. Ortega-Llebaria for their Praat scripts.

References

- Astruc, L. 2004. The intonation of extra-sentential elements. Doctoral dissertation. University of Cambridge. Available from www.astruc.info.
- Bolinger, D.L. 1958. A theory of pitch accent in English. *Word* 14, 109-149.
- Campbell, N. and Beckman, M. 1997. Stress, prominence and spectral tilt. In Botinis, A.; Kouroupetroglou, G; and G. Crayannis (eds.), *Intonation: theory, models and applications*. Proc. of and ESCA Workshop, Athens, Greece.
- Fry, D.B. 1958. Experiments in the perception of stress. *Language and Speech* 1, 126-152.
- Halliday, M.A.K. 1967. *Intonation and grammar in British English*. Mouton.
- Sluijter, A. and van Heuven, V. 1996a. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100 (4). 2471-2485.
- Sluijter, A. and van Heuven, V. 1996b. Acoustic correlates of linguistic stress and accent in Dutch and American English. Proc. of ICSL, 96. Philadelphia: Applied Science and Engineering Laboratories, Alfred I duPont Institute, 630-633.
- Turk, A. E. and Sawush, J. R. 1997. The domain of accentual lengthening in American English. *Journal of Phonetics* 25, 25-41.
- Turk, A. E. and White, L. 1999. Structural influences on accentual lengthening. *Journal of Phonetics* 27, 171-206.
- Ortega-Llebaria, M. and Prieto, P. 2006. Stress and accent in Catalan and Spanish: patterns of duration, vowel quality, overall intensity, and spectral balance. Proc. *Speech Prosody*. Dresden, Germany. 337-340.
- Vanderslice, R. and Ladefoged, P. 1972. Binary suprasegmental features and transformational word-accentuation rules. *Language*. 48: 819-38.

Word etymology in monolingual and bilingual dictionaries: lexicographers' versus EFL learners' perspectives

Zahra Awad

Department of Linguistics and Phonetics, University of Jordan, Jordan

Abstract

This paper deals with the treatment of word etymology in monolingual and bilingual dictionaries. It also investigates EFL learners' attitudes towards the importance of etymology for understanding the meaning of the words they look up in dictionaries. The data were collected through tasks of looking up Arabic loan words in English in monolingual and bilingual dictionaries. The results indicate that most monolingual and bilingual dictionaries do not provide information on word etymology. They also show that most Arab EFL learners find etymology helpful in understanding the meaning of English words of Arabic origin.

Introduction

Since the late 17th century English general-purpose monolingual dictionaries have included information about the etymology of words (Stockwell & Minkova, 2001). The origin of many English words is indicated as being French, Latin, Greek, German, Hindi and Arabic. There are perhaps as many as 10,000 English words derived from Arabic.

This study is an investigation of how English words of Arabic origin are dealt with in English monolingual dictionaries and English-Arabic bilingual dictionaries. It is also an attempt to find out how important EFL learners consider knowing the origin of such words in understanding their meaning.

Methodology

The study is based on the treatment of 10 Arabic loan words in English in 4 monolingual and 3 bilingual dictionaries. They were chosen from a list of such words given in Al-Mawrid Dictionary (1997). Their etymology was verified in the Dictionary of Words Origin (1975). The words were: masage, algebra, candy, chap, garble, shifty, syrup, safari, cumin, tariff.

The monolingual dictionaries used were: Oxford Advanced Learner's Dictionary (OALD) (1989), Longman Dictionary of Contemporary English (LDOCE) (1987), Merriam-Webster's Dictionary (WD) (1997) and the American Heritage Dictionary (AHD) (1991). The bilingual dictionaries

used were: Al Mawrid Dictionary (1997) Dictionary, Atlas Encyclopaedic Dictionary (2005), and Al Mughni Al-Akbar Dictionary (1991).

The study is also based on analyzing Arab EFL learners' performance in two writing tasks and their responses to a questionnaire and in structured interviews. These learners were 200 second year students majoring in Applied English at the University of Jordan. They were taking a course on study skills which deals partly with using dictionaries.

The students were asked to look up the meaning of the target words in a monolingual dictionary, then in a bilingual one indicating their etymology if given and mention if that helped them in understanding the meaning of these words. After submitting each task, they were interviewed to verify their answers.

The questionnaire consisted of twelve questions dealing with the students' background information and their use of dictionaries. It was administered to the students after they performed the required tasks.

Results and discussion

The look ups of the target words indicated variations in the etymological information provided in the monolingual dictionaries as shown in Table I.

Table I. Availability of etymology in monolingual dictionaries.

No.	Word	OALD	LDOCE	AHD	WD
1	Algebra	X	X	√	√
2	Candy	X	X	√	√
3	Chap	X	X	X	X
4	Cumin	X	X	X	X
5	Garble	X	X	√	√
6	Massage	X	X	√	√
7	Safari	X	X	√	√
8	Shifty	X	X	X	X
9	Syrup	X	X	√	√
10	Tariff	X	X	√	√

It is clear from Table I that WD and AHD provide the etymology for most of the words, while OALD and LDC do not.

The look ups of the words also indicated variations in the etymological information provided in the bilingual dictionaries, as shown in Table II.

Table II. Availability of etymology in bilingual dictionaries.

No.	Word	Atlas	Al-Mughni	Al-Mawrid
1	Algebra	X	X	√
2	Candy	X	X	√
3	Chap	X	X	X
4	Cumin	X	X	X
5	Garble	X	X	√
6	Massage	X	X	√
7	Safari	X	X	√
8	Shifty	X	X	X
9	Syrup	X	X	√
10	Tariff	X	X	√

It is obvious from Table II that Al-Mawrid provides the etymology for most of the words, while Atlas and Al-Mughni do not.

The analysis of the students' responses indicated variations in the students' attitude towards the importance of knowing the Arabic origin of the target words to understand their meaning, as illustrated in Table III below.

Table III: Students' attitudes towards word origin

Easier with their origin	Percent
Yes, always	20.7
Yes, sometimes	41.4
Yes, most of the time	20.7
No	13.8
No answer	3.4
Total	100.0

Table III shows that most students considered knowing the Arabic origin of the target words important in order to understand them. For example, 20.7% of them considered that always important, 20.7 % thought it was important most of the time, and 41.4% considered it sometimes important.

Conclusion

The results indicated that both English monolingual dictionaries and English-Arabic bilingual dictionaries vary in providing etymology information on Arabic loan words in English. They also show that most EFL dictionary users tend to consider knowing the etymology of such words important for understanding their meaning. It is recommended that compilers of both bilingual and monolingual dictionaries consider including the origin of Arabic loan words in English in their entries.

References

- Abbas, S. (ed.) 2005. Atlas Encyclopaedic Dictionary. Amman, Atlas Global Center for Publishing and Distributing.
- Al-Ba'alabki, M. 1998. Al-Mawrid : A Modern English – Arabic Dictionary. Beirut, Dar EL- Ilm Lil – Malayen.
- Cowie, A.P. (ed.). 1989. Oxford Advanced Learner's Dictionary of Current English. Oxford: Oxford University Press.
- De Vinne, P. B. 1991. The American Heritage Dictionary. Boston, Houghton Mifflin Company.
- Karmi, H. 1991. Al-Mughni Al-Akbar (A dictionary of contemporary English). Beirut, Libraie du Liban.
- Shipley, J. T. 1975. Dictionary of Words Origin. London: Philosophical library. URL [http:// www.questia.com /](http://www.questia.com/)
- Stevens, Mark A. 1997. Merriam Webster's Collegiate Dictionary. Springfield : Merriam Webster, Incorporated.
- Stockwell, R. and D. Minkova. 2001. English Words: History and Structure. Cambridge University Press.
- Summers, D (ed.). 1987. Longman Dictionary of Contemporary English. England: Longman Group.

Characteristics of pre-nuclear pitch accents in statements and yes-no questions in Greek

Mary Baltazani

Department of Linguistics, University of Ioannina, Greece

Abstract

In this paper I present the results of a production experiment testing the hypothesis that L*+H pre-nuclear pitch accents are indistinguishable in statements and questions in Greek. Results show that the L and the H tones of the L*+H pitch accent have different patterns of alignment in polar and in statements. These results suggest that the pitch accents are realized differently in the two utterance types under investigation. It remains to be explored, through perception experiments, whether the phonetic differences that were found are salient enough for listeners to distinguish between statements and questions. If they are, then the prenuclear pitch accents in statements and those in polar questions belong to different phonological categories.

Introduction

Analyses of the intonation system of Greek maintain that the most frequently used pre-nuclear pitch accent is L*+H across all utterance types, statements, questions, negatives, and imperatives (Arvaniti, Ladd and Mennen 1998; Baltazani 2002; Arvaniti and Baltazani 2005).

This state of affairs might give rise to processing problems on the part of the listener for the following reason: In Greek, statements and yes-no questions can be string identical, differing only in the type of nuclear pitch accent and boundary tones, according to the analyses cited above. Therefore, we can make the hypothesis that in utterances with a number of pre-nuclear pitch accents preceding a late nucleus, listeners will have to wait for the nucleus to be uttered before they can determine whether the utterance they are processing is a statement or a question. This hypothesis has not been tested yet, to my knowledge.

Experiment

To test the hypothesis that pre-nuclear pitch accents are indistinguishable in statements and questions, both production and perception tests were designed. In this paper I present the results of a production experiment.

Method

The same matrix sentence was used for the two utterance types (statement, yes-no question). Each type was preceded by a different context, to elicit the production of the desired melody. The utterances were designed to have three pitch accents, two pre-nuclear ones followed by the nuclear one. Eight speakers participated in this experiment, 3 male and 5 female, producing a corpus of 128 utterances (8 speakers X 2 matrix sentences X 2 types X 4 repetitions per speaker).

The tones of the first bi-tonal accent were labelled L1 and H1. The analysis showed a low plateau between the first and the second pitch accent, so two low points were measured for the second pitch accent, L2 and L3, at the two edges of the low plateau. Similarly, there was a high plateau during the second pitch accent, and its two edges were labelled H2 and H3. Figure 1 illustrates the four measuring points of the second pitch accent in a polar question as uttered by one of the female speakers. These points were found in both statements and polar questions.

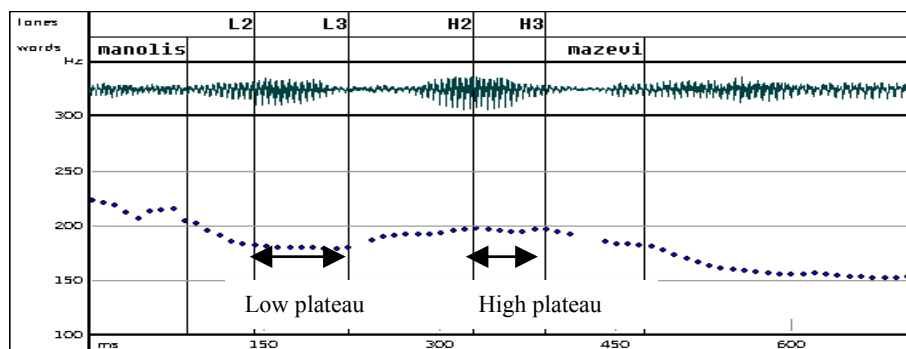


Figure 1. The four measuring points, L2, L3, H2, H3, of the second L*+H pitch accent.

The following measurements were taken. First, F0 of the tones composing the bi-tonal accent for scaling differences in statements and polars. Second, for alignment differences, I took two measurements: a) the distance in time between consecutive tones; b) the distance in time between tonal targets and segmental landmarks. In particular: b1) L1 to consonant onset of the first stressed syllable; b2) L2 to consonant onset of the second stressed syllable; b3) L3 to consonant onset of the second stressed syllable; b4) H1 to vowel offset of the first stressed syllable; b4) H2 to vowel offset of the second stressed syllable.

No statistical analysis was carried out since the criterion for significance of the results is whether or not pitch accents in the two sentence types are perceived as different by listeners.

Results

The realization of the Greek L*+H has been described as a gradual rise from a trough (the L tone) to a peak (the H tone). In general, the L is aligned at the very beginning or slightly before the onset of the stressed syllable, and the H early in the first post-stress vowel (Arvaniti, Ladd and Mennen 1998; Baltazani 2002; Arvaniti and Baltazani 2005).

The results of the present experiment show that there is a difference in the realization of the bitonal L*+H pitch accent in neutral statements and polar questions. Differences in scaling were not consistent across speakers; however, systematic differences were found in alignment. The scaling differences are presented in the first subsection below, followed by the alignment differences in the second subsection.

Scaling

Some scaling differences were found, but they depended on gender (Table 1). For male speakers, the L and the H tones are realized higher in statements than in polars. For female speakers, there is no difference for the L tones across sentence types; the H tones are realized higher in polars, that is, the reverse trend from what was found for males. No differences were found between statements and polar questions in the pitch range (that is, the difference in Hz between the L and the H).

Table 1. Average F0 values of L and H tones of the L*+H pitch accents for male and female speakers in statements and polar questions. Standard deviation is given in parentheses.

	L tones		H tones	
	statement	polar	statement	polar
Males	114 (8.6)	105 (4.5)	139 (11.2)	128 (11.7)
Females	203 (28)	205 (19.1)	244 (34)	254 (31.5)

Alignment

In the first L*+H pitch accent, no consistent differences were found in the alignment of the L tone. For 2 of the 8 speakers, the H tone occurs within the stressed vowel in polar questions. Tones with this alignment are described as L+H* in GRT0BI (Arvaniti and Baltazani 2005). For the remaining speakers, the H tone occurs in the post stress syllable and aligns consistently later in polars than in statements (on average 10 ms difference).

Recall there was a low plateau found after the first pitch accent and that its two edges were labelled L2 and L3. L2 occurred within the pre-stress vowel of the second word in both statements and polars. However, for 6 out of 8 speakers, L2 aligned on average 36 ms later in neutrals than in polars. The duration of the low plateau (the L2-L3 distance) was on average 30 ms

longer in polars. There were no consistent differences in the alignment of L3 between statements and polars.

Finally, there was a difference found in the alignment of the two H tones at the edges of the high plateau. H2 occurred within the post-stress consonant in both sentence types, but it aligned on average 40 ms earlier in neutrals than in polars. No other consistent differences were found.

Discussion and conclusion

The results suggest that the pitch accents are realized differently in the two utterance types under investigation. The low plateau that was found between the two pitch accents is longer in polars than in statements: Its left edge occurs earlier in polars than in statements and its right edge occurs later. Moreover, the high peak of the second pitch accent occurs later in statements. In geometrical terms, polar questions show a longer low plateau which has more abrupt slopes at its two edges. However, it remains to be explored, through perception experiments, whether these phonetic differences are salient enough and whether listeners actually use them to distinguish between statements and questions.

The results of this production experiment, in combination with the results of follow up perception experiments have theoretical implications. If the phonetic differences detected here turn out to be salient enough to help listeners distinguish between statements and questions, then they have the status of phonological categories. In other words, intonational theory will have to expand the inventory of pitch accents to include the type found in polar questions. If, on the other hand, listeners cannot distinguish between the prenuclear pitch accents in statements from those in polar questions, then the different realizations of the L*+H found must be merely phonetic ‘allo-tones’ of the same tonal category.

References

- Arvaniti, Amalia 2002. The intonation of yes-no questions in Greek. In M. Makri-Tsilipakou (ed.), *Selected Papers on Theoretical and Applied Linguistics*, 71-83. Thessaloniki: Department of Theoretical and Applied Linguistics, School of English, Aristotle University.
- Arvaniti, A., Ladd, R.D., and Mennen, I. 1998. Stability of Tonal Alignment: The Case of Greek Prenuclear Accents, *Journal of Phonetics*, 26: 3-25.
- Arvaniti, A., and Baltazani, M. 2005. *Intonational Analysis and Prosodic Annotation of Greek Spoken Corpora*. In Jun, S.A (ed.), *Prosodic Typology and Transcription*. Oxford University Press.
- Baltazani, M., 2002. *Quantifier scope and the role of intonation in Greek*. PhD Thesis, UCLA.

Effects of VV-sequence deletion across word boundaries in Spanish

Irene Barberia

Department of Philology, University of Deusto, Spain

Department of Spanish, University of Illinois at Urbana-Champaign, USA

Abstract

Spanish does not allow vowel deletion in unstressed syllables the way languages like English do. However, it may occur in contexts of VV-sequences across word boundaries, under durational reduction in connected speech. This paper explores the effects of vowel deletion on the perception of [-high] vowel sequences on 8 native speakers of Peninsular Spanish. The results are then compared to the speakers' production of those sequences. This double experiment suggests that, although matching between perception and production does not correlate, production cues are relevant to the perception of the words containing tautosyllabic vowels. Likewise, it is claimed that perceptual cues influence the production of vowel sequences, suggesting that perceptual distinctiveness is relevant to the understanding of acoustic and articulatory preferences.

Introduction

This study aims at contributing to the understanding of the articulatory, acoustic and perceptual cues that trigger deletion of vowels in VV-sequences. It reports the results of a psycholinguistic and an acoustic study of unstressed non-high vowel sequences in Spanish (all combinations of /a/, /e/ and /o/ vowels), in order to determine which vowel in the sequences is a preferred target for deletion. The purpose was to analyze them across word boundaries in three different contexts: in *content+content*, *function+content* and *content+function* word sequences. They were also analyzed within word boundaries, in order to compare the process of deletion across word boundaries with that at initial positions within words.

Experimental design

Eight native speakers of northern varieties of Peninsular Spanish participated in this experiment. The sample words used contained all unstressed vowel sequences and appeared adjacent to stressed syllables. The syllables containing the vowel sequences had simple onsets and were codaless. The words were either bisyllabic or trisyllabic, but not longer.

Production task

For this first task, a total of 22 samples created were implemented into a list of sentences of similar length (in *content+function* context sequences with V2 /e/ were omitted, because they require the following word to begin with a high vowel, which produces the V2 and the following high vowel to merge into diphthongization). In a sound-treated room, subjects read the list three times and were recorded with a professional-quality equipment (micro: head-mounted Shure SM10A; recording, CSL 4300B Kay Elemetrics). Prior to the recordings, subjects were trained to read the sentences with an allegretto rhythm (Harris 1969), in order to sound closer to connected speech than to a reading style.

In the data analysis, the acoustic manifestation of vowel deletion was observed and also whether V1 or V2 deletion occurred.

Psycholinguistic task

The 22 samples were recreated into two sets, one with V1 and the other with V2 extraction. These nonwords were then recorded by the author (also a native speaker of a northern variety of Peninsular Spanish) and implemented into an E-Prime computer program as sound stimuli. After the production task, subjects were asked to carry a lexical decision activity by listening to those stimuli and judging whether they sounded correct or incorrect Spanish to them, by pressing different buttons. The order of the stimuli for each subject was randomized. The E-Prime program recorded the subject's responses as well as the reaction time (i.e. the time from the end of the sound to the pressing of the button) for each stimulus.

Results

The data obtained from the psycholinguistic task refers to the judgment of correct responses and it is partially reflected in table 1. Across word boundaries, the percentage of correct responses is overall greater than that within word boundaries and also higher for V1 than for V2 deletion. Within word boundaries, the values of the correct responses are not significant enough to suggest a preference for one type of deletion.

Table 1: percentage of responses judged as correct within and across word boundaries, for all speakers.

		/ae/	/ao/	/ea/	/eo/	/oa/	/oe/
V1	within	12.5%	50%	0%	0%	12.5%	12.5%
	across	56.2%	66.6%	79.1%	50%	66.6%	93.7%
V2	within	12.5%	12.5%	12.5%	0%	12.5%	12.5%
	across	56.2%	62.5%	50%	25%	37.5%	68.7%

Reaction time (RT) measurements agree partially to the conclusions obtained from the judgment results. In order to exclude outliers, mean and standard deviation values were obtained from the correct responses, and all values higher or lower than the average \pm standard deviation were excluded. Thus, across word boundaries, significant RTs range from 31.3 ms. to 1.6 sc, showing overall the highest values in *function+content* contexts. In terms of vowel quality, there is some variation depending on the context where the vowel sequence occurs. There is a tendency for V1 deleted stimuli to have shorter RTs, and so does deletion of vowel /e/, regardless of its position. In *content+function* contexts, however, when the vowel /o/ is deleted either as V1 or V2, it also shows significantly longer RTs. Values within word boundaries showed not to be significant.

The data from the production task revealed a very low number of instances of deletion production overall (other types of hiatus resolution processes were preferred). They all occurred across word boundaries and targeted V1, especially in *content+function* contexts, where V1 deletion was performed with a much greater percentage than V2 deletion. Table 2 shows the general preference for V1 deletion across word sequences, when deletion occurred. Within word boundaries it was inexistent.

Table 2: percentage of deletion production across word boundaries, for all speakers.

	/ae/	/ao/	/ea/	/eo/	/oa/	/oe/
V1	2,8%	8,3%	50%	4,1%	37,5%	6,25%
V2	25%	0%	0%	0%	0%	6,25%

In terms of vowel quality, vowel /e/ seems to be a preferred candidate for deletion, although sequence /oe/ shows the same percentage of V1 and V2 deletion. In /ao/ and /oa/ sequences, on the other hand, there was no V2 deletion at all.

Discussion and conclusion

Overall results suggest that at the phonetic level prosodic boundaries trigger deletion: it is allowed across word boundaries and V1 deletion is preferred to V2 deletion; shorter RTs (which suggest less hesitation when answering to whether the stimuli sound “Spanish” or not) are more numerous when V1 is deleted. Although perceptual cues of vowels influence deletion production (deletion of /e/ is preferred regardless of its position), preservation of initial information of words and word type seem to be more prominent factors. The longer RTs of V1 deletion in *function+content* contexts and the greater number of V1 deletion production in *content+function* also suggest that restoration of final information in a content word is easier than information in initial position of a word or phonological phrase. Specifically, in VV-sequences that contain the vowel /o/, perceptual cues do not seem to trigger any deletion. Instead, acoustic and articulatory features are suggested to influence its behaviour. The results from this experiment show a trend that should be confirmed with a larger sample of data in the future.

Acknowledgements

Thanks are due to Jennifer Cole, Rebecca Foote, Stephen Higgins, Miquel Simonet, Lisa Pierce, Marisol Garrido and Rebeka Campos. Special thanks are for José Ignacio Hualde, who hosted and guided me throughout my research at UIUC. I am indebted to Jon Franco and Carolina González for their comments and suggestions, as well as their unconditional support.

Selected references

- Aguilar, L. 1999. Hiatus and diphthong: acoustic cues and speech situation differences. In *Speech Communication* 28, 57-74.
- Aguilar, L. 2003. Effects of prosodic and segmental variables on vowel sequences pronunciation in Spanish. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 2111-2114, Barcelona, Spain.
- Casali, R. 1997. Vowel elision in hiatus contexts: which vowel goes? *Language* 73, 493-533.
- Harris, J. 1969. *Fonología Generativa del Español*. Barcelona, Planeta.
- Hualde, J. I. and Chitoran, I. 2003. Explaining the distribution of hiatus in Spanish and Romanian. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 3013-3016, Barcelona, Spain.

Production and perception of Greek vowels in normal and cerebral palsy speech

Antonis Botinis¹, Marios Fourakis², John W. Hawks³, Ioanna Orfanidou¹

¹Dept of Linguistics, University of Athens, Greece

²Dept of Communication Disorders, University of Wisconsin-Madison, USA

³School of Speech Pathology and Audiology, Kent State University, USA

Abstract

This study investigates the perceptual consequences and differences in vowel production between cerebral palsy (CP) afflicted Greek speakers and their normal counterparts. Formant (F1 and F2) values for the five vowels of Modern Greek were extracted from productions of both male and female speakers in stressed and unstressed conditions. These same productions were presented to normal hearing Greek speakers for vowel identification. Aggregate vowel spaces constructed from the mean F1 and F2 values reveal that the male CP speakers' productions more closely resemble normal female spaces than that of normal males, but has only a slight impact on reducing intelligibility. Unstressed vowel productions from female CP speakers reduced intelligibility most significantly, in particular for the vowel [o]. Significantly greater standard errors around formant means for the CP speakers' unstressed vowel productions suggest these speakers have considerably more difficulty in controlling vocal tract shape when using less vocal effort.

Introduction

This is an experimental investigation of vowels in Greek as a function of mobility (intact versus cerebral palsy afflicted system), gender, stress and vowel category. There are five vowels in Greek ([i, e, a, o, u]) and their spectral characteristics have been described in Fourakis et al. (1999). That study examined Greek vowels as produced by male speakers under different conditions of tempo, stress, finding significant effects of these factors on most acoustic characteristics (F0, amplitude, and formant structure). However, there have been no published results on vowels produced by women. In addition, there are no data on speech produced by Greek persons with cerebral palsy (henceforth CP), male or female. Thus, the present investigation presents production data (formant frequencies) for two groups of speakers – normal and CP, and perception data (classification of the productions by persons with normal hearing). Recent results on this area are reported by Liu et al. (2005) for Mandarin Chinese.

Experimental methodology

The subjects were six persons with CP and six with no known pathologies. There were three female and three male speakers in each group. All speakers used standard Athenian Greek pronunciation. Table 1 lists the characteristics of the subjects with CP.

Table 1. Personal index of the six speakers with cerebral palsy dysfunction.

Subject	Gender	Age	Education	Classification	Severity	Mobility
1	Female	27	BA	Spastic quadriplegic	Moderate to severe	Crutches
2	Female	26	BA	Dystonic quadriplegic	Severe	Wheelchair
3	Female	18	University freshman	Spastic paraplegic	Moderate to severe	Crutches
4	Male	35	High School	Hypotonic quadriplegic	Moderate To severe	Wheelchair
5	Male	23	Technical School	Spastic quadriplegic	Severe	Wheelchair
6	Male	30	BA	Spastic paraplegic	Moderate To severe	Crutches

The stimuli were nonsense CV_1CV_2 words where V_1 was one of the five Greek vowels [i, e, a, o, u], both consonants were always [s] and V_2 was always [a]. The words were presented in the carrier phrase “to kláb sV₁sa pézi kalí musikí” (The club sV₁sa plays good music). These words were produced five times by each speaker with stress either on V_1 or V_2 and at a normal rate of speech and recorded directly to disk at 44100 Hz sampling rate and 16 bit resolution. The center frequencies of the first and second formants (F1 and F2) were measured from LPC derived spectra. The LPC used a 20 ms window and the number of coefficients was adjusted for each speaker to yield good peak separation. In the perception part of the study, twelve persons with normal hearing listened to all tokens and classified them as one of the five vowels.

Results

Spectral characteristics

Figure 1 shows the vowel spaces produced by normal male and female speakers. The thick lines connect vowel means for the stressed conditions while the thin lines connect means for the unstressed condition. The relative

locations of the three point vowels are also indicated. The vowel spaces for the male speakers are in the lower F1 and F2 frequencies while those for the female speakers are in the higher frequencies. The introduction of stress results in a shrinkage of the vowel space which is more pronounced for the female speakers. However, the five vowels remain well separated in the respective vowel spaces.

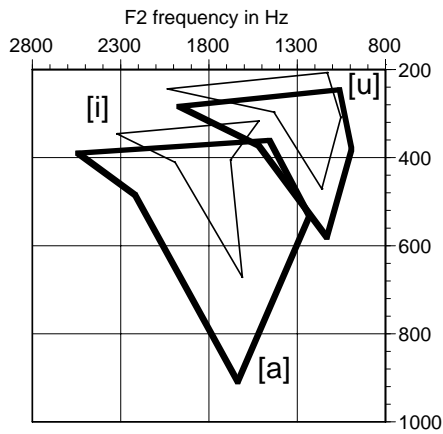


Figure1. Vowel spaces for normal male and female speakers.

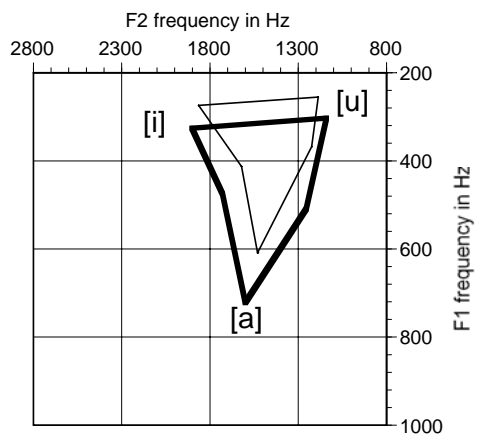


Figure2. Vowel spaces for male CP speakers.

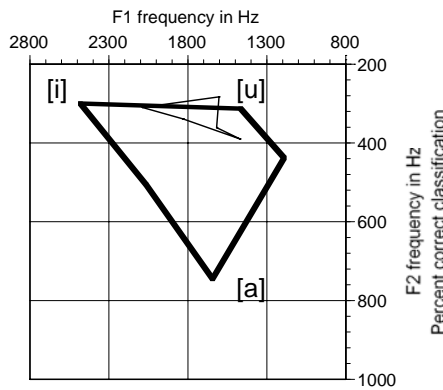


Figure 3. Vowel spaces for CP female speakers.

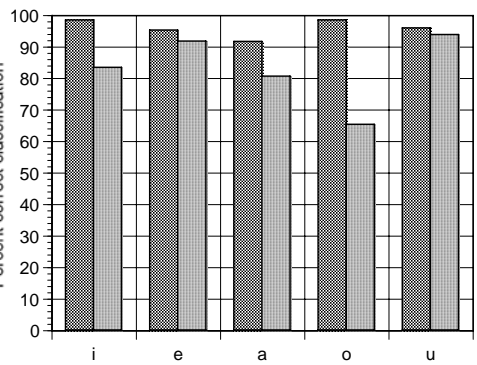


Figure 4. Classification of unstressed vowels produced by CP speakers.

Figures 2 and 3 show the stressed (thick lines) and unstressed (thin lines) vowel spaces for the male (Fig. 2) and female (Fig. 3) CP speakers. It can be seen that there is some vowel space reduction for the male CP speakers when unstressed vowels are compared to stressed ones. However, there is much more reduction for the female speakers which results (cf. discussion of Fig.4 below) in lower classification scores.

In the vowel identification experiment, the listeners with normal hearing classified vowels produced by the normal speakers with 99.7% accuracy. They classified stressed vowels produced by male CP speakers with 97.1% accuracy and female CP speakers with 99.7% accuracy. Figure 4 shows the results for unstressed vowels produced by male (checkered columns) and female (gray columns) CP speakers. Overall, vowels produced by male CP speakers were classified with 92.9 % accuracy while those produced by female CP speakers were classified with 82.6% accuracy. As can be seen in Fig.4, the vowels most affected were [i], [a], and [o] produced by the female CP speakers. Overall classification accuracy remained high, probably because of the fact that there were only five response choices and, as can be seen in Figure 3, despite the severe reduction of the unstressed vowel space, the five vowels remained well separated in the F1 by F2 acoustic space.

Discussion

The results presented here represent a first look at the effects of cerebral palsy on vowel production in speakers of Greek. In addition, acoustic and perceptual data for female Greek speakers are presented for the first time. The reduction of the unstressed vowel space relative to the stressed vowel one is much more pronounced for the female speakers than the male speakers regardless of mobility. Thus, in the CP female unstressed vowel space, which is the most reduced, there seems to be a combined effect of gender and mobility. This results in a 20% loss of intelligibility. In further research the effects on consonantal segment and whole word intelligibility will be examined using both male and female CP speakers as well as normal speakers for comparison.

References

- Fourakis M., Botinis, A. and Katsaiti, M. 1999. Acoustic characteristics of Greek vowels. *Phonetica* 56:28-43.
- Liu, H., Tsao, F. and Kuhl, P. 2005. The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *J Acoust Soc Am* 117:3879-3889.

Pre-glottal vowels in Shanghai Chinese

Yiya Chen

Department of Linguistics, Radboud University Nijmegen, the Netherlands

Abstract

This study examines the acoustic realization of two complementary sets of vowels in Shanghai Chinese: One appearing in open syllables only (OSVs) and the other in closed syllables with a glottal coda (PGVs). Two factors - consonant onset and prominence level - were controlled to test the null hypothesis that PGVs are reduced realization of the OSVs due to their short duration. Results showed spectral reduction of the PGVs but suggested that PGVs cannot be the reduced realization of OSVs as both onsetless and prominent PGVs failed to be consistently realized with a more expanded acoustic vowel space, which was predicted by the reduction account. We propose that the spectral reduction of the PGVs is an inherent feature of the vowel's phonetic implementation.

Introduction

Shanghai Chinese is a Wu dialect spoken in the city of Shanghai. One interesting feature of the language is that it has two types of vowels: One occurs in CV syllables (i.e. open-syllable vowels, hereafter OSV) and the other in closed syllables with a glottal coda (i.e. pre-glottal vowels, hereafter PGV). It is clear that PGVs are much shorter than OSVs (Zee & Maddieson 1979). The phonemic status of the PGVs, however, has been of much controversy (Chen & Gussenhoven, submitted, and references therein). This paper seeks acoustic evidence to test the null hypothesis that PGVs are the reduced and more centralized realization of some of the OSVs, given the effect of short duration on vowel production (Lindblom 1963).

Two factors, consonant onset and prosodic prominence, which have been commonly recognized as important determinants of vowel reduction, were therefore carefully controlled. Studies on the effect of consonant environment on vowel formant patterns show that vowels in CVC are in general more centralized than that in the /hVd/ context, the classical "null" environment (Peterson & Barney 1952), which is comparable to vowels in isolation (Stevens & House 1963, Hillenbrand & Clark 2001, among others). As for the effect of prosodic prominence, it has been shown that vowels occurring in a prosodically prominent position are longer and implemented with more effort, and consequently the underlying vowel target is better approximated (Fourakis 1991, van Bergem 1995, Moon & Lindblom 1994, among others). If indeed PGVs are the reduced and more centralized realization of some of

the OSVs, we then predict that PGVs with an onset and without focus should be the most centralized ones while onsetless OSVs with focus should show the most expanded vowel acoustic space.

Methods

The stimulus material comprises two sets of words: one set include five OSVs (i.e. i, u, ε, ə, a) and the other three PGVs (i.e. ɪ, ʊ, ʌ). Each set consisted of two subsets. One includes onsetless vowels and the other the same vowels with the onset /d/. Both sets are associated with a rising tone, the OSVs with Tone 3 and the PGVs with Tone 5 (indicated with superscripts), in terms of the traditional Chinese tonology (Xu and Tang, 1988).

The carrier sentence for the test stimulus (*X*) is shown in (1). Bracket indicates boundaries of the tonal domain (where only the first tone of the domain is marked). Two types of questions were asked. One is about which word is on certain row and the other about which row a specific word is on. The same stimulus sentence would be uttered accordingly, with two prosodic patterns, one with the target word focused and the other without focus.

- (1) Carrier sentence: Ξ [$\lambda\text{æ}^{\text{5}}\lambda\text{æ}^{\text{5}}$] [$\tau\text{ɪ}^{\text{2}}\tau\text{ə}\text{ə}\text{u}\text{fi}\text{ɑN}$].
X locative marker (L-M) 9th row.
 ‘*X* (target morphemes) is on the 9th row.’

Thirteen native speakers of Shanghai Chinese, ten female and three male, participated in this experiment. They were born between 1935 and 1950 and have lived most of their lives in Shanghai, mainly in the Xuhui District.

The stimuli were presented in Internet Explorer via a JAVA program, with randomized order of the stimuli. Two repetitions of the same stimulus in the same focus context were elicited. The recordings were made in stereo at the Shanghai Normal University, on a PMD660 Marantz recorder, with an AKG C477 microphone, at a 44100 Hz sampling rate, and later downsampled to 22050 Hz in GoldWave.

In Praat, the first and second formant frequencies of the vowels at the one-third and two-thirds temporal points (over the vowel duration) were computed using the LPC (autocorrelation) algorithms with its default settings for females (5500 Hz for 5 formants) and males (5000 Hz for 5 formants) respectively. The LPC spectra were occasionally recomputed with a larger number of formants (6 or 7) to separate merged formants.

For statistical analyses, the formant frequencies in Hertz were converted to the Bark scale using the formula by Traunmüller (1990) ($\text{Formant}_{\text{Bark}} = [(26.81/(1+1960/\text{Formant}_{\text{Hertz}})]-0.53$). Euclidean distance of each vowel to the center of a subject’s F1-F2 vowel space was calculated.

Results and Conclusion

To test the hypothesis that the PGVs are the reduced and centralized realization of the OSVs, Euclidean distance of each vowel was subjected to a linear mixed-effect model with Subject and Vowel as crossed random effects. The independent fixed-effect predictors included 1) position of the measurements (i.e. one-third vs. two-thirds), 2) vowel type (i.e. PGV vs. OSV), 3) onset (i.e. onsetless vs. onset 'd'), and 4) focus condition (i.e. focus vs. without focus). We observed significant three-way interactions for Focus*Type*Position [$F(1, 1650) = 4.93, p = 0.026$], Onset*Type*Position [$F(1, 1650) = 47.23, p < 0.0001$], as well as for Onset*Focus*Position [$F(1, 1650) = 5.45, p = 0.020$].

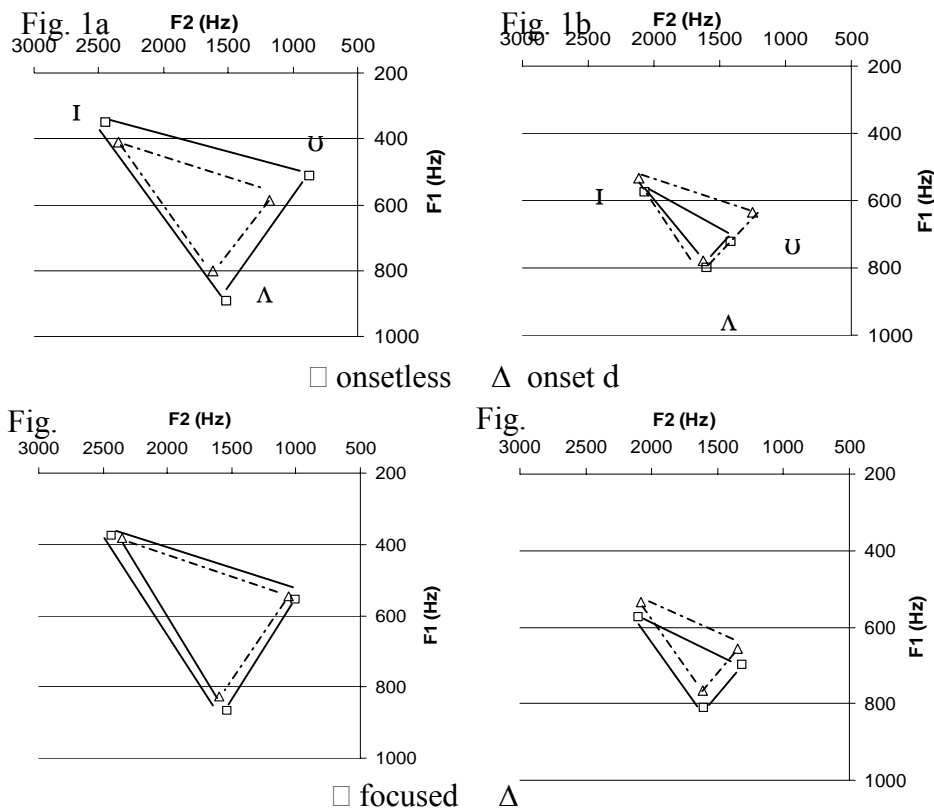


Figure 1 & 2. Effect of onset and focus on PGV formants at one-third (a) and two-thirds (b) time points.

As shown in Figures 1-2, the interactions mainly reflect the different effects of onset and focus on the PGVs at the two-thirds time point (b) from that on the PGVs at the one-third time point (a) which was similar to that on

the OSVs. Specifically, OSVs, measured at both time points, and PGVs, measured at the one-third time point, both showed more expanded acoustic vowel space in onsetless and focused conditions. PGVs measured at the two-thirds time point, however, showed that their vowel space was more expanded when following the onset /d/ and lowered when under focus, whose different magnitudes resulted in the significant interaction of focus and onset. We therefore conclude that PGVs in Shanghai Chinese should not be the reduced and centralized realization of the OSVs. Rather, their spectral reduction may be seen as an inherent feature of the vowel's phonetic implementation.

Acknowledgements

I thank Carlos Gussenhoven whose insightful comments pointed me to the right direction, Joop Kerkhof who provided much help with data extraction, Harald Baayen for his generosity with both his expertise and time on the regression model, Wuyun Pan and Qin Gu for assistance with data collection, and my subjects, of course. Usual disclaimers apply. Support by a VENI grant from the NWO is also gratefully acknowledged.

References

- Chen, Y. and Gussenhoven, C. Shanghai Chinese. Submitted.
- Fourakis, M. 1991. Tempo, stress, and vowel reduction in American English. *JASA* 90: 1816-27.
- Hillenbrand, J., and Clark, M. 2001. Effects of consonant environment on vowel formant patterns. *JASA* 109: 748 – 763.
- Lindblom, B. 1963. Spectrographic study of vowel reduction. *JASA* 35: 1773-1781.
- Moon, S. and Lindblom, B. 1994. Interaction between duration, context, and speaking style in English stressed vowels. *JASA* 1992: 40-55.
- Peterson, G. E., & Barney, H.L. 1952. Control methods used in a study of the vowels. *JASA* 24: 175-184.
- Stevens, K., & House, S. 1963. Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech Hearing Research* 6: 111-127.
- Trautmüller 1990. Analytical expressions for the tonotopic sensory scale. *JASA* 1990: 98-100.
- van Bergem, D. 1995. Acoustic and lexical vowel reduction. *Studies in Language and Language Use* 16. IFOTT: Amsterdam.
- Xu, B. and Tang, Z. (eds.) 1988. A description of the dialect of the Shanghai City. Shanghai: Shanghai Education Press.
- Zee, E. and Maddieson, I. 1979. Tones and tonal sandhi in Shanghai: Phonetic evidence and phonological analysis. *UCLA Working Papers in Phonetics* 45: 93-129.

Distinctive feature enhancement: a review

George N. Clements¹ and Rachid Ridouane^{1,2}

¹Laboratoire de phonologie et phonétique
(UMR 7018, CNRS/Sorbonne Nouvelle, Paris)

²ENST/TSI/CNRS-LTCI (UMR 5141, Paris)

Abstract

We continue the review of some of the basic premises of Quantal-Enhancement Theory (Stevens 1972, 1989, etc.) initiated in Clements and Ridouane (2006). While Quantal Theory proposes to account for similarities in feature realisation across speakers and languages, Enhancement Theory proposes to account for regular patterns of cross-linguistic variation. In this sense these two theories may be regarded as complementary modules of a more comprehensive feature theory.

Expressing variability within quantal theory

A major challenge to most feature theories comes from the existence of *variability* in speech output. The realization of a given speech form is never quite the same across utterances, and varies considerably when we take differences in speech rate, style, and speaker into account. It is easily observed that a given distinctive feature may be incompletely realized, or not realized at all, in certain utterances. Indeed, whole segments can be deleted in rapid speech, leaving no traces of their defining features. Furthermore, the realization of a given feature is not necessarily the same from one language to another.

Variability is not in the first instance a problem for Quantal Theory but for the notion of *invariance* (e.g. Perkell and Klatt 1986). Quantal Theory attempts to provide a basis for explaining why some articulatory and acoustic dimensions are favored over others for feature definitions across languages, but is not logically committed to the claim that all features are realized with their defining properties in all contexts, situations, and languages. The two views are orthogonal to each other: one may maintain a strong form of Quantal Theory while developing independent explanations for variability in feature realization.

Enhancement Theory (e.g. Stevens et al. 1986, Stevens and Keyser 1989, Diehl 1991) has precisely this mission. Starting out from the premise that much crosslinguistic variation is not random but systematic, it attempts to seek the reasons for which some languages systematically prefer one type of feature realization while others prefer another. According to this theory, when the acoustic difference between two sounds is insufficiently great,

risking confusion, a supplementary gesture may be introduced to increase the acoustic difference between them. In some cases, this gesture corresponds to a redundant feature, as when the feature [+rounded] is introduced to enhance the difference between back vowels and front vowels (Stevens et al. 1986). This feature has the effect of increasing the auditory difference between front and back vowels by increasing their difference in F2 frequency.

In other cases, the enhancing gesture may be subfeatural, as in the case of the nondistinctive lip-rounding usually added to palato-alveolar sibilants like /ʃ/ in English, increasing their auditory difference from alveolar sibilants like /s/. In this case, too, the enhancing gesture targets the inherent acoustic parameter distinguishing the two sounds and adds a gesture that increases the difference along this parameter. In the case of coronal sibilants, the universal correlate of a post-alveolar [-anterior] fricative appears to be turbulence noise in the region of F3. Adding lip-rounding to such a sibilant accents the spectral prominence in this frequency region and increases its perceptual distance from its [+anterior] counterpart /s/. Such enhancement would not, of course, be expected in languages that have no contrast of this sort.

Many examples of this sort are provided by Keyser and Stevens (2006). We provide a further example showing that not only feature contrasts, but skeleton-based contrasts between simple and geminate speech sounds can be enhanced in the same way. Tashlhiyt Berber, like many other languages, has a lexical contrast between two types of consonants, singleton and geminate, distinguished phonologically by their association to one vs. two skeletal positions (Dell and Elmedlaoui 1997). As illustrated in (1), this contrast is attested in all positions and concerns all types of consonants:

- | | | | | |
|-----|------|-----------|-------|---------------|
| (1) | tut | 'she hit' | ttut | 'forget him' |
| | imi | 'mouth' | immi | 'mother' |
| | ifis | 'jackal' | ifiss | 'he is quiet' |

The distinction between simple and geminate consonants is carried not just by duration but by a combination of properties. The primary property is the extra duration of geminates, since this property appears in every context in which the contrast occurs, even in voiceless stops following pause where the closure duration is extra-long even though it has no direct acoustic manifestation (cf. *tut* vs. *ttut* above). In addition, the singleton/geminate contrast is enhanced by further acoustic attributes such as shorter preceding vowel duration, higher RMS amplitude, and complete stop closure (Ridouane, in press). These correlates, interpreted as manifestations of greater articulatory energy, serve to enhance the primary feature by contributing additional acoustic properties which increase the perceptual distance between singletons and geminates. These enhancing correlates can take on a distinctive

function in cases where the primary correlate is not perceptually recoverable. This is, for instance, the case for voiceless stops after pause, where duration differences between singletons and geminates is not detectable by listeners.

Enhancing vs. feature-defining gestures

Subfeatural enhancing gestures such as these display two properties that distinguish them from basic, feature-defining gestures (Keyser and Stevens 2006). First, unlike feature-defining gestures, subfeatural enhancing gestures are not part of the basic feature definition as such. Evidence that they may have a different status in speech production comes from a consideration of speech error phenomena, in which forms which do not exist in the lexicon can be erroneously generated (e.g. *hash or grass* becoming the erroneous *hass or grash*, Fromkin 1973). Such errors can be generated by a transposition of basic feature-defining gestures, in this case the gesture used to distinguish [+anterior] from [-anterior] consonants. However, it appears that no speech error transpositions solely involve enhancing gestures. Thus in a word like *sunshine*, in which the palato-alveolar feature is enhanced by lip-rounding, there are no recorded speech errors in which this lip-rounding appears on the initial /s/ while the spread lip configuration expected on the /s/ appears on the palato-alveolar segment.

A related characteristic of subfeatural enhancement gestures is that their implementation appears to be *graded*. In the example just given, the degree of rounding of the palato-alveolar fricative is weaker and more variable than that of the featurally initiated lip rounding in a [+rounded] vowel such as /u/. Subfeatural enhancing gestures are non-discrete and continuous, in contrast to feature-defining gestures which are discrete and quantal in nature.

Enhancement displays another rather unexpected property. While feature-defining gestures are often weakened or obliterated in casual speech, enhancement gestures tend to survive intact, preserving underlying distinctions. Here we consider an example involving assibilation. In the historical development of many Bantu languages, the raising of mid vowels to lower high vowels came to threaten the distinction between these vowels and the upper high vowels (we follow the interpretation of Mpiranya 1997). Many of these languages developed strongly assibilated variants of stops before the vowels of the higher series. Following this development, the distinction between upper and lower vowels usually disappeared; but due to assibilation, the distinction between words with earlier upper high vowels and those with earlier lower high vowels persevered. This example is significant in showing that enhancement effects may apply not only to the segments bearing threatened feature distinctions but to segments in their context as well.

In sum, Enhancement Theory offers a basis for understanding some types of regular cross-linguistic variation. Starting from the observation that

languages tend to preserve useful contrasts, it proposes that supplementary features and gestures may be marshalled to reinforce existing contrasts between two sounds or sound sequences along an acoustic dimension that distinguishes them. Once introduced, these features tend to survive, and may eventually supplant the feature which they originally served to enhance.

Summary

While Quantal Theory proposes to account for similarities in feature realization across languages, Enhancement Theory proposes to account for (some of) the differences. In this sense they may be regarded as complementary aspects of a more general feature theory.

References

- Clements, G.N. and R. Ridouane. 2006. Quantal Phonetics and Distinctive Features: a Review. Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics, 28-30 August 2006, Athens, Greece, pp. xxx-xxx.
- Dell, F and M. Elmedlaoui. 1997. Les géminées en berbère. *Linguistique Africaine* 19, 5-55.
- Diehl, R.L. 1991. The role of phonetics within the study of language. *Phonetica* 48, 120-134.
- Fromkin, V. A. (ed.) 1973. *Speech errors as linguistic evidence*. The Hague, Mouton & Co.
- Keyser, S. J. and Stevens, K.N. 2006. Enhancement and overlap in the speech chain. *Language* 82.1, 33-63.
- Mpiranya, F. 1997. Spirantisation et fusion vocalique en bantou: essai d'interprétation fonctionnelle. *Linguistique Africaine* 18, 51-77.
- Perkell, J. and Klatt, D.H. (eds.) 1986. *Symposium on Invariance and Variability of Speech Processes*. Hillsdale, NJ, Lawrence Erlbaum.
- Ridouane, R. in press. Geminata: quality or quantity? *Journal of the IPA*.
- Stevens, K.N. 1972. The quantal nature of speech: Evidence from articulatory-acoustic data. In Denes, P.B. and David Jr., E.E. (eds.), *Human Communication, A Unified View*, 51-66. New York, McGraw-Hill.
- Stevens, K.N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17, 3-46.
- Stevens, K.N. and Keyser, S.J. 1989. Primary features and their enhancement in consonants. *Language* 65.1, 81-106.
- Stevens, K.N., Keyser, S.J. and Kawasaki, H. 1986. Toward a phonetic and phonological investigation of redundant features. In Perkell and Klatt, 426-463.

Where the wine is velvet: Verbo-pictorial metaphors in written advertising

Rosa Lúcia Coimbra, Helena Margarida Vaz Duarte and Lurdes de Castro Moutinho
Department of Languages and Cultures, University of Aveiro, Portugal

Abstract

In this presentation, we intend to approach pictorial metaphors (PM) and verbo-pictorial metaphors (VPM) in written Portuguese advertising, using a corpus of newspaper and magazine advertisements. As theoretical framework, we use the Cognitive Linguistics theory of the multiple spaces and blending by Gilles Fauconnier and Mark Turner (2003) and part of Charles Forceville's (1996) approach to VPM. The collected corpus is analysed, and the main objective of this study is the presentation of a typology for the diverse ways to establish conceptual blending in VPM (Coimbra, 2000). It is argued that there are at least six categories of VPM.

Introduction

Advertising has a prominent place among the text genres that speakers most frequently read, not only due to its prevalence in mass media, but also in all urban places. Such profusion has led to increasing attempts to catch the attention of the target public, which account for more and more imaginative texts. One original resource consists on the metaphorical reading of the interaction between linguistic elements and images within the advertisement.

The main objective of this study is the presentation of a typology for the diverse ways to establish conceptual blending in VPM (for more details on this typology, cf. Coimbra, 2000).

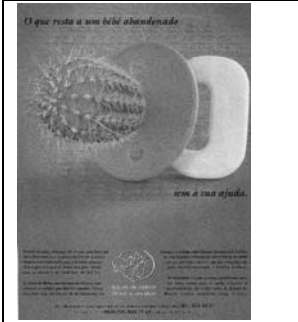





In spite of accounting for a more comprehensive phenomenon -conceptual blending- the multiple space theoretical model is appropriate to our research since VPM include the fusion of two different input spaces that allow the creation of a blended space where the new emergent reality appears. Our examples illustrate the different processes employed to operate the interaction between input domains in the VPM in written advertisements. It is argued that there are several categories which account for different ways of building and decoding the message and its polissemes.

Pictorial metaphors

The main characteristic of metaphor is that it connects things that were not previously connected, at least not in an obvious way. A conceptual metaphorical link can be expressed in a variety of ways: linguistic metaphors, pictorial metaphors, cinematic metaphors, etc. Forceville (1996: 65) suggests that “an account of pictorial metaphor should show an awareness that a metaphor has two distinctive terms, one primary subject or tenor, the other the secondary subject or vehicle”.

Our typology consists of six categories that account for the different processes employed to operate the interaction between input domains in the PM in written advertisements. They are the following:

Table 1. This table shows examples of the different categories of our typology of pictorial metaphors.

 <p><i>O que resta a um bebê abandonado sem a sua ajuda.</i></p>	 <p><i>Veludo na boca.</i></p>	 <p>ABSOLUT VODKA</p> <p>ABSOLUT SUMMER.</p>
1. Fusion	2. Context	3. Distortion
	 <p><i>Viagem para a descoberta</i></p>	 <p>JIMBO - ESPECIAL AQUECIMENTO</p> <p>Apaciguamentos a preços muito confortáveis. De 80 de Outubro a 31 de Novembro.</p> <p>Jimbo Acquer</p>
4. Superposition	5. Angle	6. Alignment

1. Fusion. On this category, we include the cases that Forceville (1996: 109-126) calls MP2s, that is, pictorial metaphors in which the two terms are merged in order to create a new, non previously existing object. The first example in table 1 shows a soother (pacifier) that is visually blended with a cactus full of thorns. The tenor is the soother, since it is an advertisement about an institution that helps abandoned babies. The words that constitute

the slogan, “What’s left to an abandoned baby without your help” explain that the thorns of the cactus mean the extreme suffering these babies feel. The metaphor is very powerful because we can imagine what one of these soothers would do to a baby’s delicate mouth.

2. Context. This category includes the examples where the metaphorical reading is not produced by a fusion, but by inserting the vehicle into a context where we would expect the tenor to be. Forceville (1996: 126-132) also studied these cases, which he calls MP1s, and which are characterised by the absence of the tenor (or source domain counterpart, or input space 1, according to Fauconnier and Turner’s terminology) which, instead, is suggested by the visual context in which the vehicle (or target domain counterpart, or input space 2) is placed and which is not the one we are used to see it in. The second example in table 1 shows a red velvet cloth placed in a visual context where we would expect red wine to be. The association is obvious: the synaesthetic sensation of smoothness.

3. Distortion. The possibility of processing images by computer software made this category one of the most used in pictorial metaphor. The two objects are captured simultaneously in a pictorial blend. Example 3 of table 1 shows us a blend between a bottle and a swimming pool by changing the latter’s typical rectangular shape.

4. Superposition. The fifth category accounts for the examples where the two counterparts are overlapped, so that we think of one in terms of the other. It is the case of the woman in example 4 table 1 whose strategic business capacities are compared to the sharp vision of the eagle.

5. Angle. Another possibility is to present the object of one domain in such a position that reminds us of the angle in which we are used to see the other object. It is the case of the office table in example 5, which is placed as a vessel. The link is underlined by the words: “Voyage to discovery”

6. Alignment. This category includes examples such as 6 where several objects are placed in such a way that they together make the shape of the absent object. In this case, 5 heaters align to form a couch.

Verbo-pictorial metaphors

Context plays an important role in differentiating pictorial from verbo-pictorial metaphors. Forceville (1996: 159) explains the difference in the following terms: “the subdivision into pictorial and verbo-pictorial metaphors is not an absolute one. (...). If one were to delete all textual material from an advertisement, and the two terms of the metaphor could still be identified, then the metaphor in question is a pictorial metaphor or simile”.

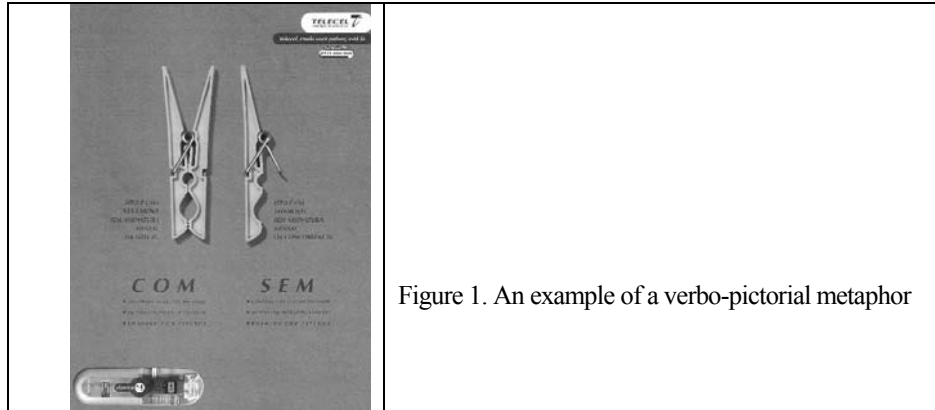


Figure 1. An example of a verbo-pictorial metaphor

It is not the case of our example in figure 1. Without the linguistic context, the reader would not understand that the picture of the two pegs required a metaphoric interpretation. In fact, the words “With Without” mean that a cell phone may be useful or useless, depending on it having or not the service that is being advertised.

Conclusion

There are several strategies to make the reader aware of the subtle presence of one of the metaphorical counterpart. In this paper, we presented some of the ways to achieve this goal. Pictorial and verbo-pictorial metaphors are, as we may see, by the examples presented, interesting ways of drawing the reader’s attention by placing an unusual reading of both image and text. Due to its pervasiveness, it is a phenomenon that should be the object of the study of discourse analysis. By its complexity, verbo-pictorial metaphors are a challenge to linguists and an open field to new an interesting research.

References

- Coimbra, R. L. 2000. Quando a Garrafa é um Porco: Metáforas (Verbo)Pictóricas no Texto Publicitário. In Castro, R.V. and Barbosa, P. (eds), *Actas do XV Encontro Nacional da Associação Portuguesa de Linguística*. vol. 1, 243-253, Faro, Portugal.
- Fauconnier, G. and Turner, M. 2003. *The Way we Think: Conceptual Blending and the Mind’s Hidden Complexities*. New York, Basic Books.
- Forceville, C. 1995. IBM is a tuning fork: Degrees of freedom in the interpretation of pictorial metaphors. *Poetics*, 23, 189-218.
- Forceville, C. 1996. *Pictorial Metaphors in Advertising*, London/ New York, Routledge.

Measuring synchronization among speakers reading together

Fred Cummins

School of Computer Science & Informatics, University College Dublin,
Ireland

Abstract

It has been demonstrated that speakers are readily able to synchronize with a co-speaker when reading a prepared text together. The means by which a high degree of synchronization is attained are still unknown. We here present a novel measure of synchrony which allows us to follow the time course of synchronization among two speakers, based on the parallel acoustic signals. The method uses traditional frame-based cepstral features and a slight variant on standard dynamic time-warping. We develop the method based on a novel corpus of synchronous speech, comparing its estimates of synchronicity with hand estimates. The method out-performs laborious manual estimation, and allows us to now begin to study the dynamics of synchronization among speakers.

Synchrony Among Speakers

It has been demonstrated that speakers are readily able to synchronize when reading prepared texts together (Cummins, 2003). The degree of synchrony achieved is remarkable (typically with lags of about 40 ms), and does not improve much with practice. Synchronization in joint activity is, of course, quite common, but typically such activities are periodic in nature. Collaboration in working a two-man saw, in juggling, dancing or playing music, all rest on a periodic basis. Despite the naive impression of speech as ‘rhythmic’, it is practically never regularly periodic (Dauer, 1983). This poses the question, then, of how speakers manage to maintain such tight synchrony without extensive practice. In order to study the process and timecourse of synchronization, a method is required to continuously assess the degree of synchrony obtaining among speakers.

Assessing Synchrony

Previous work on synchrony among speakers has relied on a point-wise estimate of synchrony. In so doing, a few points are chosen which are clearly identifiable in both speech waveforms. The lag between speakers at each of these points serves as an instantaneous estimate of synchrony.

Using this method, Cummins (2003) reported that asynchrony at phrase onsets was slightly greater than medially (62 ms, vs 40-44 ms medially). The point-wise method does not allow a continuous estimate of synchrony, and, in particular, it is difficult to see when, and how often, the lead changes between the two speakers.

Continuous assessment using Dynamic Time Warping

Dynamic Time Warping (DTW) is a well known algorithm, commonplace in speech recognition, which allows one to identify an optimal warping of one sequence onto a referent, with some common-sense constraints such as monotonicity and continuity (Meyers and Rabiner, 1981). Figure 1 (left) illustrates the path identified by DTW in aligning two short symbolic strings. As one progresses from the bottom left square, one can choose only the square to the North, East, or to the North-East as the best match at any given point.

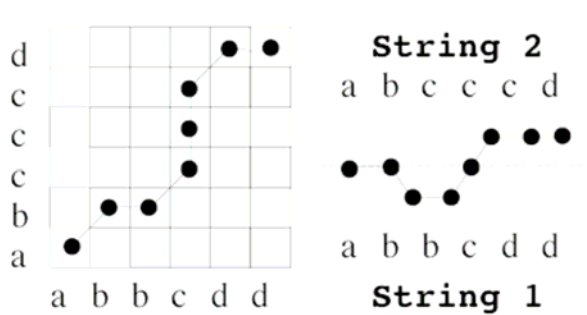


Figure 1: Illustration of standard Dynamic Time Warping path estimation (left) and conversion to an estimate of 'synchrony' (right).

The path identified by DTW can be used to derive a continuous assessment of synchrony among speakers as follows:

We first represent the speech of each of two speakers as a sequence of equally spaced vectors. Conventional MFCCs appear to do just fine for this, though one could consider using any other parametric representation such as PLP, Rasta, or even LPC coefficients. One simplification we can make is to require that each speech sample be of the same length as the other (equal number of frames). This may involve including silence at the start of one or other sample.

We then generate the optimal warping path, similar to that in Figure 2. Notice that this path veers above and below the main SW-NE diagonal.

When it is above that diagonal, the second string ('abbcdd') is ahead of the first, and when it is below the diagonal, the first leads the second.

We now redraw the path, with the SW-NE diagonal as our time axis. Steps in the DTW algorithm which move NE constitute a step of one frame width in the horizontal direction. Steps N or E each constitute deviations towards one or other string, and each such step advances $0.5 \times \text{frame width}$ along the horizontal time axis. The resulting path is illustrated in Figure 1 (right panel). It can be seen directly that String 2 leads String 1 initially, and that the lead changes, just after the mid point of the two strings.

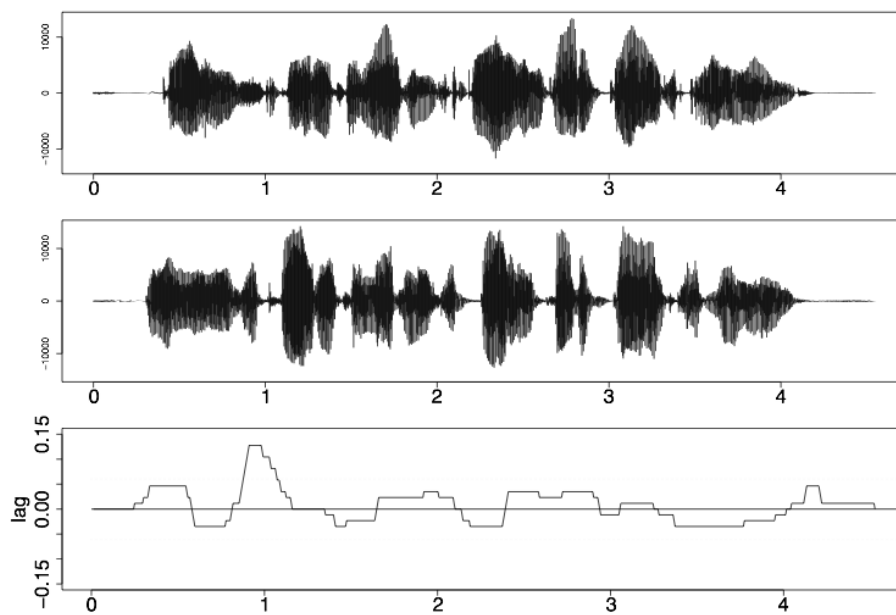


Figure 2: Two synchronous speech waveforms and associated synchrony estimate.

Figure 2 illustrates the result obtained for two phrases spoken in approximate synchrony by two male Irish speakers. When the synchrony estimate lies below the mid-line, the top speaker leads and vice versa. Dotted lines have been added at lags of ± 60 ms. It can clearly be seen from the lower synchrony estimate that the two speakers are quite closely coupled, and that the lead is exchanged several times throughout this one short phrase (The phrase was “There is, according to legend, a boiling pot of gold at one end”.)

Preliminary results

We have developed and tested this method of estimating synchrony on a subset of a database of synchronous speech we have recently collected (Cummins et al, 2006). Synchrony was estimated for 12 female and 12 male speakers reading the first paragraph of the 'Rainbow Text' in dyads. Our initial question was simply whether it was generally possible to identify a leader-follower relationship in a synchronous dyad, or whether the lead repeatedly changed from one to the other. In all 12 dyads, we found regular changes of the lead, and such changes typically happened several times within each individual phrase, as shown also in Figure 2. We were thus able to rule out the hypothesis that synchronous speaking is based on a simple leading-following relationship. Given that the vast majority of lags observed were of a magnitude less than 60 ms, and the shortest speech shadowing lags typically reported are of the order of 200 ms (Marlsen-Wilson, 1973), this is reassuring that synchronous speech does, in fact, require entrainment or coupling among speakers. However, it raises the question of how synchrony is maintained in the absence of a strictly periodic structure. Future investigations will address the following key issues:

[1] What timing information is extracted by one speaker in synchronizing with another, and

[2] To what extent does synchronous speech reveal unmarked, default speech timing, reflecting shared knowledge of what is marked and unmarked within a language/dialect.

Acknowledgements

The present work was funded by a grant from Science Foundation Ireland to the author.

References

- Cummins, F. 2003. Practice and performance in speech produced synchronously. *Journal of Phonetics* 31(2), 139-148.
- Cummins, F. 2004. Synchronization among speakers reduces macroscopic temporal variability. *Proc. 26th Annl. Meet. Cognitive Science Society*, 304-9.
- Cummins, F. and Grimaldi, M. and Leonard, T. and Simko, J. 2006. The CHAINS corpus: CHAracterizing INdividual Speakers. *Proc. SPECOM'06*. To appear.
- Dauer, R.M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11, 51-62.
- Marslen-Wilson, W. 1973. Linguistic structure and speech shadowing at short latencies. *Nature* 244, 522-523.
- Myers, C. S. and Rabiner, L. R. 1981. A comparative study of several dynamic time-warping algorithms for connected word recognition. *The Bell System Technical Journal* 60(7): 1389-1409.

Formal features and intonation in Persian speakers' English interlanguage wh-questions

Laya Heidari Darani

English Department, Azad University, Khorasgan Branch, Esfahan, IRAN

Abstract

The Unitarianist Interface Hypothesis proposed by Lotfi (ms.) concerns only one interface level -the Semantico-Phonetic Form- at which both performance systems can have access to various types of features of a lexical item. This paper hypothesizes the existence of an iconic relationship between formal and phonological features, and the influence of formal features on intonation in Persian speakers' English interlanguage wh-questions. The present study is based on computing t-test procedure on the data obtained from twenty-four Iranian L2 learners of English age-ranged between twelve and sixteen. The comparison of the data collected from the syntax and intonation tests before and after the instruction confirms the above hypotheses.

Introduction

This paper is a report on an SLA research project in Phonology and Minimalist Syntax with specific reference to formal features and phonological features.

Lotfi (ms.), contrary to Chomsky (1995), proposes a hypothesis -The Unitarianist Interface Hypothesis- according to which there is only one interface level called the Semantico-Phonetic Form at which both A-P and C-I performance systems can have access to a derivation comprising different types of features such as formal, phonological, and semantic features.

The major hypothesis of the research is that there is an iconic relationship between formal and phonological features. This hypothesis leads to two other hypotheses: 1) the acquisition of formal features has influence on the acquisition of intonation patterns for Iranian L2 learners of English; 2) the acquisition of formal features improves the phonetic performance of such participants.

Theoretical Background

Different lexical items have different types of features. These features are formal, phonological, and semantic. As different statements have different intonation patterns, Välimaa-Blum (2001) mentions that yes/no questions are

incomplete in themselves without the answers, so they use the rising tone, but *wh*-questions have a falling tone.

With regards to the existence or non-existence of an iconic relationship between formal features and phonological features, Chomsky (1995; 2000) assumes that the faculty of language consists of a cognitive system that stores information and some performance systems— A-P and C-I performance systems interacting with the cognitive system at two interface levels of PF and LF respectively—which are responsible for using and accessing information. Chomsky (1995) claims that a derivation converges finally if it converges at both PF and LF; otherwise, it crashes. With respect to the assumption that the convergence of a derivation is conditional upon its interpretability at both interface levels, he hypothesizes that there are no PF-LF interactions relevant to convergence.

Regarding what Chomsky (1995) believes in, Lotfi (2001) asserts that if there are no interactions between PF and LF, if the derivation *D* converges at PF but it crashes at LF, this means that *D* is to crash, at last. Now, how does PF 'find out' that *D* has crashed at LF that it will not be articulated phonetically? How do PF and LF communicate?

Based on the above question presented by Lotfi (ms.), he proposes the Unitarianist Interface Hypothesis, according to which there is only one interface level called the Semantico-Phonetic Form at which both performance systems can have access to a derivation.

Design of the study

Participants

A group of forty Iranian L2 learners of English as a foreign language, both male and female age-ranged between twelve and sixteen were selected randomly from among Elementary level students of a private language school in Esfahan, Iran.

The participants were given both syntax and intonation pre-tests. Those participants who failed both tests were selected as the sample population of this study. This group was 24 participants who took part in the instruction.

Materials

Different types of materials were used during the study. There were syntax and intonation pre- and post-tests each of which was the parallel form of the other. In the syntax tests the participants were asked to identify grammaticality and ungrammaticality of the statements and also to make *wh*-questions. During the intonation tests the students had to read the statements and the questions aloud while their voice was being recorded.

In addition to the above tests, two other types of materials were employed. The first one was a kind of booklet. It was researcher-made and composed of a reading passage and the grammatical point in question followed by a series of exercises. The grammatical point to be instructed to the participants was making wh-questions using the operation of movement. As it was supposed for the participants to actually see the movements, a number of cards on which different components of a statement or a question had been written were used, too. To move the cards more easily on the board a few magnets were also employed.

Procedures

During rating the tests various grammatical and phonological criteria were considered. For example, rating the syntax tests, deleting the element in question and substituting the wh-word were among the factors which had to be taken into account.

To rate participants' phonetic performance, a female native speaker's oral performance was employed as the criterion, so that participants' performance was compared using a computer software -Praat 4.1.1, © 1992-2003 by Paul Boersma and David Weenink.

In order to investigate the significance of the results of the study, the data obtained were subjected to the t-test in order to test the significance of the difference between the mean scores of the pre-tests and post-tests.

Results

To determine the participants' performance on syntax and intonation pre-tests, the mean procedure was employed. The mean score for the syntax pre-test was 20.29 while the minimum and the maximum scores were 7 and 34, respectively. The mean score for intonation pre-test was 14.8 while the minimum and maximum scores were 6 and 21. The mean scores of post-tests were also computed. The mean score of syntax post-test was 35.96 with minimum and maximum scores of 24 and 43; the mean score of intonation post-test was 22.80 with minimum and maximum scores of 16 and 27.

To figure out whether the participants' performance on syntax and intonation pre- and post-tests were significantly different before and after the instruction the t-procedure was employed. The results indicated that t-test value for the syntax pre- and post-tests with the level of significance for two-tailed test of .000 and df of 23 was 11.006 and that of the intonation post-tests with the same level of significance and df was 7.582

Conclusion

According to the results obtained from the data, it was indicated that there was an improvement in the grammatical and phonological performance of the participants before and after the instruction so that the mean scores of both post-tests were higher than those of the respective pre-tests. The results from the t-test procedures not only confirmed the usefulness of the instruction but also indicated the existence of an iconic relationship between formal and phonological features and the improvement in the acquisition of phonological features through the acquisition of formal ones.

Last but not the least, from a theoretical point of view, it can be said that the Unitarianist Interface Hypothesis was confirmed via this study; that is, different types of features in a lexical item are interpreted at the same interface level and that accordingly they have influence on each other. Moreover, empirically speaking, using movement operation as a technique to teach grammar proved to be useful in this study.

Acknowledgements

I would like to express my gratitude to Dr. A. R. Lotfi and Professor William Greaves for their helpful guidance during this research.

References

- Chomsky, N. 1995. *The minimalist program*. Cambridge, MA, MIT Press.
- Chomsky, N. 2000. *Minimalist inquiries: The framework*. In Martin, R., Michaels, D. & Uriagereka, J. (eds.) 2000, *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89-155. Cambridge, MA, MIT Press.
- Lotfi, A. R. 2001. *Iconicity: A generative perspective*. (Homepage) [Online] Available at URL: <http://www.geocities.com.arlotfi/lotfipage.htm>.
- Lotfi, A. R. (ms.), *Semantico-phonetic form: A unitarianist grammar*. Peter Lang. [online], Available at URL: <http://www.geocities.com/arlotfi/lotfipage.htm>
- Välismaa-Blum, R. 2001. *Intonation and interrogation in English: Its all a matter of default*. *Semantique syntaxe et linguistique Anglaises*. [Online], Available at URL: <http://www.unice.fr/ANGLAIS/alaes/blum.htm>.

The effect of semantic distance in the picture-word interference task

Simon De Deyne, Sven Van Lommel and Gert Storms
Department of Psychology, Leuven University, Belgium

Abstract

First, we investigated whether semantic interference in the picture-word task generalizes from close neighbours to more distant category members. Second, we investigated the operationalisation of semantic distance. The presented work manipulated closely versus distantly related distractor words, and examined to which degree various operationalisations of distance determine semantic interference. Within-category interference did occur, however depending on the quantity and quality of the semantic relationship. The effect was limited to very similar coordinates or semantic neighbours, and did not occur for less similar category coordinates. Furthermore, the associative implementation of semantic distance was most predictive for the interference effect. The findings provide a better understanding of the different measures of semantic distance as well as the process of semantic interference.

Semantic distance

In the picture-word interference task, pictures are named while attempting to ignore simultaneously presented distractor words. Semantically related distractor words have been repeatedly shown to interfere with picture naming (e.g., Lupker, 1979) and this finding contributed significantly to the modelling of speech production. The multitude of studies however failed to produce a precise definition of the required type of semantic relatedness. Semantic coordinate distractors often interfere with picture naming (Costa et al., 2003), but many studies based the selection of coordinate distractors on mere intuition. Yet not all members of the same category may invoke semantic interference, depending on the semantic distance between the two category members.

The present study investigated whether the SI effect occurs when the distractors are semantic coordinates that are neighboring category coordinates (e.g. lion-tiger, ‘neighbors’ for the remainder of the article) or more distant coordinates (e.g. lion-monkey, ‘coordinates’ for the remainder of the article) compared to unrelated distractors.

Furthermore, categorical relatedness can be operationalised in various ways (Ruts et al., 2004). The relatedness is often operationalised by collecting similarity ratings of concept pairs. But the underlying process driving these judgements has been argued to only pertain to a selection of context-

dependent dimensions, in which the objects under consideration are similar (Medin, Goldstone, & Gentner, 1993). In contrast, relatedness based on the feature- or association-patterns of a concept can be computed independent from context.

The applied semantic distances are based on previously collected norms for similarities, features and associations as described in Ruts et al. (2004). The similarity judgments are pair-wise subjective judgments of similarity judged on a scale going from 1 (totally dissimilar) to 20 (totally similar). The other two semantic distance models are based on a vector space model and correspond closely to the Word Association Model reported by Steyvers, Shiffrin, and Nelson (2004).

Experiment

Method

Each of 44 pictures appeared on a white background along with three distractor words individually superimposed on the picture: (1) a semantic neighbor (e.g., lion on the picture of a tiger); (2) a remote category member (e.g., monkey); and (3) an unrelated word (e.g., window). Each distractor occurred once in the entire set and was closely matched for word frequency. Selection of semantically close neighbors and semantically dissimilar coordinates was based upon previously collected pair-wise similarity norms (Ruts et al., 2004). The average similarity was .73 for the semantic neighbors, and .39 for the coordinates.

The participants were instructed to name the pictures as quickly and accurately as possible. A trial consisted of a fixation point, the picture with distractor, the same picture without the distractor and a blank screen during which the voice key registered responses.

Results and discussion

First, we investigated the interference effects in the three semantic distance conditions. The data were analyzed using a one-factor within-participant ANOVA. Distractor type was significant, $F_1(2,60)=13.16$, $p<.001$, $MSE=1,720.58$; $F_2(2,86)=11.04$, $p<.001$, $MSE=6,134.03$. Post hoc Tukey tests indicated that neighbors differed significantly from coordinates and unrelated distractors ($p_1<.01$; $p_2<.05$), whereas the unrelated and coordinate distractors did not. In conclusion, only neighbor distractors showed an SI effect.

Table 1. Mean Naming Latencies for Each Condition.

	Neighbors	Coordinates	Unrelated
M	957	915	912
SD	90	650	1100

Secondly, a simultaneous regression was performed on the average latencies of the 88 distractors using three semantic predictors: similarity-, associations- and feature-based semantic distances. Collinearity was verified but turned out not to characterize the data. The regression model was significant with $R^2=.19$. Only association distances were significant $\beta=.42$, $SE=65$, $p<.01$. A second model was tested where CELEX log-transformed word-form frequency was added to the list of predictors. The model was significant, $R^2=.26$. Again, there was a significant effect of the association predictor, $\beta=.34$, $SE=53$, $p<.05$.

General Discussion

Three main findings can be derived from the data. First, interference by semantic neighbors result in a 50ms processing disadvantage compared to neutral control words. Second, semantically remote distractors resulted in equally fast responses as the unrelated words. The SI effect is thus not strong enough to affect all category members of a semantic concept. This pattern contradicts early findings by Lupker (1979) where semantic interference was found between concepts that share few semantic properties with the picture. Third, the semantic interference effects correlate to a varying degree with three different measures of similarity. A semantic distance measure based on word associations accounted for a significant part of the naming latencies, but direct similarity ratings and feature based measures did not.

The first two findings indicate that a semantic relationship defined as belonging to the same basic-level category is not a sufficient condition for obtaining a SI effect. Task specific factors such as the SOA of the distractor and the possible set-size of the responses (La Heij & Van den Hof, 1995) could have attenuated the SI effect. But such explanation would retain a similar issue: Why the SI effect is more easily attenuated in coordinate words, as the present experiment setup did suffice for a SI effect in neighbor words. Another factor that needs to be explored is the proportion of related and unrelated distractors in the experiment. In the current study, only one third of the words were truly unrelated to the picture, which might increase habituation to related distractors.

The third finding indicates that the occurrence of a SI effect might depend on how semantic distance is measured (cf. Mahon et al., submitted). The lack of an SI effect might be due to the properties of the semantic dis-

tance model used. A closer investigation of the semantic distance measures can reveal some characteristics of the dimensions or depth of processing from which the SI effect originates. In contrast to the rated similarity and features model, an association model (e.g., Steyvers, Shiffrin, & Nelson, 2004) is based on automatic and more superficial responses. We argue that the semantic information available in speeded task conditions rather resembles the information captured by the association approach than by a model based on rated similarities or features. The association measure used was based only on the first three responses in a word association task (see Ruts et al., 2004). Such close associations may rather reflect the most salient characteristics of the stimuli, whereas similarity judgments are based on more central or causal features. Similarly in feature ratings, participants need to describe the concepts with at least 10 features. Task demands of the similarity and feature ratings might thus over-emphasize causal features as compared to an association task.

References

- Costa, A., Mahon, B., Savova, V., & Caramazza, A. 2003. Level of categorization effect: A novel effect in the picture-word interference paradigm. *Language & Cognitive Processes* 18, 205-233.
- La Heij, W., & Van den Hof, E. 1995. Picture-word interference increases with target-set size. *Psychological Research* 58, 119-133.
- Lupker, S. J. 1979. The semantic nature of response competition in picture-word interference task. *Memory & Cognition* 7, 485-495.
- Mahon, B., Costa, A., Peterson, R., & Caramazza, A. submitted. The effect of semantic distance in the picture-word interference paradigm: Implications for models of lexical selection.
- Medin, D. L., Goldstone, R. L., & Gentner, D. 1993. Respects for Similarity. *Psychological Review* 100, 254-278.
- Ruts, W., De Deyne, S. Ameel, E., Vanpaemel, W., Verbeemen, T., & Storms, G. 2004. Flemish Norm Data for 13 Natural Concepts and 343 Exemplars. *Beh. Res. Meth., Instruments, and Computers* 36, 506-515.
- Steyvers, M., Shiffrin, R. M., & Nelson, D. L. 2004. Word Association Spaces for Predicting Semantic Similarity Effects in Episodic Memory. In Healy, A. (ed.) 2004, *Experimental Cognitive Psychology and its Applications*. Washington DC: American Psychological Association.

Melodic contours of yes/no questions in Brazilian Portuguese

João Antônio de Moraes

Faculdade de Letras, Universidade Federal do Rio de Janeiro/CNPq, Brazil

Abstract

In this paper I describe four melodic patterns which can occur in the yes/no question, namely: i) the final rise pattern, typical of the unmarked polarity yes/no question, ii) the internal rise pattern, correlated with positive polarity questions, iii) the delayed rise pattern, indicating negative polarity, i.e. the speaker's discordance with the sentence propositional content, and iv) the double rise pattern, found in rhetorical yes/no questions and requests. The relevance of phonological representation is discussed on the grounds of auditory tests with synthesized speech.

Yes/no questions melodic contours

Questions in Brazilian Portuguese have been described as presenting three melodic contours according to their syntactic structure: rising, in the yes/no question, falling, in the wh-question and rising-falling, in the alternative question (Hochgreb 1983, Moraes 1998). A closer examination of the question intonation in spontaneous speech, however, reveals a much more complex picture as a consequence of the incidence of two pragmatic factors: the question's negative or positive polarity and its degree of dependence on the conversational context.

Together with the falling pattern, typical of statements, four interrogative patterns that characterize different types of yes/no questions were analyzed, viz.:

i) final rise, typically associated with the neutral yes/no question, considered unmarked as to the expected answer. This pattern is characterized by a medium level onset and a high level over the final stressed syllable (Fig. 1).

ii) internal rise, which is correlated with confirming questions bearing positive polarity, i.e. the speaker expects that the listener agrees with the propositional content of the question. The pattern is characterized by a rise at a high melodic level in the first stressed syllable, a level which continues to rise throughout the utterance, including the final prestressed syllable, to fall on the final stressed syllable (Fig. 2).

iii) delayed rise, which implies disbelief or doubt about what has just been said, therefore suggesting disagreement with the propositional content of the question. What distinguishes this pattern is the fact that the melodic

rising in its last stressed syllable starts only at its second third, creating a slightly concave-shaped rising (Fig. 3).

iv) double rise, which occurs in several discursive situations, among which are requests and questions that bear a “rhetorical” intention with an expected answer in the opposite direction of the propositional content expressed in the question. The pattern is characterized by a rising in the first stressed syllable, and a second, weaker, rise in the last stressed syllable (Fig. 4).

v) falling pattern, belonging to statements, is characterized by an onset at a medium level and by a low melodic level over the final stressed syllable (Fig. 5).

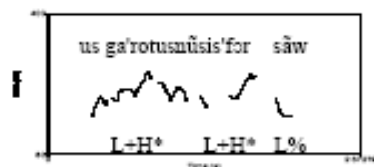


Fig 1. Final rise melodic contour

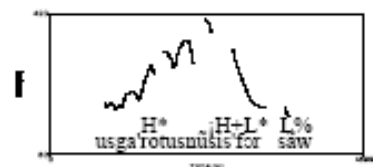


Fig 2. Internal rise melodic contour

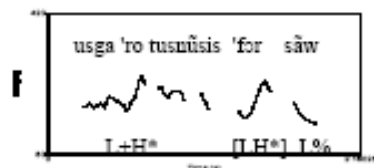


Fig 3. Delayed rise melodic contour

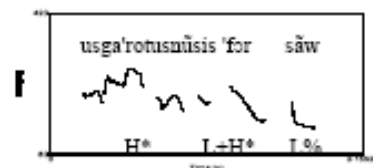


Fig 4. Double rise melodic contour

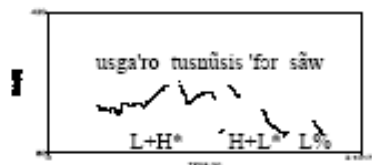


Fig 5. Falling melodic contour

Experiments with synthesized stimuli

An examination, even superficial, of these contours reveals that it is not a simple task to assign, with current notational conventions (Beckman et al. 2005), a phonological representation to these five patterns, as proposed at the bottom of the pitch curves in figures 1-5. Indeed, in order to represent the difference between these patterns, artificial solutions will have to be used, which will make the notation less phonetically transparent or ad hoc conventions, unforeseen by the system, must be established, such as the distinction between [LH*] and L + H* to indicate that the bitonal accent is located

over just one syllable or spreads over two syllables. Clearly, notations little motivated phonetically must be evaluated before being adopted. In fact, what is the real significance of the prenuclear accents melodic behavior (quite similar, in these cases) in the characterization and identification of the analyzed patterns? Besides the melodic level proper, what relevance can the intrasyllabic shape of the melodic contour have in stressed syllables?

Expecting to contribute with a tentative answer to these questions, an experiment was designed using 19 resynthesized prosodic variants of the sentence *Os garotos não se esforçam* (The boys don't try hard), with pitch modifications at prenuclear and nuclear accents. The stimuli were placed in three groups for forced choice tests, and were randomly presented to 25 subjects that estimated the effect of the melodic modifications on their meaning.

Results of the perception tests

Test I On the significance of the prenuclear accent

The first test aimed at the assessment of the importance of the prenuclear accent on the recognition of intonational meaning. The five natural sentences were resynthesized in such a way as to neutralize the melodic distinctions that occurred at the prenuclear accent. The results showed that the intentions assigned by the speaker to both the natural sentences and the sentences in which the prosodic information at the prenuclear accent was eliminated were correctly identified by the listeners in a statistically significant level ($\chi^2 = 56.51$ and $p < 0.0001$; $\chi^2 = 82.97$ and $p < 0.0001$, respectively).

Test II On the rising or falling configuration of the final H* of final rise and double rise patterns

Based on the final rise pattern with neutralized prenuclear accent, eight variants were produced, four of them with rising intrasyllabic configuration and four with the falling one, lowering by 30Hz the mean value in each curve, as it can be observed in figure 6, below:

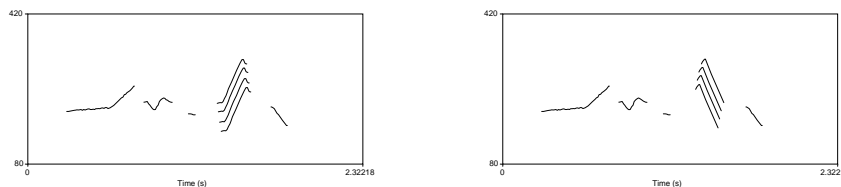


Fig. 6. Eight synthesized variants of the final rise pattern with rising and falling shape over the final stressed syllable

The tests results indicated that the melodic shape of the final stressed syllable is the feature responsible for the identification of neutral (rising) or rhetorical (falling) questions ($\chi^2 = 52.36$; $p < 0.0001$), and not the distinction represented as L+H* and H* pitch accents in the prenuclear position, as proposed in the notation on figures 1 and 4, respectively.

Test III On the rising shape of the final H* of final rise and delayed rise patterns

Based on the delayed rise pattern, six stimuli were generated, crossing the shape of the melodic rising, convex or concave, with three distinct syllabic durations: the original one, one increased by 30% and the other decreased by 30%.

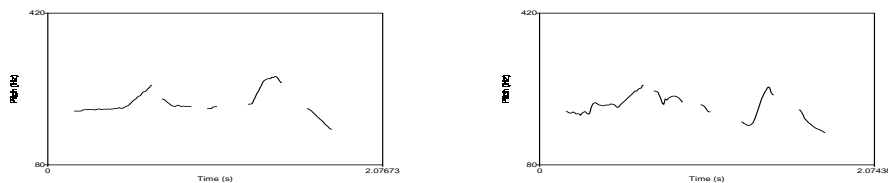


Fig. 7 Convex and concave shapes of the final rise, respectively.

The test results showed that the shape of the rising configuration on final stress syllable is the feature which leads to the identification of neutral (convex shape) or disbelief (concave shape) questions ($\chi^2 = 77.63$; $p < 0,0001$).

Final remarks

Our results strongly suggest that the current phonological representation does not account for aspects that are important to the characterization of the intonation patterns described here. On the one hand, the representation of the prenuclear accent is irrelevant, since the opposition between the patterns is concentrated on the nuclear accent. On the other hand, the direction, falling or rising, of the melodic modulation over stressed syllables, and even its sharpness must be taken into account since they are factors that effectively participate in the distinction of these patterns.

References

- Beckman, M., J. Hirschberg and S. Shattuck-Hufnagel 2005. The original ToBI system and the evolution of ToBI framework. In Jun, S.-A. (ed.) *Prosodic Phonology. The Phonology of Intonation and Phrasing*, pp 9-54. Oxford, OUP.
- Hochgreb, N. 1983. *Análise acústico-perceptiva da entoação do português: A frase interrogativa*. Tese de doutorado, Universidade de São Paulo.
- Moraes, J. A. 1998. Intonation in Brazilian Portuguese. In Hirst, D. and A. Di Cristo (eds.) *Intonation Systems: a Survey of Twenty Languages*, 179-194. Cambridge, CUP.

The phonology and phonetics of prenuclear and nuclear accents in French

Mariapaola D'Imperio¹, Roxane Bertrand¹, Albert Di Cristo² and Cristel Portes²

¹Laboratoire Parole et Langage – UMR 6057 CNRS, Aix-en-Provence, France

²Université de Provence and Laboratoire Parole et Langage - UMR 6057 CNRS, Aix-en-Provence, FRANCE

Abstract

The goal of this paper was to determine whether there is a formal difference between high-ending nuclear (IP-final) and prenuclear accents in French. After transcribing the relevant items, we then compared the different accentual and phrasal categories by analyzing, among the other things, the tonal and temporal characteristics of their tonal targets as well as durational characteristics of the target syllables. The hypothesis tested is that nuclear accents differ from prenuclear ones in terms of formal characteristics of the contour which cannot be explained, for instance, by invoking the presence or absence of an upcoming tone. We show that both alignment and scaling differences can be found between the two accents types, thus motivating a contrastive phonological analysis.

Introduction

Despite the very large body of literature on prosody in French, there is still no consensus about which and how many constituents should be included in the prosodic hierarchy as well as which differences between contours are categorical. Two influential models of French intonation posit the existence of two prosodic constituents, a higher level one which can be identified with the Intonation Phrase (IP), and a lower level which is either defined in purely tonal terms, as in the case of the AP posited by Jun and Fougeron (2002) or as a rhythmic constituent which is built on the basis of prosodic phonological rules thus with reference to the syntax. This is the case of the phonological phrase of Post (2000).

Higher in the prosodic hierarchy (being also the highest level), we find the IP. Traditionally, this last phrasing level is also the domain of the “nuclear accent”, which is positionally defined as being the last accent within this constituent, as well as being the most prominent one in the prosodic hierarchy. Any preceding accent in the intonation phrase is defined as “prenuclear”.

Within the autosegmental-metrical model of intonation, only the positional definition remains, though defining a lower phrasing level, that of the intermediate phrase. Specifically, the nuclear accent is the accent immediately preceding the phrase accent. In the models of French intonation proposed by Jun and Fougeron (2002) and Post (2000) there is no formal distinction between nuclear and prenuclear accent type. However, Di Cristo (1999), claims that a phonological distinction between these two accents should be kept for French.

This work aims at investigating the potential difference between an IP-final rising contour and an IP-internal (for instance, accentual-phrase final) rising contour, which are usually described as containing the same rising pitch accent. Specifically, the pitch accent associated with the “primary stress” position in standard French has been described as either a H* (Post, 2000) or as a LH* (Jun & Fougeron, 2002) pitch accent.

Corpus and Methodology

Our corpus (D'Imperio *et al.*, 2006) consisted of sentences in which the same syllable [mi] could occur either in IP-final or non-final position of left-dislocated sentences (e.g., *Les amis de mami, ils ne viennent que demain* “The friends of grandma, they will only come tomorrow”), and was always potentially at least AP-final. Moreover, we varied the syllable structure between open and closed as well as the number of APs within the IP (2 or 3).

Four speakers of French participated into the experiment. Here we will present data relative only to two speakers.

As for the acoustic measures, alignment and scaling were considered as indices of tonal structure. We also believed that in order to get the most complete view of the phenomenon at hand we needed to follow a double approach. That is we performed both an auditory and an acoustic analysis of the contours. The auditory transcription was performed by two experts, and aimed at distinguishing between a rising (R) versus a high (or same, S, relative to preceding syllable) level (S/H) accent at the primary stress location. Moreover, since it was particularly difficult to manually identify the location of L targets, we employed a semi-automatic procedure described in D'Imperio et al. (in press).

Results and Discussion

The results for the auditory transcription show complete agreement for IP-final accents, which were all transcribed as rises, while the agreement was much lower for non-final accents (speaker SC: Kappa = 0.02; speaker SD: Kappa = 0.1). Note that, despite the higher disagreement within the non-final

position, almost 80% of agreed H (S/H) accents were detected for speaker SD, while mostly LH rises (R) were detected for speaker SC. Hence, the accent type transcribed for the two subjects was not the same in pre-nuclear position. This called for investigating the difference acoustically.

Here we present acoustic results for one speaker (SC), since the behaviour for the other speaker was comparable. Figure 1 (upper, left) shows F0 height for the H target in both pre-nuclear (first) and nuclear (last) position.

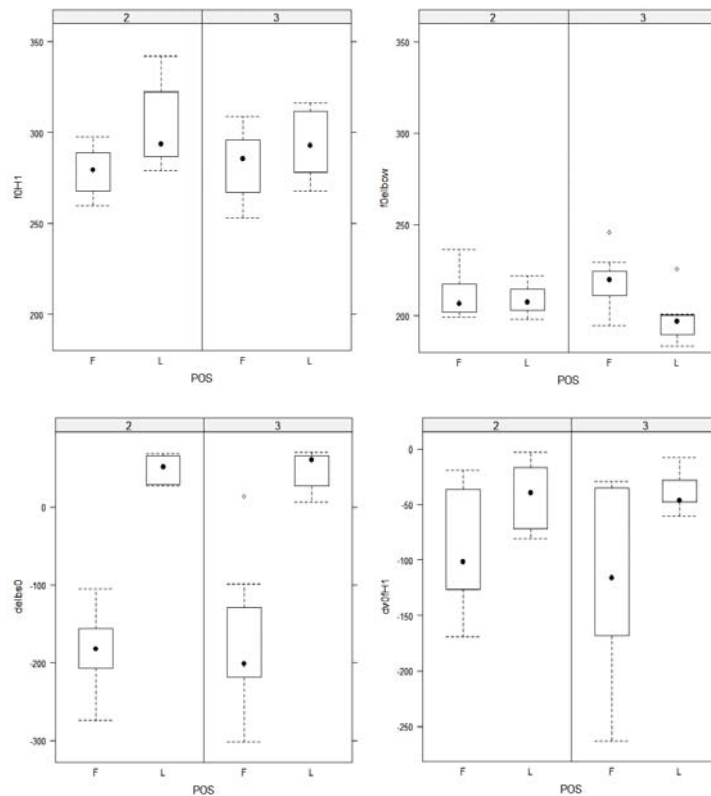


Figure. 1 Mean F0 values of the H targets (upper, left) and of the L targets (upper, right) by position within the IP (F = first; L = last) and by number of APs within the IP (2 vs. 3). The lower section shows L alignment relative to stressed vowel onset (lower, left) and H alignment relative to stressed vowel offset (lower, right).

As expected, we found higher F0 values for the nuclear (last) H. Note also a tendency for pre-nuclear Hs to be higher in 3-AP items than in 2-AP items. F0 values for the L target are also shown in Figure 1 (upper, right).

Unexpectedly, we found greater values in prenuclear position only for 3-AP items. The bottom of Figure 1 shows alignment results of the L target relative to stressed syllable onset (lower, left), while H alignment was measured relative to the offset of the stressed vowel (lower, right). Note that L alignment was earlier for prenuclear rises, while H alignment was later (closer to stressed vowel offset) for nuclear rises, though this measure is less consistent for the prenuclear H.

Our study shows that the hypothesis of the similarity between nuclear and prenuclear accent in French is not supported neither by transcription nor by acoustic data. Nuclear rises are higher, and both its H and L targets are later than in prenuclear rises. The alignment differences found cannot be simply accounted for by the presence of a H% in nuclear position. The earlier H alignment in prenuclear position is not due to the presence of a L boundary tone or phrase accent at the end of the AP, since the falling contour following the AP-final rise does not seem to have a fixed slope (Jun and Fougeron, 2002). Moreover, the F0 raising effect on the L target of prenuclear rises in 3-AP items could be due to a preplanning effect, and should be further investigated.

The old distinction between prenuclear and nuclear accents in traditional intonation studies cannot be completely dismissed for French. Hence, we propose a phonological analysis of this contour as consisting of a prenuclear H* vs. a nuclear L(+H)* gesture. The differences reported need further exploration through additional acoustic data, especially of spontaneous speech (cf., Portes and Bertrand (2005)), as well as through perception experiments, to verify whether the accent type opposition has a categorical nature.

References

- Di Cristo, A. 1999. Le cadre accentuel du français contemporain: essai de modélisation. *Langues* 2(3): 184-205, 2(4): 258-267.
- D'Imperio, M., Bertrand, R., Di Cristo, A. and Portes, C. 2006. The phonology and phonetics of prenuclear and nuclear accents in French, LSRL 2006, Rutgers, New Jersey, March 31-April 2, 2006.
- D'Imperio, M., Espesser, R., Loevenbruck, H., Menezes, C., Welby, P. et Nguyen, N. (in press). Are tones aligned with articulatory events? Evidence from Italian and French. In Cole, J. and Hualde, J. (eds.) *Papers on Laboratory Phonology IX*, Oxford : OUP Press.
- Jun, S.-A. and Fougeron, C. 2002. Realizations of accentual phrase in French intonation. *Probus* 14, 147-172.
- Portes, C. and Bertrand, R. 2005. Some cues about the interactional value of the "continuation" contour in French. *Proceedings of IDP05*, oct. 8-9, Aix-en-Provence, France.
- Post, B. 2000. *Tonal and phrasal structures in French intonation*, The Hague : Holland Academic graphics.

The influence of second language learning on speech production by Greek/English bilinguals

Niki-Pagona Efsthathopoulou
Department of Linguistics, Simon Fraser University, Canada

Abstract

This study examined 20 Greek/English bilinguals, living in the Vancouver area, Canada. The duration of aspiration for the three voiceless stops (/p/, /t/, /k/) and for the following vowel for a series of Greek and English stimuli were analyzed. Factors such as the Age of Arrival (AOA), Age of Learning (AOL), Length of Residence (LOR), everyday use of Greek and English and self-estimated proficiency in both languages were taken into consideration. A number of English sentences produced by the same speakers were also collected and rated for accentedness by native speakers of English. The degree of accent of the bilingual subjects were also examined along with the data collected. The main focus of the study is to see how the interaction of first language (L1) and second language (L2), if any, is observed concerning VOT (voice onset time) and following vowel duration.

Introduction

The focus on this study is the comparison of Greek and English VOT (Voice Onset Time) for voiceless stops consonants /p/, /t/, /k/ produced by Greek/English bilinguals. VOT was examined in the initial stressed syllable of disyllabic Greek and English words. Thus, the factors of stress and syllable ordering were controlled. The interaction between the L1 and L2 phonological systems was also explored. Factors correlating with VOT production were also taken into consideration such as the AOA (Age of Arrival), AOL (Age of Learning), LOR (Length of Residence), Education level, everyday use of Greek and English and self-estimated proficiency in both languages. The participants also read a number of sentences that were judged for accentedness by a group of native English speakers.

The expected results are that L1 affects L2 and the other way round, thus VOT in both L1 and L2 are expected to deviate from the average performance of monolinguals. Also, it is expected the accent-ratings to correlate with the degree of shift in the production of VOT. The stronger the accent that a subject has, the least affected is expected to be the Greek VOT by English/L2 and its English VOT is expected to be more affected by Greek/L2.

Theoretical background

This study follows Flege's Speech Learning Model (SLM), (Flege, 1987, 1992, 1995), as well as Best's (1993) Perceptual Assimilation Model (PAM).

Methodology

The participants in this study were 20 Greek/English bilinguals living at the time in the Greater Vancouver area, BC, Canada. Greek was their first language (L1) and English was their second language (L2). There were 11 male and 9 female subjects, mean age=58, (range=39-78) who had been living in an English speaking environment for at least 16 years (range=16-50, mean=35).

The recordings were made in the quietest room in the participants' house or working place by the same Greek speaking researcher using a tape recorder; the data were later digitized at 44KH and 16bits. The procedures were conducted mostly in the Greek language, except when the participant him/herself shifted to English. The participants first signed a consent form and then completed a language background questionnaire, where information such as their profession and age, AOA, AOL, LOR, Education level, self-estimated proficiency in both languages and self-estimated everyday use of Greek were collected.

Then, the participants wearing a head-mounted microphone read a list of 20 Greek and 20 English disyllabic words. Each list was read three times. The English stimuli were read in the carrier phrase 'I say__again', while the Greek carrier phrase was "Ipa__pali". Therefore, the data collected were [20X15X3] =900 instances of VOT and vowels for Greek and English respectively. At the end of the session, the participants read three English sentences¹ three times so the data collected were [20X3X3]=180 sentences.

In the accentedness experiment four native speakers of English rated the sentences for accentedness using a 9-point scale ranging from 1= no accent to 9= extremely strong accent (mean age=40, range 26-59). The researcher scanned the sentences for overall quality selecting three from each subject, so the stimuli presented were 20X3=60. The stimuli were randomized for each of the three times that each judge listened to them.

Results - discussion

As shown in studies of Italian/English (MacKay & al, 2001), French/English (Flege, 1987, 1991), English/Portuguese (Major, 1992), Greek/English bilinguals (Efstathopoulou, 2006), it is expected that the bilingual subjects will produce intermediate VOT values for English VOT, because of L1 influence, but also Greek VOT values longer than those of monolinguals because

of L2 influence. The influence is expected to be similar for vowel length. However, the VOT-to-vowel ratio will be examined. Since values for Greek VOT (Fourakis, 1986) and vowels (Hawks & Fourakis, 1995), are shorter in comparison with the English ones (Klaa, 1975, Lisker & Abramson, 1967).

Also, the accent rating of the subjects will be taken into account along with the VOT and vowel measurements. The four native speakers/judges gave accent rating for the 20 bilingual subjects, as shown in the following table along with the subjects' age, age of arrival, length of residence, estimated percentage of use of Greek.

Table 1. Accent ratings.

SJ#	Age	AOA	AOL	LOR	% of Use	Overall Mean
G1	65	27	20	38	40	7.28
G2	62	20	20	42	20	4.33
G3	71	25	17	46	90	6.67
G4	39	23	16	16	60	6.94
G5	39	21	4	18	40	1.58
G6	42	25	10	17	15	3.22
G7	57	24	12	33	50	5.78
G9	55	27	28	28	80	2.22
G10	68	18	18	50	60	6.47
G11	57	19	14	38	60	7.19
G12	62	23	23	39	40	5.33
G13	60	27	15	33	0.5	7.00
G14	55	22	14	33	50	5.42
G15	64	19	19	45	60	5.58
G16	65	27	27	38	95	7.64
G17	57	19	19	38	95	8.22
G18	52	17	14	35	30	4.67
G19	61	32	16	29	80	5.78
G20	54	15	13	39	70	5.69
G21	78	29	20	49	10	7.33

Notes

1. The sentences, a subset from Munro & Derwing (1995) were: a) The Queen of England lives in London, b) Some people love to eat chocolate, c) Ships travel on the water.

Acknowledgements

I thank my subjects for their collaboration. I am grateful to my supervisor Dr. Munro for his constant guidance and support. I am solely responsible for any mistakes.

References

- Efstathopoulou, P.N. 2006. The influence of second language learning on speech production by Greek/English bilinguals: A pilot study, the case of VOT, Presented at NWLC 2006.
- Flege J. E. 1992 “Speech Learning in a second Language” in *Phonological Development, Models, Research, and Applications*, ed. By C. Ferguson, L. Menn, and C. Stoel-Gammon, Parkton MD: York Press, 565-604
- Flege J. E. 1987. The production of ‘new’ and ‘similar’ phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 45-67.
- Flege J. E., 1991. Age of learning affects the authenticity of voice onset time (VOT) in stops consonants produced in a second language. *Journal of Acoustical Society of America* 89, 395-411.
- Flege J. E. 1995. Second language speech learning: theory, findings and problems. In W. Strange, ed., *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium, MD: York Press, 237-277.
- Fourakis, M. 1986. A timing model for word-initial CV syllables in Modern Greek. *Journal of the Acoustical Society of America* 79, 1982-1986.
- Hawks, J. & Fourakis, M., 1995. The Perceptual Vowel Spaces of American English and Modern Greek: A Comparison. *Language and Speech* 38 (3), 237-252.
- Klatt, D. 1975. Voice onset time, frication and aspiration in word-initial clusters. *Journal of Speech and Hearing Research* 18, 686-705.
- Ladefoged, P. 1993. *A course in Phonetics*, 3rd ed., Harcourt Brace & Company.
- Lisker, L. & Abramson, A., 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20, 384-422.
- Lisker, L. & Abramson, A., 1967. Some effects of context of voice onset time in English stops. *Language Speech* 10, 1-28.
- Major, R., 1991. Losing English as a first language. *The Modern Language Journal*, 76, 190-208.
- McKay I., Flege J., Thorsten P. & Carlo S.. 2001. Category restructuring during second-language speech acquisition”, *Journal of Acoustical Society of America*, 110(1), 516-528.
- Munro M. & Derwing T. 1995, “Processing time, accent, and comprehensibility in the perception of native and foreign accented speech.” *Language and Speech* 38 (3), UK: Kingston Press Services, 289-306.

Aspectual composition in Modern Greek

Maria Flouraki

Department of Language and Linguistics, University of Essex, UK

Abstract

The rich aspectual system in Modern Greek involves both morphologically expressed grammatical aspect and eventuality types. Particular emphasis is paid to the interaction between grammatical aspect and eventuality types, since it is due to this interaction that the verbal predicate acquires distinct meanings. The main aim of this paper is to explain potential changes in the meaning of the eventualities caused by the interaction with grammatical aspect and provide a formal analysis of this interaction. I propose an analysis within Head-Driven Phrase Structure Grammar (HPSG), using Minimal Recursion Semantics (MRS) for the semantic representations. I argue that grammatical aspect is a function which takes as arguments particular eventualities. When the arguments are different from the required ones, then there are instances of reinterpretations, which are not instances of ungrammaticality. This can be explained with the introduction of subeventual templates, where grammatical aspect combines with eventuality types and selects eventualities or subeventualities appropriate to its selection restrictions, using information that is already there in the denotation of the eventualities.

Introduction

Aspectual composition occurs when grammatical aspect (perfective and imperfective) and eventuality types (accomplishments, achievements, processes, states) carried by the verb along with its arguments combine to trigger particular meanings. This aspectual composition may change the denotation of the eventuality type resulting to aspectual shifts. (Moens and Steedman 1988, Jackendoff 1990, Pustejovsky 1995, Krifka 1998, de Swart 1998, Giannakidou 2002, Egg 2002, Michaelis 2004).

An instance of this phenomenon is found in Modern Greek (M.G.) where there is a contrast between perfective and imperfective aspect, being overt in the morphology of the verb. The information, grammatical aspect presents, is affected by the eventuality type it combines with, which is implicit in the meaning of the verb phrase.

In (1) there is a process eventuality, which denotes a situation where *Giannis loves Anna* but is not clear when this loving situation starts and when and whether it finishes. When this eventuality occurs with imperfective aspect in (1a), it gets the default meaning of the eventuality, where no culmination point is denoted and no visible endpoints. In (1b) the same eventuality combines with perfective aspect, which may focus either on the

initial stages of the eventuality in which case it acquires an inchoative reading or simply adds both endpoints, in which case we get a bounded reading.

(1a) O Giannis agapous -e tin Anna.
 the Giannis love.imperf -past.3sg the Anna
 `Giannis was loving Anna' / `Giannis used to love Anna`

(1b) O Giannis agapis -e tin Anna.
 the Giannis love.perf -past.3sg the Anna
 `Giannis loved Anna (and does not love her any more)'(bounded)
 `Giannis fell in love with Anna' (inchoative)

The aspectual shifts involved are subtypes of type shifts, which in the literature are formalised with the usage of a functor argument relation: $f(a)$, where f is the functor and a the argument. In the case of aspectual shifts, there is a functor-argument relation between grammatical aspect and eventuality types (2a). Aspect is further instantiated into the imperfective functor which combines with processes and states (2b).

- (2a) aspect (eventuality type)
 (2b) imperfective(process \vee state)

There are cases where the argument is not the appropriate input for the functor as in (1b). However, there is no ungrammaticality involved but just re-interpretations occur, which remedy the conflict.

An explanation for these re-interpretations lies in the sphere of extralinguistic knowledge. The general relation $f(Op(a))$ is used, where the operator Op added, is given by pragmatic context. A major drawback of these approaches is that these operators can not be appropriately constrained, so that they occur only where and when needed.

Following Michaelis (2004) and Pustejovsky (1995), we provide an alternative, where we develop a highly constructed inventory of eventuality types, which consists of eventualities as well as their subeventualities. These interact with grammatical aspect, which adds or selects the whole or subparts of the eventualities according to its selection restrictions.

The Analysis

The analysis proposed follows the framework of Head-Driven Phrase Structure Grammar (HPSG) (Pollard and Sag 1994), using Minimal Recursion Semantics (MRS) for the semantic representations (Copestake et al. 2000). Following (MRS) architecture, we introduce a number of relations, which represent both the grammatical aspect functor and the eventuality type argument given in (2a).

The *aspect-rel* has the features L(a)B(e)L and BINDS as indicated in (3). The LBL identifies the relation and shows its scopal connection with the other relations whereas the BINDS feature shows the eventuality the *aspect-rel* has to bind with. The *aspect-rel* combines with an eventuality through the BINDS feature and gives back the same or a different eventuality through the EVENT-STR(ucture) feature.

- | | |
|-------------------------|----------------------------------|
| (3) [<i>aspect-rel</i> | (4) [<i>eventuality-rel</i> |
| LBL <i>handle</i> | LBL <i>handle</i> |
| EVENT-STR | EVENT-STR [EVENT1 <i>process</i> |
| <i>event-str</i> | EVENT2 <i>state</i> |
| BINDS <i>event-str</i> | RESTR 1 < 2]] |

The eventualities are decomposed into subparts so as grammatical aspect to be able to select the appropriate subpart in each case. Following Pustejovsky (1995), we support that each *eventuality-rel* has an event structure (EVENT-STR), whose value is a feature structure, that consists of different subeventualities indicated by the features EVENT1 and EVENT2.

The *eventuality-rel* in (4) introduces apart from the attributes LBL, the attribute EVENT-STR, which has two subeventualities: An EVENT1 with value a *process* type and an EVENT2 with value the *state* type. Their temporal ordering is guaranteed through the RESTR(iction) attribute, which states that there is a precedence temporal relation between EVENT1 and EVENT2.

Following Michaelis 2004, we support that as in the Romance languages, imperfective aspect in M.G. is a type-selecting operator, which reflects the eventuality type of its arguments. It modulates when it is necessary the aspectual properties of its argument and denotes eventuality types and place constraints upon the types it combines with. This kind of combination is guaranteed by the *Aktionsart preservation principle*, where no extra material is needed intervene in the functor argument relation.

Hence, the imperfective functor takes as argument particular eventualities and when the argument is not the appropriate input, the functor selects or adds a subpart to the eventuality it combines with. It selects process eventualities and returns an output of the same eventuality as the input as in (5a). When it combines with transition eventualities, it selects only the process subeventuality which is appropriate for its selection type as it is shown in (5b), where when the input is a transition then the output is just the process subeventuality.

- (5) $F_{\text{imperf}}(X, Y) = Z$
 (5b) if $Y = [\text{EVENT1 } \textit{process}]$, then $Z = Y [\text{EVENT1 } \textit{process}]$
 (5c) if $Y = [\text{EVENT1 } \textit{process}$
 $\text{EVENT2 } \textit{state}]$, then $Z = [\text{EVENT1 } \textit{process}]$

Conclusion

Through the account provided we have shown that eventualities consist of subeventual templates and grammatical aspect selects each time an appropriate subeventuality as input according to its selectional restrictions. Particular meanings are inferred which are already there in the denotation of the eventuality and they just need to be picked up by grammatical aspect.

Acknowledgements

I would like to thank my PhD supervisor Prof. Louisa Sadler and Dr. Doug Arnold. This research was supported by ESRC funding. Contact details: maria.flouraki@gmail.com

References

- Copestake, A., Flickinger D., Pollard C. and Sag, I. A. 2000. Minimal Recursion Semantics: an Introduction. *Language and Computation* 1, vol.3, 1-47.
- De Swart, H. 1998. Aspect shifts and Coercion. *Natural language and Linguistic Theory* 16, 347-385.
- Egg, M. 2002. Semantic Construction for Reinterpretation Phenomena. *Journal of Semantics* 15(1), 37-82.
- Giannakidou, A. 2002. A puzzle about Until and the present perfect. In Alexiadou, A., Rathert, M. and von Stechow, A. (eds.), *Perfect explorations*, Mouton de Gruyter.
- Jackendoff, R. 1997. *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.
- Krifka, M. 1998. The origins of Telicity. In Rothstein, S. (ed.), *Events and Grammar*, 197-235, Great Britain.
- Michaelis, L. 2004. Type shifting in Construction Grammar: An integrated approach to Aspectual Coercion. *Cognitive Linguistics* 15, 1-67.
- Moens, M, and Steedman, M. 1988. Temporal ontology and Temporal Reference. *Computational Linguistics* 14, 15-28.
- Pollard, C. J. and Sag, I. A. 1994. *Head-Driven Phrase Structure Grammar*. Stanford: CSLI publications.
- Pustejovsky, J. 1995. *The Generative Lexicon*. Cambridge, MA: MIT Press.

Investigating interfaces: an experimental approach to focus in Sicilian Italian

Raffaella Folli¹ and Elinor Payne²

¹ School of Communication, University of Ulster, UK

² Department of Phonetics and Linguistics, University College London, UK

Abstract

This paper describes an investigation of the syntax-prosody interface in the expression of focus in Sicilian Italian. It finds evidence of a generalised reversal of the articulation of information structure, with new information often preceding old, and of intonational differentiation between focus types.

Introduction

A pressing question in current linguistic theory is how different components of language interact. Our investigation considers the prosody and syntax interaction in the expression of focus. A major challenge for interface studies is reconciling differences in methodology and theoretical approach inherent in each discipline. This is particularly true for an investigation of focus, a phenomenon particularly difficult to elicit under experimental controls.

Our study examines focus in two varieties of Sicilian Italian. This variety is of particular interest because, although the ‘default’ word order (i.e. for ‘broad focus’) is VO, as in other varieties of Italian, a very frequent OV order is attested with the expression of New Information Focus (see Poletto and Benincà, 2002 and Cruschina, 2006). This again contrasts with Italian where the typical partition of a sentence sees the old information preceding the new information (see example 1), while the preverbal position is typically available only for Contrastive Focus (see example 2):

- (1) A. Cosa mangia Gianni?
‘What does John eat?’
B. Gianni mangia una mela
‘John eats an apple’
- (2) A. Gianni mangia una pera?
John eats a pear?
No, una mela mangia
No, an apple he eats

It has been argued that in Sicilian Italian a preverbal position can be targeted not only by contrastively focused material, but also by new information focus.

For example, Cruschina (2004: 33) reports that in the dialect of Mussomeli (in the province of Caltanissetta) the following order is attested:

- (3) A. Chi successi?
What happened?
B. Ruppi a seggia (Broad Focus)
He broke the chair
- (4) A. Chi ruppi Salvo?
What did Salvo break?
B. A seggia ruppi (New Information Focus)
The chair he broke
C. A SEGGIA ruppi (Contrastive Focus)
THE CHAIR he broke

(4) shows that the preverbal position can be occupied both by the constituent carrying new information and contrastive focus, giving rise to an OV order in (4)B which is unattested in Standard Italian. Cruschina claims that, though these two types of focus are semantically distinct, in Sicilian they are indistinguishable syntactically. He claims, from impressionistic analysis, that the two are, nevertheless, prosodically distinct (i.e. they have a different intonational contour) and argues that this distinction at the prosodic level confirms the existence of two distinct syntactic positions.

We conducted a detailed instrumental analysis of these two types of focus to investigate whether such a prosodic difference actually exists and gain insight into whether or not they can be conflated into a unified syntactic position (Frascarelli, 2000). Our investigation also extended to other kinds of syntactic environment, varying verb type (transitives, unaccusatives, unergatives) to test for sensitivities at the interface to the different relationship between phrase constituents.

Experiment

Material

The material consisted of sentences constructed to provide data that conformed both to syntactic/semantic requirements and phonetic/phonological conditions (e.g. all segments voiced, to allow tracing of pitch contour).

Procedure

Subjects

The 12 subjects were all students at Catania University¹, from *Catanese* families, and familiar with dialect. Subjects were chosen on the basis of their

performance in an initial test, conducted to ascertain how naturally they produced OV order in the relevant context.

Elicitation technique

Subjects were presented with a series of questions, and asked to construct answers using words printed on small cards which were placed in front of them, in random order. For example, for the question ‘*cosa fa Anna?*’, subjects saw the cards ‘MANGIARE’ (eat) + ‘ANNA’ + ‘MELA’ (apple).

To elicit the most natural responses possible, one experimenter (from Catania) was dedicated to asking the questions. This was to avoid any interference that hearing non-Sicilian accents could import, and to focus her attention on building a rapport with the subject. Another experimenter manipulated the priming cards, and another controlled the recording.

Results

Preliminary results from an analysis using PRAAT reveal intonation differences between new information and contrastive focus, for focus on subject, verb or object. Figure 1 shows this difference for focus on unergative verbs.

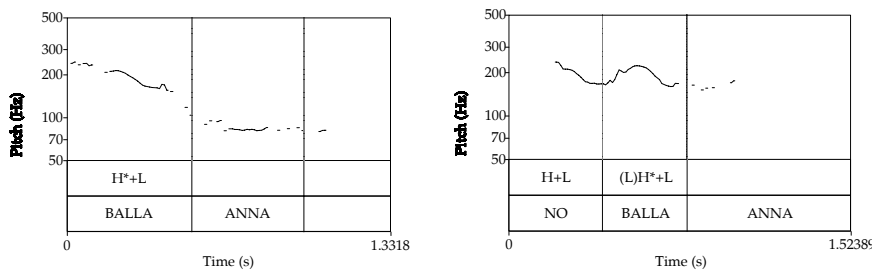


Figure 1. Pitch contours of new information focus on *balla* (left) and contrastive focus (right) on *balla*

New Information Focus on *balla* (‘dances’) attracts a FALL (H*+L), aligned with the beginning of the stressed syllable. When focus is contrastive, the H* is aligned later in the stressed syllable, indicating the same underlying phonological sequence (H*+L), but different phonetic alignment. Some evidence indicates also a preceding L, which would be interpretable either as a pragmatically specific Phrase Accent (L H*+L), or part of what is actually a tritonal, RISE-FALL Nuclear Pitch Accent (L+H*+L).

For subjects, New Information Focus appears to have the same basic pitch accent (H*+L) as Broad Focus. The difference is that in the latter, this is a pre-nuclear pitch accent, and the nucleus falls at the end of the phrase (e.g. (H*+L) *Anna ha il* (H*+L) *libro*), while in the former the pitch accent on *Anna* is nuclear and everything else that follows is de-accented (e.g.

(H*+L) *ANNA ha il libro.*). As with verbs, contrastive focus on subjects attracts what is either a RISE-FALL (L+H*+L), or an L Phrase Accent (L H*+L).

Preliminary conclusions

The evidence points to a reversal in the articulation of information structure in this variety when compared with Standard Italian: often in *Catanese* Italian new information precedes old information. This appears to be true not just for object-initial sentences, but also when other constituent types are fronted (e.g. subjects or verbs).

From an initial analysis, it would appear that the two types of focus, new information and contrastive, are indeed distinguishable prosodically at the production level, though this distinction remains to be confirmed perceptually. It also remains to be shown whether this distinction lies within the Nuclear Pitch Accent itself, or is expressed through a form of Phrase Accent associated with contrast. It is noted that some form of pragmatic markedness is associated with the left peripheral new information focus, whereby this order is often employed to reconfirm or strongly emphasise a piece of new information. This opens the question, which we are currently investigating, as to whether these two types of focus target two distinct syntactic positions in the left peripheral C field, or whether the prosodic difference is the only cue to interpretation. In other words, could this be an example of a trading relation between intonation and syntax?

Notes

- ¹ Recordings were also made in Palermo, where it proved far more difficult to elicit OV order, and these data will serve principally as a benchmark for the intonation analysis. The confounding factor was most likely sociolinguistic, with perceived greater aspirations among subjects to standardise their speech.

Acknowledgements

This research was funded by a British Academy Small Research Grant. We would also like to thank Sebastiano Grasso and Maria Catena Costa from Catania University and Mari D'Agostino from Palermo University for their kind assistance with recordings.

References

- Cruschina, S. (forthcoming), Informational focus in Sicilian. To appear in M. Frascarelli (ed.), *Phases of Interpretation*. Berlin, Mouton.
- Frascarelli, M. 2000. *The Syntax-Phonology Interface in Focus and Topic Constructions in Italian*. Kluwer Academic Publishers.
- Poletto, C. and Benincà, P. 2002. Topic, Focus and V2: defining the CP sublayers, in L. Rizzi (ed.) (2004) *The structure of IP and CP*. Vol. 2, OUP.

A corpus based analysis of English, Swedish, Polish, and Russian prepositions

Barbara Gawronska¹, Olga Nikolaenkova^{1,2} and Björn Erlendsson¹

¹ School of Humanities and Informatics, University of Skövde, Sweden

² Department of Linguistics, University of Athens, Greece

Abstract

In this study, the use of most frequent spatial prepositions in English, Polish, Swedish, and Russian is analyzed. The prepositions and their contexts are extracted from corpora by means of concordance tools. The collocation strength between the prepositions and the most frequent nouns in the PPs (Gries et al. 2005) is then computed in order to get a more detailed picture of the contexts in which a given preposition is most likely to appear. The results of the investigation are then analysed within the framework of cognitive semantics, especially Croft and Cruse's (2004) taxonomy of construal operations, and Talmy's (2005) classification of spatial images.

Background and aim

Prepositions define relations between objects, or, rather, conceptualizations of objects. In order to define their meaning it is necessary not only to describe the relation the preposition expresses but also the objects involved. Still, a description based on geometrical notions (the dimensionality of the objects) does not cover all aspects of the semantics of prepositions. Croft and Cruse (2005) propose a model that enriches the geometrical descriptions with construals as focus, scale of attention, perspective and viewpoint.

The main difficulty in cross-linguistic description of preposition semantics is due to the fact that cross-language differences in prepositional systems increase as we move from physical senses of prepositions into the metaphorical extensions of their meaning. The meaning chains (Brugman 1981, Gawronska 1993, Taylor 1988) have different shapes in different languages. Another difficulty lies in different degrees of lexicalization and in formulating criteria for regarding prepositions as parts of lexicalized multiword entries. These problems are of central importance for language technology application, especially Machine Translation (MT).

In the present study, we investigate whether the collocation analysis, proposed by Gries et al. (2005), may be helpful in identification of nouns and noun classes that tend to co-occur with certain prepositions in 4 different languages, and whether the values of the collocation strength may contribute to a better understanding of similarities and differences among prepositional systems. Another question we are concerned with is whether and how the results may improve the treatment of prepositions in MT.

Method and results

The frequency of prepositions in English, Swedish, Polish, and Russian was investigated using large corpora:

- A corpus of modern English prose (31 234 174 word tokens)
- A corpus of modern Swedish prose (43 634 620 word tokens)
- A corpus of modern Polish prose (49 478 901 word tokens)
- A corpus of modern Russian (25 000 000 word tokens)

The English, Swedish, and Polish corpora were obtained from LexwareLabs AB (www.lexwarelabs.com), and the Russian one – from the Russian Academy of Science. PPs with three most frequent spatial prepositions for each language were subjects for further investigation. The preposition were: *in, at, on* (English), *i, på, till* (Swedish), *w, na, przez* (Polish), and *в, на, через* (Russian).

The colostrucural strength between the prepositions and the nouns was computed according to the formula in Figure 1. The variables are explained in Table 1. Tables 2-3 show the results for the ten most frequent nouns co-occurring with with the English and Polish prepositions investigated here.

$$CS = -\text{Log}_{10} \left(\frac{\binom{(a+c)}{a} \times \binom{(b+d)}{b}}{\binom{N}{(a+b)}} \right)$$

Figure 1. The formula for calculation of colostrucural strength (CS).

Table 1. An example showing the values of the variables in Figure 1

	Construction C	Other Constructions	Row Totals
Word = bank	a = f(på + bank)	b = f(other P + bank)	a+b
Other Words	c = f(på+other noun)	d = f(other P+other noun)	c+d
Col. Totals	a+c	b+d	N=a+b+c+d

A comparison of the obtained CS-values showed that the following nouns and noun categories displayed either both high frequency value and high (> 0.6) colostrucural strength, or high collocational strength in at least three of the four languages: WORLD, EARTH, COUNTRY, NET/WEB/INTERNET, WAY/ROAD, TIME, TIME PERIOD, TIME BOUNDARY (start/end), AUTHORITIES, TEXT/NEWSPAPERS/LITERATURE SUBJECT/ MATTER, HEAD.

Table 2. English: Ten most frequent noun co-occurring with in, on and at in P+N and P+Det+N phrases. F= frequency in thousands, CS= collostructural strength.

In (90 093 occurrences)			On (20 119 occurrences)			At (32 262 occurrences)		
Noun	F	CS	Noun	F	CS	Noun	F	CS
world	4,13	2,90	deck	5,57	8,25	moment	3,03	3,57
way	3,57	2,53	account	1,63	1,80	length	2,99	2,74
spite	2,51	2,30	board	1,51	1,80	home	2,89	2,74
front	2,27	2,30	earth	1,00	1,80	end	2,34	2,74
fact	2,37	1,86	contrary	1,19	1,80	rate	1,42	1,60
morning	2,49	1,86	subject	1,83	1,51	night	1,52	1,31
midst	1,69	1,15	ground	1,39	1,34	door	1,66	1,14
house	2,08	1,14	side	1,97	1,34	time	0,82	0,06
love	1,75	0,64	foot	0,62	0,01	hand	0,82	0,02
time	1,96	0,58	deck	5,57	8,25	work	0,73	0,01

Table 3. Polish: Ten most frequent nouns co-occurring with w, na, and przez

w 388 056 occurrences			na 154 907 occurrences			przez 41 067 occurrences		
noun	F	CS	noun	F	cs	noun	F	CS
case	24,5	3,74	ground	23,0	14,20	sms	6,44	7,16
service	23,0	3,38	sake	10,0	3,92	authorities	2,40	1,63
end	17,2	3,31	territory	8,07	3,39	person	2,34	0,99
august	8,82	1,54	subject	7,55	3,41	moment	3,06	0,93
content	6,68	1,49	example	4,18	2,38	period	1,05	0,45
matter	15,7	1,32	earth	7,30	1,70	author	0,81	0,43
Google	6,91	1,15	side	8,21	1,39	people	1,57	0,42
case	9,28	1,07	conclusion	5,70	1,38	head	0,41	0,31
newspapers	6,93	1,05	principle	4,38	1,22	moment	0,39	0,06
goal	17,8	0,80	world	5,27	1,07	agency	0,66	0,03

Conclusions

Our results confirm Gries' et al.(2005) claim that the collocstructural strength value outperforms raw frequency data in corpus-based analysis. For example, although the Polish and Swedish nouns TIME are not among the 10 most frequent nouns after *på/na*, the CS-values between *på* and TIME and *na* and TIME are higher than the values of the 10th most frequent nouns co-occurring with these prepositions, which is intuitively correct. Nevertheless, a high CS-value cannot be used for automatic selection of translation equivalents in Machine Translation without further refinement. Both

Swedish *i* and English *in* have high CS-values in connection to the noun MORNING, but the Swedish phrase *i morgon* is equivalent to *tomorrow*. Collocations with the lexeme TIME should be coded in the lexicon as patterns like:

P + TIME (= the noun "time")+ [viewpoint]

The same is true about the collocations with the categories TIME PERIOD, TIME BOUNDARY.

Furthermore, our analysis reveals certain different conceptualizations of common-experience concepts:

WORLD is a 'surface' in Polish and a 'container' in Swedish, Russian, and English

WAY/ROAD is 2-dimensional both in the spatial and the metaphorical sense in Swedish and English (*vara på väg*, be on the way); however, in English it is 2-dimensional if the travellers viewpoint is preserved, and 3-dimensional from an outside perspective (in this way). In Polish, the situation is opposite: WAY is 3-dimensional from the traveller's point of view (*jestem w drodze* – 'I am on the way'), and 2-dimensional otherwise.

Low collocation values (<0.5) seem to indicate either valence-boundness of the type V + P or A + P, or a particular syntactic construction on sentence level (e.g. passive; cf. the results for the Polish *przez*, which is used as agent marker in passive). This hypothesis has to be tested in further research.

References

- Brugman, Claudia M. 1981, Story of OVER, Master's thesis, university of California, Berkeley. Trier: LAUT 1983.
- Croft W. and Cruse D. A. 2004: Cognitive linguistics. Cambridge: Cambridge University Press.
- Gawronska, B. 1993. Entailment in Logic and in the Lexicon. In: Martin-Vide, C. Current Issues in Mathematical Linguistics. Amsterdam: Elsevier, pp. 239-248.
- Gries, S.Th, Hampe, B., and Schönefeld, D. 2005. Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. Cognitive Linguistics 16-4, pp. 635-676.
- Talmy, L. 2005, The fundamental system of spatial schemas in language, In Hampe, B. (ed.), From perception to Meaning, Berlin/New York: Mouton de Gruyter, pp 199-234.
- Taylor, J. R. 1988, Contrasting Prepositional Categories: English and Italian, In Rudzka-Ostyn, B. (ed.), Topics in Cognitive Linguistics, Amsterdam/Philadelphia, John Benjamins Publishing Company, pp 299-326.

Evaluation of a virtual speech cuer

Guillaume Gibert, Gérard Bailly and Frédéric Elisei
Institut de la Communication Parlée, INPG/Univ. Stendhal, Grenoble Cedex,
France

Abstract

This paper presents the virtual speech cuer built in the context of the ARTUS project aiming at watermarking hand and face gestures of a virtual animated agent in a broadcasted audiovisual sequence. For deaf viewers that master cued speech, the animated agent can be then incruised - on demand and at the reception - in the original broadcast as an alternative to subtitling. The paper presents the multimodal text-to-speech synthesis system and the first evaluation performed by deaf users.

Introduction

Listeners with hearing loss and orally educated typically rely heavily on speechreading based on lips and face visual information. However speechreading alone is not sufficient due to the lack of information on the place of tongue articulation and the mode of articulation (nasality or voicing) as well as to the similarity of the lip shapes of some speech units (so called labial *sosies* such as [u] vs. [y] for French). Indeed, even the best speechreaders do not identify more than 50 percent of phonemes in nonsense syllables (Owens and Blazek 1985) or in words or sentences (Bernstein, Demorest et al. 2000). Cued Speech (CS) was designed to complement speechreading. Developed by Cornett (1967; 1982) and adapted to more than 50 languages (Cornett 1988), this system is based on the association of speech articulation with cues formed by the hand. While uttering, the speaker uses one of his hand to point out specific positions on the face with a hand shape). Numerous studies have demonstrated the drastic increase of intelligibility provided by CS compared to speechreading alone (Nicholls and Ling 1982) and the effective facilitation of language learning using CS (Leybaert 2000; Leybaert 2003). A large amount of work has been devoted to CS perception but few works have been devoted to CS synthesis. We describe here a multimodal text-to-system driving a virtual CS speaker and its first evaluation by deaf users.

The multimodal text-to-speech system

The multimodal text-to-speech system developed in the framework of the ARTUS project converts a series of subtitles into a stream of animation pa-

rameters for the head, face, arm and hand of a virtual cuer and an acoustic signal. The control, shape and appearance models of the virtual cuer (see Figure 1) have been determined using multiple multimodal recordings of one human cuer. The different experimental settings used to record our target cuer and capture its gestures and the complete text-to-Cued speech system are described in our JASA paper (Gibert, Bailly et al. 2005).



Figure 1: Chronogram for the word “Bonjour” pronounced and cued by our virtual speaker.

Evaluation

A first series of experiments have been conducted to evaluate the intelligibility of this virtual cuer with skilled deaf users of the French cued speech. The first evaluation campaign was dedicated to segmental intelligibility and the second one to long-term comprehension.

Segmental intelligibility

This test was conducted to assess the contribution of the cueing gestures in comparison with lip reading alone.

Minimal pairs The test mirrors the Modified Diagnostic Rime Test developed for French (Peckels and Rossi 1973): the minimal pairs do not here test acoustic phonetic features but gestural ones. A list of CVC word pairs has thus been developed that test systematically pairs of consonants in initial positions that differ almost only in hand shapes: we choose the consonants in all pairs of 8 subsets of consonants that are highly visually confusable (Summerfield 1991). The vocalic substrate was chosen so as to cover all potential hand positions while the final consonant was chosen so that to avoid rarely used French words or proper names, and test our ability to handle coarticulation effects. Due to the fact that minimal pairs cannot be found in all vocalic substrates, we end up with a list of 196 word pairs.

Conditions Minimal pairs are presented randomly and in both order. The lipreading-only condition is tested first. The cued speech condition is then presented in order to be able to summon up cognitive resources for the most difficult task first.

Stimuli In order to avoid a completely still head, head movements of the lipreading-only condition are those produced by the text-to-cued speech synthesizer divided by a factor of 10 (in fact the head accomplished on average 16.43% of the head/hand contact distance). We did not modify segmental nor suprasegmental settings that could enhance articulation.

Subjects Height subjects were tested. They are all hearing impaired people who have practised French CS (FCS) since the age of 3 years.

Results Mean intelligibility rate for “lipreading” condition is 52.36%. It is not different from haphazard way of response that means minimal pairs are not distinguishable. Mean intelligibility rate for “CS” condition is 94.26%. The difference in terms of intelligibility rate between these two conditions shows our virtual cuer gives significant information in terms of hand movements. In terms of cognitive efforts, the “CS” task is easier: the response time is significantly different $F(1,3134)=7.5$, $p<0.01$ and lower than for the “lipreading” one.



Figure 2: Eye gaze for one subject captured during the comprehension test using an eye tracker system: (left) teletext, (right) video incrustated.

Comprehension

To evaluate the global comprehension of our system, we asked the same 4 subjects to watch a TV program where subtitles were replaced by the incrustation of the virtual cuer. Ten questions were asked. The results show all the information is not perceived. On average, the subjects replied correctly to 3 questions. The difficulties of the task (proper names, high speaking rate) could explain these results. We conducted further experiments using a Tobii eye tracker. We asked 4 deaf people to watch a TV program divided in 2 parts: one part subtitled and another part with the inlay of a cuer video. The results show the subjects spend 56.36% of the time on the teletext and 80.70% on the video of the cuer with a significant difference $F(1,6)=9.06$, $p<0.05$. A control group of 16 hearing people spend 40.14% of the time reading teletext. No significant difference was found.

Conclusions

The observation and recordings of CS in action allow us to implement a complete text-to-Cued Speech synthesizer. The results of the preliminaries perceptive tests show significant linguistic information with minimal cognitive effort is transmitted by our system. This series of experiments must be continued on more subjects and other experiments must be added to quantify exactly the cognitive effort used. Discourse segmentation and part of speech emphasis by prosodic cues (not yet implemented) is expected to lighten this effort.

Acknowledgements

The authors thank Martine Marthouret, Marie-Agnès Cathiard, Denis Beautemps and Virginie Attina for their help and comments on building the perceptive tests. We also want to thank the 25 subjects who took part to the evaluation.

References

- Bernstein, L. E., M. E. Demorest and P. E. Tucker 2000. Speech perception without hearing. *Perception & Psychophysics* **62**: 233-252.
- Cornett, R. O. 1967. Cued Speech., *American Annals of the Deaf* **112**: 3-13.
- Cornett, R. O. 1982. Le Cued Speech. Aides manuelles à la lecture labiale et perspectives d'aides automatiques. F. Destombes. Paris, Centre scientifique IBM-France.
- Cornett, R. O. 1988. Cued Speech, manual complement to lipreading, for visual reception of spoken language. Principles, practice and prospects for automation. *Acta Oto-Rhino-Laryngologica Belgica* **42** (3): 375-384.
- Gibert, G., G. Bailly, D. Beautemps, F. Elisei and R. Brun 2005. Analysis and synthesis of the 3D movements of the head, face and hand of a speaker using cued speech. *Journal of Acoustical Society of America* **118** (2): 1144-1153.
- Leybaert, J. 2000. Phonology acquired through the eyes and spelling in deaf children. *Journal of Experimental Child Psychology* **75**: 291-318.
- Leybaert, J. 2003. The role of Cued Speech in language processing by deaf children: an overview. *Auditory-Visual Speech Processing*, St Jorioz - France: 179-186.
- Nicholls, G. and D. Ling 1982. Cued Speech and the reception of spoken language. *Journal of Speech and Hearing Research* **25**: 262-269.
- Owens, E. and B. Blazek 1985. Visemes observed by hearing-impaired and normal-hearing adult viewers. *Journal of Speech and Hearing Research* **28**: 381-393.
- Peckels, J. P. and M. Rossi 1973. Le test de diagnostic par paires minimales. Adaptation au français du 'Diagnostic Rhyme Test' de W.D. Voiers. *Revue d'Acoustique* **27**: 245-262.
- Summerfield, Q. 1991. Visual perception of phonetic gestures. Modularity and the motor theory of speech perception. I. G. Mattingly and M. Studdert-Kennedy. Hillsdale, NJ, Lawrence Erlbaum Associates: 117-138.

Broad vs. narrow focus in Greek

Stella Gryllia
LUCL, Leiden University, The Netherlands

Abstract

This paper reports the results of a production and a perception experiment about focus in Greek.

Introduction

Let's look at the example in (1).

O Kostis kerdise to lahio.
the-NOM Kostis-NOM win-3SG the-ACC lottery-ACC
'Kostis won the lottery.'

The word order SVO (ex.1) is a felicitous answer to an all focus question, '*What's up?*', a verb phrase focus question, '*What did S do?*' and a object focus question, '*What V S?*'. Thus, example (1) is multiply ambiguous and its focus breadth varies. The focus' breadth depending on the preceding question varies among [SVO]_F, S[VO]_F and SV[O]_F. The focus' breadth [SVO]_F can be described as broad, whereas SV[O]_F as narrow. However, broad and narrow focus are relative terms. Verb phrase focus S[VO]_F is broad compared to SV[O]_F and narrow compared to [SVO]_F. Narrow focus has often been associated with contrastive interpretation. In this paper, following Cohan (2000), the terms broad and narrow focus are taken to refer to the breadth of focus.

A question that emerges is whether the breadth of focus is reflected on the phonetic realization of the utterances. More specifically, do speakers produce a difference among sentence focus, verb phrase focus and object focus? Do listeners perceive any difference? To tackle these questions, a production and a perception experiment were performed.

Production experiment

In GrToBI, Arvaniti & Baltazani (2000) reported that narrow focus is signaled by a L+H* nuclear accent, whereas broad focus by a H* nuclear accent. We thus expect to find a difference among the three types of focus.

Methods

Stimuli. A list of 13 sets of four question-answer (QA) pairs was constructed. For the first three QA pairs, the word order of the answer was kept constant, it

was namely SVO, whereas the question varied. There were three types of questions: an all focus (2a), a verb phrase focus (2b) and an object focus (2c). In the fourth QA pair the question was an object focus question (2d) and the word order of the answer was OVS. A sample is given in (2).

2a	ti jinete; 'What's up?'	[_S I Eleni meloni mila.] _F 'Helen smears honey on apples.'
2b	Ti kani i Eleni; 'What does Helen do?'	I Eleni [_{VP} meloni mila] _F 'Helen smears honey on apples.'
2c	Ti meloni i Eleni; 'On what does Helen smear honey?'	I Eleni meloni [_{NP} mila] _F 'Helen smears honey on apples.'
2d	Ti meloni i Eleni; 'On what does Helen smear honey?'	[_{NP} Mila] _F meloni i Eleni. 'On apples, Helen smears honey.'

Procedure. A self-paced stimulus presentation was used. Utterances were directly recorded via a head-mounted close-taking microphone (Shure SM10A) on computer disk using an Abode Audition Software. Forty native speakers of Athenian Greek participated in the experiment.

Analysis

Four sets out of the total 13 were analyzed. The productions of all 40 participants (640 utterances) were analyzed, using Praat. The autocorrelation pitch extraction method (Boersma 1993) was used to determine the fundamental frequency (F0) as the primary correlate of vocal pitch. All F0 curves in the materials were stylized in this way. Afterwards, pivot points were defined. For each utterance, ten pairs of time points and their correspondent pitch ($t_1, p_1 \dots t_{10}, p_{10}$) were obtained. The time-frequency coordinates of the pivot points were automatically extracted and stored in a database for off-line statistical processing. Thereafter, one-way analyses of variances were performed. The independent variable was the type of focus. The dependent variable was the frequency of one specific pivot point. Ten one-way analyses of variance were run, one for each of the ten pivot points.

Results

Comparing the results of ten different one-way ANOVAs for four types of focus with the results of ten different one-way ANOVAs for three types of focus, excluding the cases of preverbal object focus, it can be concluded that speakers do not produce any significant differences with respect to sentence focus, verb phrase focus and postverbal object focus. The difference Δ between p_3 (pitch peak on the first content word) and p_6 (pitch peak on the second content word) was established. Δ corresponds to the downstep in pitch between p_3 and p_6 . 160 accentual downsteps per focus type were obtained, stored in a database for off-line statistical processing and compared by per-

forming paired t-tests. The downstep in sentence focus differs significantly from the downstep in verb phrase focus. The same holds for the downstep in $[_S\text{SVO}]_F$ and $\text{SV}[_{\text{NP}}\text{O}]_F$. The downstep in verb phrase focus differs significantly from preverbal object focus. The same holds for the downstep in postverbal object focus and preverbal object focus. It should be noted that the downstep in verb phrase focus does not differ from the downstep in postverbal object focus. The size of the accentual downstep was also analyzed as a function of the four focus types per gender. The accentual downstep in $[_{\text{NP}}\text{O}]_F$ VS is large, namely 22Hz by females and 25Hz by males. In contrast to the large downstep in preverbal object focus, the accentual downstep in $\text{S}[_{\text{VP}}\text{VO}]_F$ is small, 6Hz by females and 7.4Hz by males. Female and male speakers differ with respect to the accentual downstep in $[_S\text{SVO}]_F$ and $\text{SV}[_{\text{NP}}\text{O}]_F$. In $[_S\text{SVO}]_F$ the female speakers downstep by 26Hz, whereas the male speakers by 7Hz. In $\text{SV}[_{\text{NP}}\text{O}]_F$ the female speakers downstep by 15Hz, whereas the male speakers' downstep is 0Hz. Four ANOVAs were run to evaluate the differences between male and female speakers. The independent variable was gender. The dependent variable was the accentual downstep. Female and male speakers differ significantly with respect to the accentual downstep in $[_S\text{SVO}]_F$. The difference in accentual downstep between female and male speakers in $\text{SV}[_{\text{NP}}\text{O}]_F$ is marginally significant.

Perception experiment

Methods

Stimuli. 24 stimuli produced by a male and a female speaker, who participated in the production experiment were used. The set of twelve stimuli was the same for the male and the female speaker. These twelve stimuli consisted of four sets of three sentences: $[_S\text{SVO}]_F$, $\text{S}[_{\text{VP}}\text{VO}]_F$, $\text{SV}[_{\text{NP}}\text{O}]_F$.

Procedure. The 24 stimuli were made audible with a fixed interstimulus interval of 0.3sec (offset-onset). Listeners were supplied with an answering sheet containing a list of questions in sets of three. Each set contained a sentence focus, a verb phrase focus and an object focus question. Listeners were instructed to tick off the question which according to them corresponded best to the declarative sentence they were listening to. Forty native speakers of Greek, twenty females and twenty males participated in the experiment. These forty speakers had not participated in the production experiment.

Results

960 responses were analyzed. Listeners seem to perceive some differences among the types of focus. Sentence focus is perceived below chance level, verb phrase focus is perceived just above chance level, while postverbal object focus is perceived well above chance level. More specifically, when the focus intended by the speakers was sentence focus, then 14.1% perceived it as such. When the intended focus was verb phrase focus, it was correctly

perceived by 42.2% of the listeners. When the intended focus was postverbal object focus, 74.7% of the listeners perceived it correctly. The distribution of responses differs significantly across focus types also in terms of incorrect responses. When the intended focus type is postverbal object focus, then sentence focus is hardly ever chosen as a response. However, when the intended focus type is verb phrase focus, then the distribution of responses is much more balanced. More specifically, out of 960 utterances, sentence focus was chosen as a response 80 times, i.e. 8.3%, while verb phrase focus was chosen as a response 304 times, i.e. 31.7% and postverbal object focus was chosen as a response 576 times, i.e. 60%. These results show that there is a preference for choosing postverbal object focus as an answer and a dispreference for sentence focus. This preference and dispreference might be interpreted as a response bias. However, the preference for postverbal object focus might not be related to the acoustic properties of the stimuli. Crain et al. (1994) have experimentally shown that adults follow the least effort strategy for ambiguity resolution, reducing the risk of making commitments that will need to be changed later. In this sense, the dispreference for sentence focus is not so surprising. When sentence focus is selected as a response, then it coincides with the focus intended by the speaker at 14.1%. This percentage is almost double than the incorrect response verb phrase focus (7.2%). When verb phrase focus is selected as a response, then it coincides with the focus intended by the speaker at 74.7%. This is 20% higher than the incorrect response sentence focus.

Acknowledgements

Special thanks to Vincent van Heuven. I wish to acknowledge Lingua, Elsevier and LUF for financial support. Parts of the production experiment will appear in Proceedings of CamLing 2006.

References

- Arvaniti, A. and Baltazani, M.. 2000. GREEK TOBI: A System for the Annotation of Greek Speech Corpora. In Proceedings of 2nd International Conference on Language Resources and Evaluation Vol. 2: 555-62.
- Boersma, P. 2003. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In IFA Proceedings 17: 97-110.
- Boersma, P. and Weenink D. 2005. Praat: doing phonetics by computer.
- Cohan, J. 2000. The Realization and Function of Focus in Spoken English. PhD dissertation, University of Texas.
- Crain, S.W.Ni, and Conway, L. 1994 Learning, parsing and modularity. In C. Clifton, L. Frazier and K. Rayner (eds.) Perspectives on sentence processing. Lawrence Erlbaum: 443-467.

Incremental interpretation and discourse complexity

Jana Häussler and Markus Bader
Linguistics Department, University of Konstanz, Germany

Abstract

We will present two self-paced reading studies that provide new evidence on the incrementality of semantic interpretation, in particular with regard to the notion of discourse complexity as introduced by Gibson's distance-based dependency locality theory (DLT; cf. Gibson, 1998, 2000). More specifically, we will focus on the contribution of referential processing on sentence complexity. The experiments compared the processing of simple definite DPs like *der Lehrer* ('the teacher') and complex DPs containing a possessive NP like *Peters Lehrer* ('Peter's teacher'). While simple DPs introduce only a single discourse referent, complex DPs introduce two discourse referents and some relation between them. This additional processing effort is reflected by increased reading times.

Introduction

In prior work (Bader and Häussler 2005) we investigated simple and complex DPs in experiments focussing on the processing of number ambiguous nouns in German. A number ambiguous noun like *Lehrer* is usually disambiguated by a determiner (*der Lehrer* 'the teacher' vs. *die Lehrer* 'the teachers'), but remains ambiguous when specified by a possessive proper name (*Peters Lehrer* 'Peter's teacher(s)'). We found that number ambiguous nouns cause garden-path effects on the clause-final disambiguating verb. Reading times on the verb correlate substantially with number preferences determined by a corpus analysis. The more often a noun occurs in the singular, the more often it is analyzed as a singular noun during first-pass parsing, leading to difficulties when encountering a verb specified for plural.

Furthermore, reading times at the noun were longer in ambiguous DPs than in unambiguous DPs. As an alternative or in addition to ambiguity, the increased reading times might be attributed to complexity. While unambiguous DPs introduce only a single discourse referent, ambiguous (= possessive) DPs introduce two discourse referents and some relation between them. This might cause additional processing effort which is reflected in reading times.

Since complexity and ambiguity were confounded in these experiments we conducted two further experiments separating both factors by using two types of unambiguous material. Both experiments used a word-by-word non-cumulative self-paced reading procedure.

Experiment 1

Experiment 1 compares ambiguous nouns within a simple DP introduced by a determiner (*der eine Lehrer* 'the one teacher' and *die beiden Lehrer* 'the two teachers') and ambiguous nouns within a complex DP introduced by a possessive proper noun (*Peters Lehrer* 'Peters teacher(s)'). In contrast to the experiments reported above, the disambiguating verb precedes the ambiguous noun. A full set of examples is shown in (1).

- (1) a. Zum Glück war Peters Teilhaber mit der Klausel ebenfalls einverstanden.
to fortune was P.'s associate with the clause also agreeable
'Fortunately, Peter's associate also agreed with the clause.'
- b. Zum Glück war der eine Teilhaber mit der Klausel ebenfalls einverstanden.
to fortune was the one associate with the clause also agreeable
'Fortunately, the one associate also agreed with the clause.'
- c. Zum Glück waren Peters Teilhaber mit der Klausel ebenfalls einverstanden.
to fortune were P.'s associates with the clause also agreeable
'Fortunately, Peter's associates also agreed with the clause.'
- d. Zum Glück waren die beiden Teilhaber mit der Klausel ebenfalls einverstanden.
to fortune were the both associates with the clause also agreeable
'Fortunately, the two associates also agreed with the clause.'

Reading times on the ambiguous noun and thereafter in sentences like (1a) and (1c) were substantially longer than reading times on the noun in sentences like (1b) and (1d). Again, the increase already starts at the possessor (cf. Figure 1).

Since the verb precedes and therefore disambiguates the number ambiguous DP, we attribute these increased reading times to complexity instead of ambiguity.

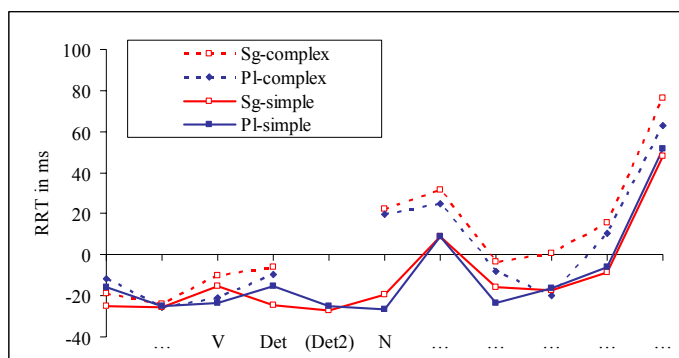


Figure 1. Residual Reading Times in Experiment 1.

Experiment 2

Experiment 2 compares complex and simple DPs with either ambiguous masculine nouns as in the experiments before or unambiguous feminine counterparts. The resulting DP is the subject of an embedded verb-final clause. A set with unambiguous nouns is given in (2). Ambiguous DPs contained the corresponding masculine noun *Teilhaber* ('associate').

- (2) Es hat sich jetzt herausgestellt, ... ('It turned out now')
- a. dass Peters Teilhaberin mit der Klausel ebenfalls einverstanden war
that P.'s associate (fem.) with the clause also agreeable was
'that Peter's associate also agreed with the clause'
 - b. dass die eine Teilhaberin mit der Klausel ebenfalls einverstanden war
that the one associate (fem.) with the clause also agreeable was
'that the one associate also agreed with the clause'
 - c. dass Peters Teilhaberinnen mit der Klausel ebenfalls einverstanden waren
that P.'s associates (fem.) with the clause also agreeable were
'that Peter's associates also agreed with the clause'
 - d. dass die beiden Teilhaberinnen mit der Klausel ebenfalls einverstanden waren
that the both associates (fem.) with the clause also agreeable were
'that the two associates also agreed with the clause'

On the noun and thereafter, reading times for complex DPs were longer than for simple DPs, for both the ambiguous masculine nouns and the unambiguous feminine nouns. Since feminine DPs show the same specifier effect as masculine DPs, the increased reading times cannot be attributed to ambiguity but must rather reflect complexity differences between simple DPs and complex DPs. This supports the DLT's claim that introducing new discourse referents consumes processing resources and thereby slows down processing. Interpreting *Peter's associate* requires the integration of two new discourse referents and establishing a dependency between them whereas interpreting *the one associate* requires the integration of only one new discourse referent.

In addition we found a main effect of gender and an interaction of number and gender. Reading feminine nouns took longer than reading masculine nouns. Furthermore, reading a feminine plural noun took longer than reading a feminine singular noun. For masculine nouns no such number effect was observed. This is probably a frequency effect: the corpus frequency for the experimental masculine nouns is about 7 times higher than for feminine singular nouns and 14 times higher than for feminine plural nouns.

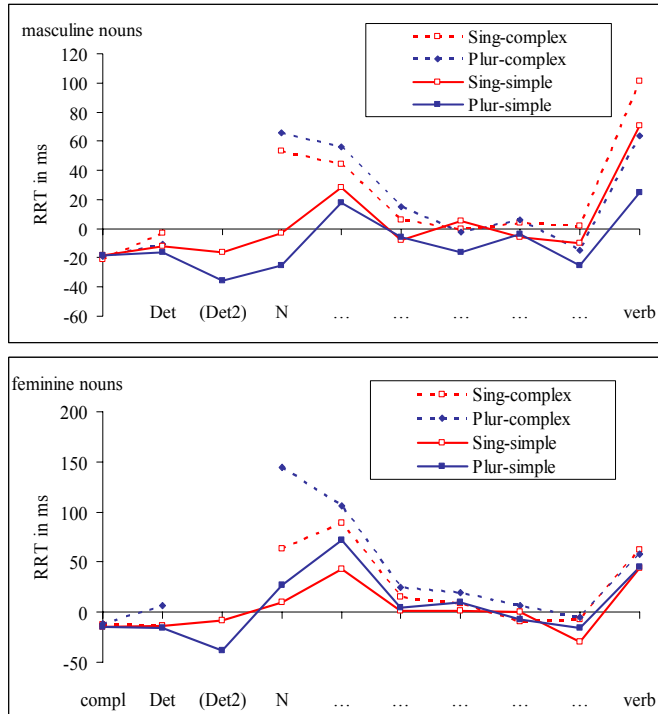


Figure 2. Residual Reading Times in Experiment 2

Summary and Conclusions

In sum, our results support the notion that discourse processing modulates on-line sentence complexity. Finding increased reading times immediately on the noun shows that syntactic structures are rapidly interpreted.

References

- Bader, M. and Häussler, J. 2005. World-Knowledge and Frequency in Resolving Number Ambiguities. Presented at the 11th Annual Conference on Architecture and Mechanisms for Language Processing, Ghent, Belgium.
- Gibson, E. 1998. Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, 1-75
- Gibson, E. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In Marantz, A., Miyashita, Y. and O'Neil, W. (eds.) 2000, *Image, language, brain. Papers from the first Mind Articulation Project Symposium*, 95-126. Cambridge, MIT Press.
- Warren, T. and Gibson, E. 2002. The influence of referential processing on sentence complexity. *Cognition* 85, 79-112.

Dynamic auditory representations and phonetic processing: The case of virtual diphthongs

Ewa Jacewicz, Robert Allen Fox and Lawrence L. Feth
Department of Speech and Hearing Science, Ohio State University, USA

Abstract

Auditory spectral integration in the perception of dynamic acoustic cues in speech was examined experimentally. The potential role of a dynamically changing center of gravity in the perception of diphthongs /ui/ (as in *we*) and /iu/ (as in *you*) was verified. Listeners identified the effective frequency changes well, showing that movement of the spectral center of gravity can provide the cues necessary for the identification of dynamic events in speech such as F2 transitions. Most models of vowel perception propose that vowels are identified on the basis of formant peaks. Our results indicate that perception of dynamic events in speech is to a large extent attributable to central auditory processes such as spectral integration.

Introduction

The concept of auditory spectral integration (ASI) refers to the improvement in detection or discrimination of complex sounds when their bandwidth exceeds a certain value (the “critical” band). Our present interest is in understanding how ASI functions in the coding of acoustic speech segments.

In speech perception, the effects of ASI have been studied primarily with reference to formants and perceived vowel quality. For example, early research demonstrated that two closely spaced formants could be matched to a single “intermediate” formant whose frequency depends on the specific relationship between the frequencies and amplitudes of the two close formants (Delattre *et al.*, 1952). The predictable shift in the matching frequency of the single formant occurs only within a larger bandwidth of about 3.5 bark (e.g., Chistovich and Lublinskaja, 1979). This center of gravity (COG) effect was interpreted as an indication of a central processing such as ASI.

A significant limitation of this early research was that the integration effects were examined only in static vowels. Auditory processing of these unnatural speech sounds is entirely focused on the frequency domain. However, human speech is inherently dynamic (in terms of both frequency and amplitude changes in time) and listeners are very sensitive to these dynamic changes. In a landmark study, Lublinskaja (1996) showed that the auditory system could attend to the dynamic spectral COG created by modifying relative formant amplitudes (but not formant frequencies) over time. The present paper assesses the efficacy of this “moving COG” in producing per-

ceived dynamic frequency changes to signal glide (i.e., diphthong) differences.

Experiment: Perception of dynamic cues

This experiment examined and verified the potential role of a dynamically changing COG in the perception of the diphthongs /ui/ (as in *we*) with a rising F2 transition and /iu/ (as in *you*) with a falling F2 transition.

Stimuli: Actual F2 transitions (FT)

Nine basic F2 contours in diphthong stimuli were created using HLSYN with the .kld option. There were three tokens with a steady-state F2, three with rising F2, and three with falling F2. For all stimuli, F1 remained at 300 Hz for the entire token. F2 onset was 1800, 2000 or 2200 Hz; F2 offset was 1800, 2000 or 2200 Hz (see Figure 1). There were three different durations (50, 100 and 150 ms) for each of the nine formant patterns. F0 remained steady at 100 Hz for the entire token.

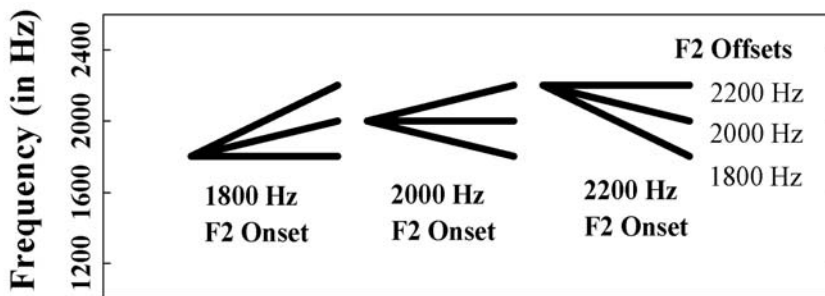


Figure 1. Schematic of nine basic F2 contours in diphthong stimuli. These basic formant patterns produced /i/, /u/, /ui/, or /iu/ percepts.

Stimuli: Virtual F2 transitions (VT or “virtual diphthongs”)

A base token was created from the FT series which contained F1 only. Two “resonances” were then created. The first contained the 17th and 18th harmonics only; the second contained the 22nd and 23rd harmonics only of steady-state F2s created using the parallel branch of the synthesizer. Harmonics were isolated using sharp FIR filters. The amplitudes of these resonances were modified as appropriate to allow the COG to follow the frequency changes to the F2 in the FT series. Overall rms of these sets of resonances was matched to F2s in the FT series and inserted into the base token.

Listeners and procedure

Six native speakers of American English aged 19-28 years listened to the signals over headphones while seated in a sound-attenuating booth. A single-interval 4AFC identification task was used with the response choices /i/ /u/ /ui/ and /iu/ displayed on the computer screen. The tokens were blocked by stimulus type (FT and VT). In each block there were 270 stimuli presented randomly in each session (3 durations x 9 formant patterns x 10 repetitions).

Results

Table 1. Percentage of /i/, /u/, /ui/ and /iu/ responses to each stimulus token for the steady-state (S), rising F2 (R), and falling F2 (F) series.

	Onset (Hz)	Offset (Hz)	/i/	/u/	/ui/	/iu/	/i/	/u/	/ui/	/iu/
			Actual F2 Transition				Virtual F2 Transition			
S	1800	1800	5.6	83.9	1.1	9.4	6.1	91.1	1.7	1.1
	2000	2000	63.9	19.4	6.1	10.6	44.4	45.0	9.4	1.1
	2200	2200	95.6	0.0	3.9	0.6	95.0	2.2	2.8	0.0
R	1800	2000	13.3	7.8	75.6	3.3	3.9	15.6	80.0	0.6
	1800	2200	2.8	0.6	95.6	1.1	0.6	0.0	99.4	0.0
	2000	2200	55.0	0.0	43.3	1.7	30.0	0.6	69.4	0.0
F	2000	1800	2.2	31.7	1.1	65.0	0.6	38.9	0.0	60.6
	2200	1800	3.9	12.8	0.0	83.3	0.0	1.1	0.0	98.9
	2200	2000	33.3	9.4	2.8	54.4	21.1	18.9	0.6	59.4

In both series, listeners most often identified the 1800 Hz steady-state tokens as /u/ and the 2200 Hz steady-state tokens as /i/ (see Table 1). A paired-samples t-test showed no significant difference as a function of stimulus type. However, for the 2000 Hz tokens, there was a significantly greater number of /i/ responses to the FT tokens than to the VT tokens showing that the amplitude variation in the two resonances for VT did not produce as high a perceived F2 as in the FT condition.

As expected, the percentage of /ui/ responses for both FT and VT was highest when the decrease in F2 frequency was the largest. An ANOVA of the number of /ui/ responses with the within-subject factors series (FT and VT) and token showed a significant effect of token ($F(2,10)=12.9, p=0.002, \eta^2=0.546$), but no significant main effect of series nor a significant series by token interaction.

The percentage of /iu/ responses for both FT and VT was highest when the increase in F2 frequency was the largest. An ANOVA of the number of /iu/ responses with the within-subject factors series (FT and VT) and token showed a significant main effect of token ($F(2,10)=7.99, p=0.008, \eta^2=0.615$). The number of /iu/ responses was significantly higher for the 2200-1800 Hz token, while the remaining two were not different.

Discussion and conclusions

The results show that listeners were equally sensitive to both the actual and the virtual frequency changes in making their vowel identifications. The differences between responses to the dynamic formant transitions and virtual transitions were not significant, indicating that movement of the spectral COG did provide the cues necessary for the identification of F2 transitions comparably with the actual formant transitions.

For both types of signals (i.e., FT and VT), the highest proportion of expected responses was obtained for the greatest frequency differences between the diphthongal onsets and offsets, which produced the clearest percepts of either /ui/ (the 1800-2200 Hz token) or /iu/ (the 2200-1800 Hz token). The proportion of the expected responses to spectral changes decreased with each smaller frequency separation between onsets and offsets, and the signals were identified as stationary vowels /i/ or /u/ when there was no frequency change.

Most models of vowel perception propose that vowels are identified on the basis of formant peaks. This approach will not work with the signals utilized here. Our approach to modelling is to examine dynamic auditory excitation patterns thought to result from the acoustic signals presented to the listener. Perception of dynamic events in speech is to a large extent attributable to central auditory processes such as auditory spectral integration explored here. Moreover, the perception of formants in vowels is almost certainly a result of the spectral integration of the energy of their harmonics.

Acknowledgements

Work supported by NIH R01DC00679-01A1 (L.L. Feth, PI). We thank Marc Smith for help with data collection.

References

- Chistovich, L. A. and Lublinskaja, V. V. 1979. The 'center of gravity' effect in vowel spectra and critical distance between the formants: psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research* 1, 185-195.
- Delattre, P., Liberman, A., Cooper, F. and Gerstman, L. 1952. An experimental study of the acoustic determinants of vowel color. *Word* 8, 195-210.
- Lublinskaja, V. V. 1996. The 'center of gravity' effect in dynamics. In Ainsworth, W. and Greenberg, S. (eds.), *Proc. of the Workshop on the Auditory Basis of Speech production*, 102-105, ESCA.

Syntactic abilities in Williams Syndrome: How intact is ‘intact’?

Victoria Joffe¹ and Spyridoula Varlokosta²

¹Department of Language and Communication Science, City University, UK

²Department of Mediterranean Studies, University of the Aegean, Greece

Abstract

The present paper investigates the syntactic abilities of a group of individuals with Williams Syndrome to explore the debates surrounding the dissociation between language and cognition as well as possible dissociations within the language module in this population. Experimental linguistic measures that assess interpretation of passives, pronouns and reflexives as well as elicitation and comprehension of wh-questions were employed. Overall results for the WS group reveal little evidence of the reported spared linguistic abilities in this population thus challenging the idea of a relatively intact language system in WS.

Background assumptions and aims

Neuro-developmental disorders like Williams Syndrome (WS) have been shown to present with a non-linear relationship between cognitive and linguistic abilities, showing better performance in language relative to cognition (Bellugi, Wang & Jernigan 1994, Clahsen & Almazan 1998). This has been used as evidence for a dissociation between language and cognition. WS has also been used to support the existence of submodular dissociations within the language faculty. It has been reported that individuals with WS perform better on grammatical tasks (passives, negation, interpretation of reflexives and pronouns, conditionals, formation of regular past tense and plurals) compared with tasks involving lexical processing (irregular past tense and plurals) (Bellugi et al. 1994, Clahsen & Almazan 1998, 2001).

However, the claims about both kinds of modularity evidenced in WS have not gone unchallenged. Whilst many researchers would agree that the linguistic abilities of individuals with WS are in advance of their cognitive skills (Mervis et al. 2004, among others), the claim of relatively intact linguistic processing has been contested by researchers that present evidence for impaired morphosyntactic abilities along with other linguistic and non-linguistic abilities (Karmiloff-Smith et al. 1997, among others). Also, there is currently a heated debate on whether any intra-linguistic dissociations exist in WS (Thomas et al. 2001).

In an attempt to further explore the debates surrounding WS, Joffe and Varlokosta (under review) report preliminary new results (from a larger on-

going study) on the verbal and non-verbal abilities of a group of nine WS individuals, aged 8.8 to 23.5 years, using a range of standardised and non-standardised measures. These preliminary results reveal little evidence of the reported spared linguistic abilities in WS as well little support for intra-linguistic dissociations. Any superior performance in rule-governed operations found in the WS cohort, was evident in the two control groups, Down Syndrome (DS) and Typically Developing (TD).

In the present paper we extend the report of Joffe & Varlokosta (under review) by providing preliminary results from a series of experimental measures that aimed to assess complex computational operations (such as movement) in the same group of WS individuals.

Method

Participants

Nine (2 female; 7 male) English-speaking individuals with WS, aged 8.8 to 23.5 years (mean age 12.5), with mean performance IQ of 49.7, participated in the study. Their performance was compared to two control groups: a chronological and cognitive age-matched DS group (9 [3 females; 6 males], aged 8.8 to 16.11 years, mean age 12.4, mean performance IQ 49.7) and a younger TD group (9 [3 females; 6 males], aged 3.8 to 7.11 years, mean age 4.3). Further details about the participants can be found in Joffe and Varlokosta (under review).

Materials and procedure

Syntactic abilities were assessed through experimental tasks that tested: (a) interpretation of passives: the task included six verbs presented twice in four sentence conditions (active transitive, full verbal passive, short progressive passive and ambiguous passive) yielding a total of 48 sentences; (b) interpretation of pronouns and reflexives: the binding task was a picture selection task with 48 sentences comprised of four conditions: name-pronoun, quantifier-pronoun, name-reflexive, quantifier-reflexive; (c) elicitation and comprehension of *wh*-questions: 16 sentences were tested (4 *who*-questions with a subject gap, 4 *who*-questions with an object gap, 4 *which-NP*-questions with a subject gap and 4 *which-NP*-questions with an object gap).

Testing was administered over approximately three-four sessions of around 40 minutes in length usually within a 4-week period.

Results and discussion

A series of one and two factor Anova's revealed no significant group differences with the WS group performing at the same level across all tasks as the cognitive age-matched group (DS). There is a trend for the WS group to obtain higher scores than the DS group on the binding, *wh*-elicitation and passive tasks. However, these differences were not found to be significant. No difference is evident on *wh*-comprehension with both groups scoring similarly. On the whole, WS and TD groups perform at similar levels (see Table 1).

Better performance was evident for WS and TD groups on the binding task compared with the other two comprehension tasks (passives and *wh*-questions). There was a significant main effect for group with planned comparisons revealing significantly better performance on *wh*-comprehension than *wh*-elicitation ($F(1,21) = 72.905, p < .001$). This finding follows the typical developmental trend of reception preceding expression. With respect to the binding task, the WS and TD group appear to perform better than the DS group. Nonetheless, our WS subjects performed worse than Clahsen & Almazan's (1998) subjects (mean chronological age 13.1 and mean IQ 53) on a similar task that assessed anaphoric dependencies. Similarly, our WS subjects performed worse than Clahsen & Almazan's (1998) subjects in the passive task, reflecting the heterogeneity of the disorder.

Table 1. Correct responses in percentages for all tasks.

Group	Binding	Passives	Wh-elic	Wh-comp
WS	72	49	17	58
DS	59	41	10	60
TD	74	60	21	59

Results for the WS group reveal little evidence of the reported spared linguistic abilities in this population thus challenging the idea of a relatively intact language system. However, patterns of performance may indicate some superiority in language for WS over DS, differences which may prove to be significant with an increase in subjects. Further WS data is being collected to explore these trends further. The language functioning of WS does appear to be complex and the debates are far from resolved.

Acknowledgements

The authors are grateful for the support of the Williams Syndrome Foundation in the UK, to all the WS, DS and TD participants and their families, and to Amy Riddett, who collected most of the data. The study was funded by an ESRC grant to both authors (RES-000-22-0656).

References

- Bellugi, U., Wang, P. and Jernigan, T. 1994. Williams syndrome: an unusual neuropsychological profile. In Broman, S. and Grafman, J. (eds.) 1994, *Atypical Cognitive Deficits in Developmental Disorders. Implications for Brain Function*, 23-56. Hillsdale NJ, Erlbaum Press.
- Clahsen, H. and Almazan, M. 1998. Syntax and morphology in Williams Syndrome. *Cognition* 68, 167-198.
- Clahsen, H. and Almazan, M. 2001. Compounding and inflection in language impairment: evidence from Williams Syndrome (and SLI). *Lingua* 111, 729-757.
- Joffe, V. and Varlokosta, S. under review. Language abilities in Williams Syndrome: exploring comprehension, production and repetition skills. *Advances in Speech-Language Pathology*.
- Karmiloff-Smith, A., Grant, J., Berthoud, I., Davis, M., Howlin, P. and Udwin, O. 1997. Language and Williams Syndrome: how intact is 'intact'? *Child Development* 68, 274-290.
- Mervis, C.B., Robinson, B., Rowe, M., Becerra, A.M. and Klein-Tasman, B.P. 2004. Relations between language and cognition in Williams syndrome. In Bartke, S. and Siegmüller, J. (eds.) 2004, *Williams Syndrome Across Languages*, 63-92. Amsterdam/Philadelphia, John Benjamins Publishing.
- Thomas, M., Grant, J., Barham, Z., Gsodl, M., Laing, E., Lakusta, L., Tyler, L., Grice, S., Paterson, S. and Karmiloff-Smith, A. 2001. Past tense formation in Williams Syndrome. *Language and Cognitive Processes* 16, 2/3, 143-176.

Experimental investigations on implicatures: a window into the semantics/pragmatics interface

Napoleon Katsos

Research Centre for English and Applied Linguistics, University of Cambridge, UK

Abstract

It is traditionally assumed in the linguistic-pragmatic literature that Scalar Implicatures (A or B \gg either A or B but not both; some of the Fs \gg at least one but not all of the Fs) are explicitly defeasible, structure-dependent and defeasible in context. We present three off-line studies that demonstrate the psycholinguistic reality of these properties of Scalar Implicatures (henceforth SIs). We then present two on-line text comprehension experiments that investigate the time-course of generating SIs and support a pragmatic account of SIs, according to which SIs are generated only when both structural and contextual constraints license them. We aim to demonstrate how an experimental approach can be informative on core issues in the semantics/pragmatics literature.

Introduction

Certain linguistic expressions form entailment scales where terms on the right of the scale are informationally stronger than terms on the left (measured by number of entailments, e.g. <some, most, all>, <or, and>; see Horn, 1984 i.a.). Scalar expressions may trigger Scalar Implicatures (SIs) when the assertion of an informationally weaker term conversationally implies (" \gg ") the negation of the stronger terms in (1a & b):

- (1) a Mary: Who is representing the company at the court hearing?
John: Turner or Morris. \gg Either Turner or Morris but not both.
b Mary: How is our candidate doing in the polls?
John: He has managed to overtake some of his opponents that have little funding. \gg *At least one but not all them.*

Characteristic properties of SIs include explicit defeasibility, structure-dependency and defeasibility in context (Gazdar, 1979; Horn, 1984 i.a.). The fact that SIs share some properties of grammatical inferences has given a rise to a debate on how to classify them: as structure-based default inferences (Chierchia 2004; Levinson 2000 i.a.) or truly context-dependent pragmatic inferences (Atlas, 2005; Carston, 2002; Grice, 1975; Hirschberg 1991;

Recanati, 2003; Sauerland, 2004; Sperber & Wilson, 1995 i.a.). In the first part of the paper we present three off-line studies that demonstrate the psycholinguistic reality of the properties of SIs. In the second part of the paper we present two on-line studies that address the debate between default and pragmatic theories.

On the properties of SIs

Explicit defeasibility

The first off-line study investigates whether SIs are defeasible inferences, i.e. whether they can be explicitly revised without giving rise to contradictions. The baseline condition for defeasibility is the entailment, a grammatical inference whose contradiction ought to give rise to a strong contradiction.

Participants were asked to rate short question/answer pairs for coherence. The critical items consisted of a question and an answer that came in two utterances. The first utterance of the answer contained a disjunction in upward-entailing structure that licenses the generation of an SI. The second utterance of the answer revised an aspect of the meaning of the disjunction. In the Implicature condition, the second utterance contradicted the content of the SI of the disjunction (2a). In the Entailment condition, the second utterance contradicted the content of the entailment of the disjunction (2b).

- (2) a The director asked his consultant: Who is representing our company at the court hearing? His consultant replied: Turner or Morris. In fact, both of them are.
- b The director asked his consultant: Who is representing our company at the court hearing? His consultant replied: Turner or Morris. In fact, none of them are.

Analyses of variance indicate that revising the implicature is significantly more acceptable than revising an entailment (5.7 vs 1.3 on a 7 point scalar where 7 indicates that the answer is 'perfectly coherent'; $F1(1, 24) = 445.3, p < .001$; $F2(1, 16) = 990.2, p < .001$). This is evidence that implicatures are explicitly defeasible in a way that truly grammatical inferences are not.

Structure-dependency and defeasibility in context

It is also argued in the linguistic literature that SIs are constrained by structural and contextual factors. SIs are generated in conditions where both structure and context license them, i.e. in Upward-entailing structures with

Upper-bound contexts; in this condition the disjunction should be interpreted exclusively, with an SI (3a). However, SIs are not available in conditions where contextual constraints do not license them, i.e. in Upward-entailing structures with Lower-bound contexts, where the Lower-bound context biases towards an inclusive interpretation of the disjunction without an SI (3b). Furthermore, SIs are not generated when linguistic structure doesn't license them, i.e. in Downward-entailing structures (e.g. in the antecedent of a conditional, as in 3c):

- (3) a UB: The director asked his consultant: Who is representing our company at the court hearing? His consultant replied: Turner or Morris from the Legal Department.
- b LB: The director asked his consultant: Who is available to represent our company at the court hearing? His consultant replied: Turner or Morris from the Legal Department.
- c DE: The director asked his consultant: Who is representing our company at the court hearing? His consultant replied: I believe that if Turner or Morris from the Legal Department do so, we need not worry too much.

In the second off-line study, participants were asked to rate on a scale whether they believe that the answer implies 'X or Y *but not both of them*', or whether they believe that the answer implies 'X or Y and *even both of them*'. In the third off-line study, participants were asked to fill in a verb inflected for number at the end of the last utterance. We assumed that if they interpreted the disjunction with an SI, they would use a verb form inflected in singular (e.g. X or Y *is*), whereas if they interpreted the scalar term without an SI, they would use a verb inflected in plural (e.g. X or Y *are*).

With regards to the second study, the disjunction was judged as exclusive in UB and inclusive in LB and DE (2.9 vs 5.2 and 5.1 respectively in a 7 point scale, where 1 indicates that the disjunction was exclusive). Analyses of variance indicate a main effect of condition $F1(2, 26) = 37.5, p < .001$; $F2(2, 14) = 23.7, p < .001$. Planned comparisons reveal that UB is judged significantly more exclusive and that LB and DE are equally inclusive¹.

With regards to the third study, participants used a verb in singular agreement 82.1% in UB, 49.1% in LB and 47.9% in DE. There was a main effect of Condition ($F1(2, 38) = 77.3, p < 0.001$; $F2(2, 14) = 48.4, p < 0.001$). Planned comparisons show that there was a significant difference between the UB and the LB conditions and the UB and DE conditions whereas the difference between the LB and DE conditions was not significant¹. We conclude that SIs are generated in UB, where the inference is li-

censed both by context and structure, but not generated in LB and DE conditions where either the context or the structure don't license the SI.

The debate between default and pragmatic accounts

The off-line studies show that SIs are indeed explicitly defeasible, structure-dependent and defeasible in context. In the final part of the paper we present two on-line studies that address the debate on the default vs pragmatic nature of SIs. Default accounts (Chierchia 2004; Levinson 2000 i.a.) claim that SIs are generated by default when licensed by structural constraints (in UB and LB), and may have to be cancelled in subsequent stages if not licensed by the context (in LB). Pragmatic accounts claim that SIs are generated only when both structure and context license them. In case the context doesn't license the SI, the SI is simply not generated, rather than generated and then cancelled. Two studies investigated the on-line processing of disjunctions (with items similar to 3a & b) and the existential quantifier.

Reading time results for the disjunction indicate that processing the scalar term with an SI in the UB condition is more time consuming than processing the scalar term without an SI in the LB condition (811ms vs 761ms; $F_1(1,36) = 6.053, p < 0.02$; $F_2(1, 15) = 8.612, p = 0.01$). Similar results were obtained for the existential quantifier¹. It is impossible to argue that in the LB condition the SI was generated by default and then cancelled, in less time than the SI was generated in the UB condition. On the other hand, these findings are consistent with pragmatic accounts that predict that SIs are generated only when licensed by both structure and context in UB, whereas they are simply not generated at all when they are not licensed by context in LB.

Conclusion

We presented three studies that demonstrate the psycholinguistic reality of traditional linguistic-pragmatic intuitions on SIs. We also presented two on-line studies that are informative on the recent debate on the default vs contextual nature of SIs. We discuss these findings with regards to literature in sentence processing and the acquisition of semantic/pragmatic competence, and we illustrate how the experimental approach can contribute to issues in the core of linguistic theory.

Notes

Please contact the author for appendices with the items, statistical analyses of the planned comparisons and a list of references.

On learnability and naturalness as constraints on phonological grammar

Hahn Koo and Jennifer Cole

Department of Linguistics, University of Illinois at Urbana Champaign, USA

Abstract

We report six experiments on learnability of four non-adjacent phonotactic constraints which differ in their attested frequency and phonetic conditioning factors; liquid harmony, liquid disharmony, backness harmony, and backness disharmony. Our results suggest that such phonotactic constraints can be implicitly learned from brief experience and that learnability of a phonological grammar may be independent of its attested frequency and phonetic basis.

Introduction

The phonological structures of language are partly shaped by phonetic processes. For example, assimilation is attributed to undershoot and overlap of articulatory gestures (Browman & Goldstein 1992; Ohala 1990). Learnability is another source of constraint: sound patterns become grammaticized only to the extent that they define robust generalizations over the lexicon that are learnable from the speech environment (Bybee 2001). Thus, we predict that frequently attested phonotactic patterns are those which are both perceptually salient and easily learned. This paper tests this prediction using the experimental technique of artificial grammar learning, where adult subjects learn phonotactic dependencies based on brief exposure to nonce words, as in Onishi et al. (2002).

We compare evidence of learning across four phonotactic constraints involving harmony and disharmony between non-adjacent liquids and high vowels: liquid harmony, liquid disharmony, backness harmony targeting high vowels, backness disharmony targeting high vowels. These constraints have plausible bases in processes of speech production and/or perception, but they differ in their attested frequency in natural languages. Liquid disharmony is more commonly attested than liquid harmony (Hansson 2001), while vowel harmony is more common than vowel disharmony (Pycha et al. 2003). Furthermore, constraints involving non-adjacent vowels are more common than those involving non-adjacent consonants. Our hypothesis is that the lower frequency constraints will also show lesser evidence of learning.

Auditory repetition experiments

We tested learnability of each of four constraints (liquid harmony, liquid disharmony, backness harmony, backness disharmony) in separate experiments (Experiments 1A-1D) through an auditory repetition task. In the study phase, subjects hear and repeat words that instantiate the phonotactic constraint (study words), and in the test phase they hear and repeat new words that are consistent with the constraint (legal words) and words that violate the constraint (illegal words). Mean reaction time measures are compared for legal and illegal words. Evidence of learning was interpreted as significantly shorter mean latencies to legal words than to illegal words.

Methods

Four experiments investigated learnability of the respective four phonotactic constraints. 15 adult native speakers of English participated in each experiment and received course-credit for compensation.

Stimuli were nonce words of the form $C_1V_1.C_2V_2.C_3V_3$, produced by a male native speaker of English. The first syllable (C_1V_1) was either /sa/ or /ke/. Elsewhere, consonants and vowels were chosen from {s, k, l, r, a, e, i, u}. For each experiment, the words were classified into four types of words: study, legal, illegal, and filler (distracter) words. For example, for the experiment on learnability of backness disharmony, study words and legal words had two high vowels of conflicting backness, illegal words had two high vowels of same backness, and filler words were the remaining words. For each experiment session, 16 study words, 18 legal words, 18 illegal words, and 40 filler words were pseudo-randomly chosen. The chosen words were then distributed across five blocks. Study words recurred in each of five blocks. Legal and illegal words were evenly distributed in the last three blocks. Filler words were evenly distributed in all five blocks.

In each trial, subjects listened to a word through a headset and repeated it as quickly and accurately as possible into a microphone. Latency was measured from the stimulus offset to the response onset for each trial. The entire session was audio-taped for analysis of response accuracy.

Results

For each subject, latencies were averaged per block and per stimulus type after the following were excluded: (1) errors, (2) responses not detected by the microphone in the first attempt, (3) responses initiated before the word was presented through its penultimate syllable, (4) responses with reaction times of 2.5 standard deviations away from the mean.

For each of the four experiments, a within-subject ANOVA was conducted with block (blocks 3~5) and legality (legal vs. illegal) as factors. The

effect of legality was significant in experiments on liquid harmony and disharmony, but not in experiments on backness harmony and disharmony. This suggests that subjects learned to generalize the liquid constraints to new instances but failed to learn to generalize the backness constraints. The results of analyses are summarized in Table 1.

Table 1. Results of within-subject ANOVAs.

Constraint (Experiment)	Legality	Legality \times Block
Liquid Harmony (1A)	$F(1,14)=6.278,$	$F(2,28)=1.487,$
Liquid Disharmony (1B)	$F(1,14)=8.435,$	$F(2,28)=0.111,$
Backness Harmony (1C)	$F(1,14)=0.164,$	$F(2,28)=0.289,$
Backness Disharmony	$F(1,14)=0.188,$	$F(2,28)=0.390,$

Grammaticality judgment experiments

The lack of evidence of learning for the backness constraints may have been task-specific in that repetition facilitation may have been present, but not strong enough to result in significantly shorter repetition latencies with these stimuli. We ran two further experiments to test learnability of liquid harmony (Experiment 2A) and backness harmony (Experiment 2B) with the same nonce word stimuli, but with a grammaticality judgment task in the test phase. After a study phase with an auditory repetition task, subjects were asked to decide for each test word if it belonged to the language exemplified in the study phase (i.e., if it was ‘grammatical’). Evidence of learning was interpreted as the subjects’ ability to discriminate legal words from illegal words.

Methods

Subjects in each of the two experiments comprised 15 adult native speakers of English, who received course credit for compensation. The materials were identical to the ones used in the auditory repetition experiments. A session in 2A and 2B comprised a study phase with three blocks and a test block, and each study block contained the study words and filler words from 1A and 1C, respectively.^[j1] The subjects performed the auditory repetition task during the study phase. In the test block, which comprised legal, illegal and filler word, subjects judged for each test word whether it belonged to the language of the study phase by responding “Yes” or “No”.

Results

A d' -score was computed for each subject, where hit-rate was defined as the proportion of “Yes” responses to legal words and false-alarm rate was defined as the proportion of “Yes” responses to illegal words. The log-linear

rule (Hautus 1995) was applied to account for extreme cases where subjects had a hit-rate of 1.0.

A one-sample *t*-test with $d'=0.0$ as the null hypothesis showed that subjects discriminated legal words from illegal words significantly in both the liquid harmony experiment ($t(14)=3.717$, $p=0.002$) and the backness harmony experiment ($t(14)=3.399$, $p=0.004$). The scores between the two experiments were not significantly different ($t(28)=1.089$, $p=0.285$). Thus, subjects appeared to have learned to generalize backness harmony as well as liquid harmony.

Conclusion

Our findings from six experiments show that subjects implicitly learned phonotactic constraints from brief experience, and that constraints that differ in attested frequency and phonetic conditioning factors are learned equally well. We conclude that the learnability of a constraint is independent of its phonetic basis and attested frequency in natural languages. However, we also find differences in learning that relate to differences on the experimental task, indicating that future research must consider task factors in relation to phonetic factors in assessing the role of learnability on phonological grammar.

References

- Browman, C. and Goldstein, L. 1992. Articulatory phonology: an overview. *Phonetica* 49, 155-180.
- Bybee, J. 2001. *Phonology and Language Use*. Cambridge, CUP.
- Hansson, G. O. 2001. Theoretical and typological issues in consonant harmony. Doctoral dissertation, Department of Linguistics, University of California at Berkeley.
- Hautus, M. 1995. Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavioral Research Methods, Instruments, and Computers* 27, 46-51.
- Ohala, J. 1990. The phonetics and phonology of aspects of assimilation. In Kingston, E. and Beckman, M. (eds.) 1990, *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, 258-275, Cambridge, CUP.
- Onishi, K., Chambers, K., and Fisher, C. 2002. Learning phonotactic constraints from brief auditory experience. *Cognition* 83, B13-B23.
- Pycha, A., Nowak, P., Shin, E., and Shosted, S. 2003. In Garding, G. and Tsujimura, M. (eds.) 2003, *Proceedings of the XXIIth West Coast Conference on Formal Linguistics*, 423-435, Somerville, M.A., U.S.A.

Prosody and punctuation in *The Stranger* by Albert Camus

Mari Lehtinen

Department of Romance Languages, University of Helsinki, Finland

Abstract

The aim of this paper is to provide an overview of the results of a study concerning the prosodic interpretation of punctuation. The focus is placed on the two most common punctuation marks, that is to say, full stops and commas. The findings are based on a comparison between *The Stranger* by Albert Camus the way it is published as a book and the way it is read out by its author in two French radio broadcasts in 1954. My findings suggest that the oral interpretation of the punctuation clearly is subject to a certain code. However, the most interesting findings of this study concern a non-negligible number of cases in which the code is not observed.

Introduction

The relationship between the spoken and the written languages constitutes one of the eternal questions repeatedly taken up in different fields of linguistics as well as within other disciplines. This paper approaches the question in the perspective of oral equivalencies of written punctuation marks. For this purpose, I have chosen to study *The Stranger* (*L'étranger*) by Albert Camus. The data concerning the punctuation comes from a French edition of the book, published by Gallimard in Paris in 1996 (1st edition: 1942). The prosodic findings in turn are based on two French radio broadcasts in which Camus himself reads out the first half of the volume. The broadcasts were recorded and transmitted by *La Radiodiffusion française* (RDF) in 1954. The length of the reading time examined for this paper is 87 minutes, which corresponds to six chapters or 95 pages of the text. These pages include 2309 punctuation marks. The number consists of full stops, commas, question marks, exclamation points, colons, dashes, semicolons, points of suspension and parentheses. Because of the limited number of pages available, only the occurrences of full stops and commas are taken into consideration in this paper; these two marks cover together about 94 % of the total number of punctuation marks in my corpus. Each occurrence has been analysed both melodically and rhythmically. The prosodic analyses have been carried out with the help of the computer programs Sound Forge and Praat.

According to my findings, full stops tend to be read out with a pitch fall followed by a pause, whereas commas rather entail a pitch rise instead of a fall. This might not seem surprising. But it is interesting to notice that there

also are a certain number of cases in which full stops and commas are 1) interpreted melodically the opposite way compared with their prototypical oral equivalent, 2) interpreted melodically in the standard manner but passed by without a pause, that is to say, ignored from the rhythmical point of view, 3) interpreted melodically the opposite way with regard to the prototype and in addition to that, without a pause, 4) melodically ignored, that is to say, passed over without any noticeable pitch change, or 5) ignored both melodically and rhythmically, that is to say, passed by without a pitch change or a pause.

The atypical prosodic interpretations of punctuation marks have as a result a prominent intonation the functions of which often seem to be similar to those of certain prosodic contextualisation phenomena of speech: their occurrences often seem to be related to stylistic and discourse-structuring factors that, in turn, contribute to the construction of an interpretational framework for what is being said.

Prosodic interpretation of full stops and commas. Prototypes and atypical cases

In this chapter, I will briefly present some general observations and hypotheses concerning different prosodic interpretations of full stops and commas. However, for lack of space, only an overall description of some general aspects can be provided in the current paper: unfortunately examples cannot be given and only the prototypical prosodic equivalent as well as the most common atypical interpretation of each of the two punctuation marks can be treated.

Prototypical interpretations

As it could be expected, the full stop is the most common punctuation mark in the corpus: it occurs 1185 times and thus covers over 50 % of all punctuation. In 79 % of the cases full stops are interpreted with a remarkable pitch fall associated with a clearly perceptible pause. This seems natural, because as it is largely known, a pitch fall typically indicates conclusion in spoken French (*cf.* references), whereas a full stop basically finishes an autonomous whole in written language (*e.g.* Catach 1996). The prototypic functions of a pitch fall and a full stop thus seem similar both syntactically and semantically with respect to their concluding nature. Concerning the pause, which is the second constituent of the prototype, it is known that when its occurrence is related to discourse-structuring, it gives the listener the time required to construct the meaning of the preceding unit (Morel & Danon-Boileau 1998). The functions of a pause are thus consistent with those of a pitch fall and a

full stop (when a pause is used as a discourse-structuring marker, which is obviously not always the case; a pause can also be related to utterance-formulation processes, lexical research, stylistic factors, etc.). According to my findings, when a full stop is interpreted in this prototypical manner, it gives an independent position to the action described in the sentence that it concludes. It also “closes” the situation in a way; that is to say, it does not create expectations about any other relevant situational factors but rather presents the sentence as an exhaustive description of the action.

In 64 % of the cases examined a comma entails a rising pitch combined with a clearly perceivable pause. A comma is typically used to structure the syntactic and semantic whole formed by a sentence by separating its constituents (Catach 1996). As a comma always occurs inside a sentence and not at the end of it, its functions are generally not concluding, but, on the contrary, continuative. In regard to this aspect of continuity, the role of a comma thus seems in some measure analogous with that of a pitch rise in French speech (*cf.* references).

It is important to note that especially in languages like French in which the use of a comma is not subject to strict rules (unlike in Finnish, for example), the structuring functions of the mark do not reflect only the syntax of the sentence but also, for instance, its internal informational relationships as well as the stylistic choices of the author. In the light of my findings, it is frequent that when two or more separate acts are described in consecutive clauses within one sentence so that each clause is finished by a comma interpreted in the prototypical manner, the rising pitch emphasizes the situational dependence and closeness of the acts which are being described; often a full stop and / or a falling pitch between the clauses would mark too big an independence and temporal distance between the acts.

Most common atypical interpretations

Although most of the full stops occurring in my corpus entail a falling pitch and a pause, it is, nevertheless, interesting to observe that 14 % of them are interpreted by the speaker with a remarkable pitch rise followed by a pause. In these cases, the prosodic interpretation is not analogous with the prototypical use of a full stop, but it is rather reminiscent of that of a comma. Accordingly, it seems that when a full stop is interpreted with a pitch rise instead of a fall, it indicates structural and situational continuity. All cases of this kind naturally occur inside a textual paragraph.

Generally speaking, a rising pitch seems to serve as a link between two or more clauses or sentences that describe acts or circumstances that are temporally and spatially closely related to each other. The difference between full stops and commas in this respect is that commas interpreted with

a rising pitch most often separate acts or circumstances that are simultaneous or nearly simultaneous, whereas those separated by full stops are consecutive: clauses distinguished by “rising commas” typically refer to a same situation, whereas sentences concerned by “rising full stops” rather form a continuum of closely related situations. In both cases, the rising pitch anticipates activity; it creates an interpretational framework for an active situation.

As already mentioned above, when a full stop is interpreted in the prototypical manner, it normally indicates finality and autonomy. Analogously, when a comma is interpreted with a pitch fall and a pause (6 % of the cases), it loses its continuative value. The difference between full stops and commas in this regard is that the concluding effect of a comma is weaker than that of a full stop: where a full stop implies finality and autonomy, a comma implies stagnation. Often commas entailing a falling pitch and a pause occur in “passive” contexts, such as descriptions of the milieu. In cases of this kind, the pitch falls have a stylistic function: they typically convey a nuance of coldness and of informational irrelevance and contribute to contextualise the contents of the sentence in question as a mere description of circumstances, and not, for example, as a meaningful event.

References (selection)

- Catach, N. 1996. [1994]. *La ponctuation*. Paris, PUF.
- Delattre, P. 1966. Les dix intonations de base du français. *French Review* XL (I), 1-14.
- Di Cristo, A. 1998. Intonation in French. In Hirst, D. and Di Cristo, A. (eds.) 1998, *Intonation Systems. A Survey of Twenty Languages*, 195-218. Cambridge, Cambridge University Press.
- Fónagy, I. and Fónagy, J. 1983. L’intonation et l’organisation du discours. *BSLP* 78, 1, 161-209.
- Morel, M.-A. and Danon-Boileau, L. 1998. *Grammaire de l’intonation. L’exemple du français oral*. Paris / Gap, Ophrys.
- Rossi, M. 1999. *L’intonation, le système du français : description et modélisation*. Paris / Gap, Ophrys.

An acoustic study on the paralinguistic prosody in the politeness talk in Taiwan Mandarin

Hsin-Yi Lin¹, Kwock-Ping John Tse² and Janice Fon³

¹Department of English, National Taiwan Normal University, Taiwan

²College of Foreign Language and Literature, Providence University, Taiwan

³Graduate Institute of Linguistics, National Taiwan University, Taiwan

Abstract

The relationship between the interlocutors is known, in Sociolinguistics and Pragmatics, to be influential on speakers' tone of voice. This study examined this phenomenon from acoustical aspects by measuring the duration and pitch of the dialogues made by pairs of talkers. Shorter word duration is found between strangers' dialogues and lower pitch register is found in female talkers' speech to males. The findings indicate that the prosody of talkers indeed varies with their familiarity with hearers and also with hearers' gender.

Introduction

Speakers' speech styles in conversations have been found, in order to be considered polite, to be adjusted to the relationship between the speaker and the hearer, and the influential factors are hearers' gender, the power relationship, and familiarity between the interlocutors (Brown & Levinson, 1987; Holmes, 2001). The present study, using familiarity and hearers' gender as factors (Ofuka, McKeown, Waterman, & Roach, 2000), aims to find out their effects on the duration and pitch of young females' speeches in Taiwan Mandarin.

Method

The current study used the Word Card Display game, which is a slightly varied version of the Shape Display Task (Fon, 2006). Modifications were made to obtain a better control for data elicitation.

Participants

16 female talkers participated in the study. Half of the talkers had partners of the same gender, and half had partners of different gender. Each of the talkers participated twice, once with their good friends, and once with strangers matched by the first author. In total, there were 12 dialogues collected, four

of which were of the same gender, and eight of which were of different genders.

Equipment

In the Word Card Display Game, three kinds of equipment were needed: a display board with a 2 x 3 grid, twelve word cards, and a game pocket with the twelve word cards in it. Each word card contained a Chinese character. The characters were chosen so that they were all in Tone 4, which is a high-falling tone.

Two head-mounted microphones (SONY MRD 7520), and a recorder (BurnIt CDR830) were used for recording.

Stimuli

The 12 word card characters contained 6 pairs of homophones: *mi*⁴ ‘honey’ and ‘secret’, *la*⁴ ‘spicy’ and ‘cured food’, *qi*⁴ ‘container’ and ‘air’, *dian*⁴ ‘electricity’ and ‘shop’, *mao*⁴ ‘exuberant’ and ‘hat’, and *wu*⁴ ‘thing’ and ‘mist’.

Procedure

The main goal for the talkers and their participants was to complete the task by making the word card display on their display boards look exactly the same. Detailed procedure can be referred to Fon (2006).

Measurement

The concerned prosodic cues in the present study are duration and pitch. For duration, the lengths of the stimuli characters were measured. The stimuli characters would be spoken by the participants when they tell their partners what character on the word card they got. For example, when *mi*⁴ was picked, they would say it was *fong*¹*mi*⁴*de*⁰*mi*⁴ ‘the *mi*⁴ as in honey’, in which way the *mi*⁴ would not be mistaken as other homophones by the hearers. As shown in the example, the target character would appear in two positions in the phrases, one is at the beginning or middle position (presented as “W1” hence), and the other at the final position (presented as “W2” hence).

For pitch, the pitch ranges and registers of the phrases which contained the stimuli characters were measured. Therefore, the phrases which have the pattern as *fong*¹*mi*⁴*de*⁰*mi*⁴ were our targets. The phrasal pitch ranges were obtained by computing the range between the peak pitch in W1 and the valley pitch in W2, and the registers that were occupied by these pitch ranges were also analyzed by observation.

Results

Duration

Figure 1 shows the duration of W1 and W2. It appears that the durations of W2 are longer in dialogues between familiar speakers, and the effect of familiarity on W2 duration is significant in the 2-way ANOVA test ($F(1, 134) = 3.657, p < 0.05$). That the final lengthening (W2 duration – W1 duration) is longer in familiars' dialogues can also be observed ($F(1, 134) = 3.179, p = 0.077$). Though it fails to reach significance at present, it should be a worth observing trend in the future study. The result suggests that duration manipulation is made according to the different degrees of familiarity between interlocutors.

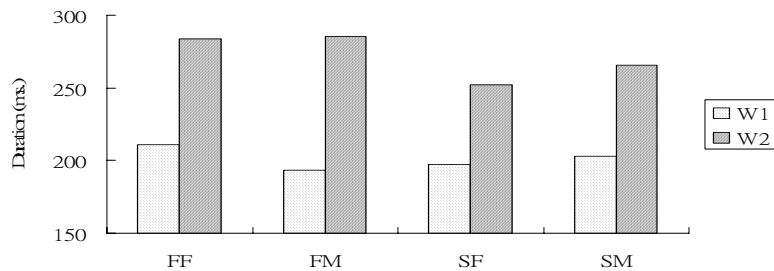


Figure 1. Word duration of W1 and W2. (FF = with a familiar female; FM = with a familiar male; SF = with a strange female; SM = with a strange male).

Phrasal pitch range and pitch register

The phrasal pitch ranges were obtained by computing the range between the W1 peak pitches and W2 valley pitches. Figure 2 shows that neither familiarity nor hearers' gender has effect on the pitch ranges. However, as can be seen in Figure 2, when speakers are talking to male hearers, both the W1 peak and W2 valley fall to a lower register than those used to female hearers. The statistic analysis shows that this effect of hearers' gender on pitch register is significant (2-way ANOVA, W1: $F(1, 134) = 17.080, p < 0.05$; W2: $F(1, 134) = 7.733, p < 0.05$), which suggests that the female talkers would accommodate their male hearers by lowering their own pitch registers.

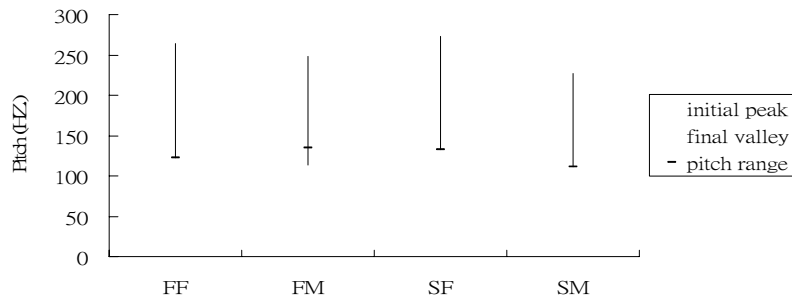


Figure 2. The average initial peak pitch and final valley pitch of the target phrases. The top-end and the bottom-end on each vertical line represent the initial peak and the final valley respectively. (Abbreviations have the same meanings as those in Fig 1.)

Conclusions

This paper presents a study on the effect of social relationship between the interlocutors on their speech prosody. The results show that, for young female speakers in Taiwan, when they talk to strangers, their phrasal final length is shorter, and when they talk to the opposite gender, which is male in this study, their pitch registers are lower. These findings suggest that (1) the phrasal final length is manipulated by female talkers according to their familiarity with hearers, and (2) the pitch register is adjusted to accommodate to the hearers' pitch registers. Therefore, the effect of social relationship between the interlocutors on their speech prosody is demonstrated.

References

- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (2 ed.). London: Cambridge University.
- Fon, J. (2006). Shape Display: task design and corpus collection. Paper presented at the The 3rd Speech Prosody, Dresden, Germany.
- Holmes, J. (2001). Gender, politeness and stereotypes. In G. Leech & M. Short (Eds.), *An Introduction to Sociolinguistics* (pp. 284-316): Longman.
- Ofuka, E., McKeown, J. D., Waterman, M. G., & Roach, P. J. (2000). Prosodic cues for rated politeness in Japanese speech. *Speech Communication*, 32, 199-127.

Analysis of stop consonant production in European Portuguese

Marisa Lobo Lousada¹ and Luis M. T. Jesus²

¹ Escola Superior de Saúde da Universidade de Aveiro and Secção Autónoma de Ciências da Saúde, Universidade de Aveiro, Portugal

² Escola Superior de Saúde da Universidade de Aveiro and Instituto de Engenharia Electrónica e Telemática de Aveiro, Universidade de Aveiro, Portugal

Abstract

This paper investigates acoustic features correlated with voicing (VOT, stop duration, closure duration, release duration, voicing into closure duration, duration of preceding vowel and duration of following vowel) and with the place of articulation (spectral peaks) of stop consonants /p, b, t, d, k, g/. A corpus with these stops in initial, medial and final word position was recorded for six native speakers of European Portuguese.

Introduction

The present study examines the acoustic properties correlated with voicing and with place of articulation for European Portuguese stops.

Andrade (1980) compared VOT of homorganic stops, in initial position, before a vowel, in words produced by a speaker of European Portuguese. Results showed that some voiced stops had a period of prevoicing (120 ms to 130 ms) followed by a devoiced period (10 to 20 ms), and that VOT was larger for velars, than for labials and dentals, as in English (Klatt, 1975).

Stops are often devoiced (Alphen and Smits, 2004) and there multiple acoustic properties related with voicing distinction. Viana (1984) and Veloso (1995) observed that stop duration and duration of the preceding vowel, were acoustic properties that cued voicing in European. Fuchs (2005) also suggested closure duration, duration of following vowel, duration of preceding vowel, and voicing into closure duration, as voicing cues.

The shape of the spectrum of the stop release was analyzed by Blumstein and Stevens (1978). Labial stops had a diffuse-falling or diffuse-flat pattern, and alveolar stops, also had a diffuse spread of peaks of energy, but the amplitudes of these peaks were greater at high frequencies (diffuse-rising pattern). Velar stops had a mid-frequency spectral peak (compact pattern). Labial and alveolar stops shared the property of diffuseness and were distinguished by the shape of the spectral energy distribution.

Recording method

A corpus of fifty four Portuguese real words containing /p, b, t, d, k, g/ was recorded using a Philips SBC ME 400 unidirectional condenser microphone located 20 cm in front of the subject's mouth. A laryngograph signal (Lx) was also collected using a laryngograph processor (model EG-PC3 produced by Tiger DRS, Inc., USA). The acoustic and Lx signals were pre-amplified (Rane MS 1-b) and recorded with a Sony PCM-R300 DAT recorder, each with 16 bits and a sampling frequency of 48 kHz.

The corpus contained an equal number (eighteen) of words with stops in: initial position, followed by the vowels /a, i, u/; medial position, preceded by the vowels /a, i, u/ and followed by the vowel /ə/; final position, preceded by the vowels /ə, a/. The words were produced without any context and within the frame sentence "Diga,... por favor." by six native speakers of European Portuguese (three men and three women).

Analysis method

Temporal analysis

All corpus words were manually analyzed to detect the: beginning of the preceding vowel; end of preceding vowel and beginning of closure; voice offset; end of closure and beginning of release; beginning of prevoicing; end of the release and beginning of the following vowel; end of following vowel.

The following measurements were obtained: duration of preceding vowel, closure duration, voicing into closure duration, release duration, type of voicing (voiced, partially devoiced or voiceless), VOT, stop duration and duration of following vowel.

Spectral analysis

Multitaper spectra were calculated with 11ms windows left aligned to the release of the stop. We also calculated the frequency (F) at which the spectral amplitude was maximum, excluding the fundamental and its harmonics in voiced stops. It provided an endpoint for line fits used to determine the spectral slope. The average values for all Corpus 1 stops produced by speakers ML e LJ were: $F_{/p,b/} = 3,7$ kHz, $F_{/t,d/} = 3,9$ kHz and $F_{/k,g/} = 4,6$ kHz.

Results

Temporal analysis

Results of temporal analysis showed that when speakers ML and IM (female), LJ and HR (male) produced the words in a frame sentence, the stop duration, as shown in Figure 1, and the closure duration was longer for voiceless than for voiced stops in all word positions. The voicing into closure duration, the duration of preceding vowel and the duration of following vowel were generally shorter for voiceless than for voiced stops. VOT was generally shorter for bilabials than for dentals, and shorter for dentals than for velars except in final-word position.

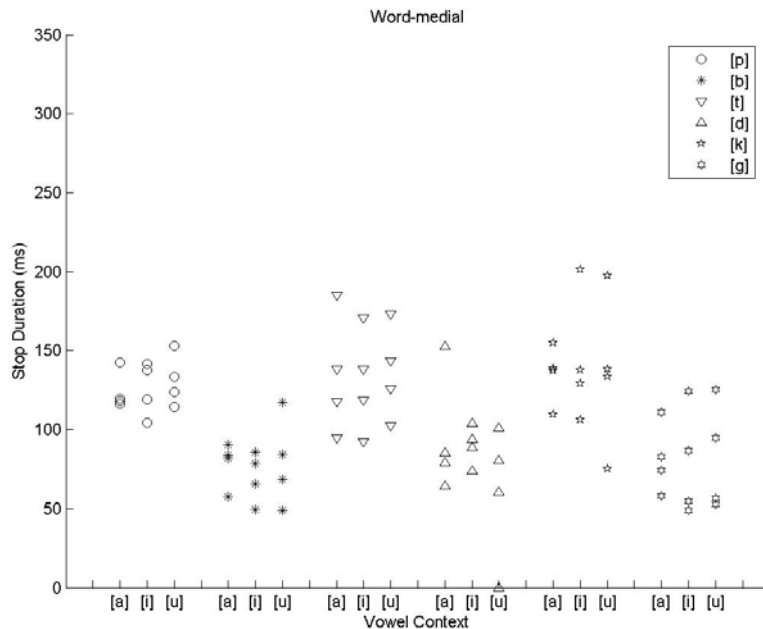


Figure 1. Stop duration for words in medial position produce by speakers ML, IM, LJ and HR.

Spectral analysis

[p] had spectral troughs at 0.8-4.6 kHz, and a broad peak at 1.4-5.6 kHz. [b] had spectral troughs at 0.7-5.0 kHz, and a broad peak at 1.5-5.6 kHz. [t] had spectral troughs at 1.7-7.0, a first peak at 0.3-3.7 kHz, a second peak at 0.2-4.6 kHz, and a broad peak at 6.0-10.4kHz. [d] had spectral troughs at 1.5-5.5 kHz, a first peak at 0.3-1.7kHz, a second peak at 2.4-5.6kHz, a first broad peak at 5.2-9.9 kHz, and a second broad peak at 11.0-12.8 kHz. [k] had

spectral troughs at 2.9-6.3 kHz, a first peak at 0.6-3.8 kHz, a second peak at 3.9-5.4 kHz, a first broad peak at 7.1-9.4 kHz, and a second broad peak at 10.0-13.2 kHz. [g] had a spectral troughs at 0.8-7.4 kHz, a first peak at 1.0-2.6 kHz, a second peak at 3.9-4.9 kHz, a first broad peak at 6.8-9.1, and a second broad peak at 12.0-13.6 kHz. Results of spectral analysis also showed that: [p, b] had a concentration of energy in the low frequencies (500 to 1500 Hz); [t, d] had flat spectrums or a concentration of energy in the high frequencies (above 4000 Hz); [k, g] had a concentration of energy in intermediate frequency regions (1500 to 4000 Hz).

Voiceless bilabial stops, in initial position, followed by vowel [a] had spectra with steeper negative slopes than dentals and velars. Spectra of velars followed by vowel [i] had positive slopes, bilabials were mostly flat and dentals had a negative slope. Dentals followed by vowel [u] had a positive, sometimes flat slope, and bilabials and velars had a negative slope. Velars in medial position had a less negative slope than bilabials and dentals. Word-final dentals had a less negative slope than bilabials and velars.

Conclusions

The results of stop duration agree with those presented by Viana (1984) and Veloso (1995). We observed the same correlation between place of articulation and VOT previously reported by Klatt (1975) and Andrade (1980). The results showed that different acoustic properties are important for voicing distinction in European Portuguese stops. We were not able to observe the spectral patterns reported by Blumstein and Stevens (1978).

References

- Alphen, P. and Smits, R. 2004. Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of pre-voicing. *Journal of Phonetics* 32, 455-491.
- Andrade, A. 1980. Estudos experimentais aerodinâmicos, acústicos e palatográficos do vozeamento nas consoantes. CLUL, Lisboa, Portugal.
- Blumstein, S. and Stevens, K. 1979. Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, 66(4), 1001-1017.
- Fuchs, S. 2005. Articulatory correlates of the voicing contrast in alveolar obstruent production in German. *ZAS-Papers in Linguistics*, 41.
- Klatt, D. 1975. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 18(4), 686-706.
- Viana, M. 1984. *Étude de Deux Aspects du Consonantisme du Portugais: Fricatisation et Devoisement*, Doct., U. Sciences Humaines de Strasbourg, France.
- Veloso, J. 1995. Aspectos da percepção das “oclusivas fricativizadas” do Português: Contributo para a compreensão do processamento de contrastes alofônicos. Universidade do Porto, Portugal.

Towards multilingual articulatory feature recognition with Support Vector Machines

Jan Macek, Anja Geumann and Julie Carson-Berndsen
School of Computer Science and Informatics, University College Dublin,
Ireland

Abstract

We present experiments on mono-lingual and cross-lingual articulatory feature recognition for English and German speech data. Our goal is to investigate to what extent it is possible to derive and reuse articulatory feature recognizers, whether particular features are better suited to this task. Finally whether this goal is practically achievable with the chosen machine learning technique and the selected set of speech signal descriptors.

Introduction

Earlier results on articulatory feature recognition suggest that Support Vector Machines perform better than Hidden Markov Models (HMM) (Macek et al., 2005). In frame based articulatory feature recognition the HMM approach does not benefit from its ability to model probabilistic dependencies, which makes it very useful in the task of speech recognition based on phone level descriptions (Kanokphara et al., 2006). This is due to limited dependency between adjacent frames in articulatory feature recognition.

We present experiments with support vector machines that use different types of kernels, a linear and two polynomial of different orders. All parameters were automatically extracted using the PRAAT analysis tool (Boersma and Weenink, 2006). Extraction of descriptive attributes was performed to obtain values of MFCCs with first and second order differences, formants with first order differences and bandwidths, distance between adjacent formants (F_3-F_2 , F_2-F_1 , F_1-f_0), and fundamental frequency. To extract these values we analyzed the data as sequences of 25ms windows with 10ms step length. We used the TIMIT corpus for English and the Phondat2 corpus for German. Both corpora are based on read speech.

In this paper, for TIMIT only dialect region 3 is used as the training set (102 speakers, 10 sentences each) while the whole core test set is used as the test set. The core test set, which is the abridged version of the complete test set, consists of 8 utterances from each of 24 speakers.

The Phondat2 corpus was split up into 11 speakers for training and 5 speakers for testing. Only sentences for which manual annotations were available were used (64 sentences per speaker).

Experiments

In this paper, SVMs with polynomial kernel were used for extraction of articulatory features from the speech signal. SVMs (Schoelkopf and Smola, 2002) learn separating hyperplanes to classify instances in the feature space that are mapped from the input space of the classified data. The mapping from input space to feature space is performed with application of a kernel. The dimension of the feature space is typically much higher than that of the original input space which allows for separability of the data.

The performance of the SVMs on feature recognition in German and English was compared. Performance of other methods on the same task was reported in (Kanokphara et al., 2006; Macek et al., 2005). The SVM classifiers were run with the SVMLight implementation (Joachims, 1999).

The speech signal was classified in a frame-by-frame manner, where every 10 ms for each frame of 25 ms length a set of descriptors was extracted. The non-speech parts, such as silence, are excluded in both feature sets in the training and evaluation. The performance of the articulatory feature recognition was evaluated for different orders of polynomial kernels of the SVMs up to order 3. The performance improved consistently with increasing order of the kernels. In this work we experimented with two sets of descriptors. The first consisted of 12 MFCC values, its first and second order differences, 5 formants (F1–F5) with first order differences and bandwidths, and pitch (f0). The second set of attributes extended the first one by distance in frequency between formants (F3–F2, F2–F1) and lowest formant and pitch (F1–f0).

The comparison of the two sets showed only negligible differences. The results presented in Table 1 are for the second set of descriptors used together with SVMs with polynomial kernel of order 3.

Results and discussion

To evaluate the performance of the recognizers we used two measures, namely the accuracy that gives overall performance regardless of the class distribution in the data, and the F_1 -measure, which is the harmonic mean of the class dependent values of precision and recall. The F_1 -measure gives a better picture of the actual performance of the recognizer for the relevant class. The F_1 -measure is given priority for reasoning about the results.

In the mono-lingual setting, there are large differences in the performance, e.g. robust features such as [CONSONANTAL], [CONTINUANT], [SONORANT] and less well recognized features as [DORSAL], [LABIAL], [LATERAL]. In most cases a feature that is robustly recognized in the monolingual English setting is recognized robustly in the monolingual German setting, and nonrobust features are nonrobust in either of the two languages.

Table 1. Mono-lingual and cross-lingual recognition results.

Feature	E-E		G-G		E-G		G-E	
	Acc. in %	F ₁ -measure	Acc. in %	F ₁ -measure	Acc. in %	F ₁ -measure	Acc. in %	F ₁ -measure
-anterior	91.01	0.923	87.37	0.914	63.37	0.672	82.53	0.870
+anterior		0.892		0.764		0.586		0.734
-atr	92.75	0.957	79.18	0.843	56.39	0.661	77.20	0.861
+atr		0.758		0.690		0.390		0.374
-back	89.02	0.894	92.83	0.916	81.59	0.770	81.67	0.817
+back		0.886		0.937		0.847		0.816
-consonantal	89.10	0.885	94.24	0.931	88.36	0.851	85.11	0.853
+consonantal		0.896		0.951		0.904		0.849
-continuant	90.26	0.865	91.86	0.903	68.51	0.716	78.26	0.655
+continuant		0.924		0.930		0.646		0.841
-coronal	84.01	0.753	86.88	0.909	45.13	0.455	62.98	0.622
+coronal		0.882		0.764		0.447		0.637
-distributed	98.85	0.994	99.55	0.998	98.81	0.994	97.98	0.990
+distributed		0.710		0.792		0.265		0.160
-dorsal	93.02	0.962	91.62	0.953	88.55	0.937	88.59	0.938
+dorsal		0.574		0.583		0.348		0.228
-high	87.96	0.899	87.65	0.884	74.76	0.726	71.51	0.799
+high		0.850		0.868		0.766		0.511
-labial	84.56	0.906	91.55	0.952	82.86	0.897	81.26	0.888
+labial		0.561		0.667		0.495		0.425
-lateral	97.78	0.989	98.65	0.993	97.17	0.986	97.30	0.986
+lateral		0.368		0.133		0.003		0.004
-low	87.31	0.925	92.94	0.950	76.55	0.855	78.45	0.857
+low		0.581		0.880		0.385		0.565
-nasal	97.92	0.989	95.86	0.965	92.59	0.940	95.88	0.978
+nasal		0.810		0.949		0.904		0.652
-round	92.01	0.956	95.06	0.973	89.72	0.943	88.51	0.935
+round		0.603		0.718		0.468		0.491
-sonorant	96.69	0.978	96.35	0.972	93.23	0.947	85.50	0.909
+sonorant		0.937		0.948		0.905		0.642
-strident	92.99	0.704	97.19	0.984	59.78	0.699	77.31	0.439
+strident		0.960		0.904		0.393		0.858
-vocalic	93.12	0.936	94.29	0.952	84.84	0.883	88.10	0.880
+vocalic		0.926		0.929		0.783		0.882
-voiced	93.59	0.883	94.64	0.914	86.88	0.745	90.25	0.834
+voiced		0.956		0.961		0.912		0.931
-vot	90.01	0.884	93.84	0.966	64.36	0.776	41.67	0.561
+vot		0.912		0.656		0.128		0.130

In the cross-lingual settings, the most obvious pattern is that a recognizer that performs poorly in the mono-lingual setting will perform even worse in the cross-lingual setting. However, good performance in the mono-lingual setting is not general indicator for good cross-lingual performance. The features [CONSONANTAL], [VOCALIC], [VOICED] and [SONORANT] are the only ones that appear to be robust across all language combinations.

As can be seen in Table 1 accuracy can be a misleading measure of performance in the case of highly uneven class distribution in the data, an extreme example is the feature [LATERAL].

Although the original motivation for the articulatory features is their language independence, the comparison of performances on different features here suggests that some of the features are more language independent than others. However, studies such as that presented here do provide indications as to how features could be suited for the purposes of multilingual speech recognition.

Acknowledgements

This material is based upon works supported by the Science Foundation Ireland under grant No. 02/INI/II00. The opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of SFI.

References

- Garofolo, JS, Lamel, LF, Fisher, WM, Fiscus, JG, Pallett, DS and Dahlgren, NL. 1993. DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM. Technical Report NISTIR 4930, National Institute of Standards and Technology, Gaithersburg, MD, USA
- Kanokphara, S, Macek, J, and Carson-Berndsen, J. 2006. Comparative Study: HMM & SVM for Automatic Articulatory Feature Extraction. In Proc. of the 19th Intern. Conf. on Industrial, Engineering & Other Applications of Applied Intelligent Systems, Annecy, France, June 2006. Springer Verlag.
- Macek, J, Kanokphara, S and Geumann, A. 2005. Articulatory-acoustic Feature Recognition: Comparison of Machine Learning and HMM methods. In Proc. of the 10th Intern. Conf. on Speech and Computer SPECOM 2005, vol. 1, 99–103. University of Patras, Greece, 2005.
- Joachims, T. 1999. Making large-scale SVM learning practical. In Schoelkopf, B, Burges, CJC and Smola, AJ (eds.), *Advances in Kernel Methods—Support Vector Learning*, 169–184, Cambridge, MA, MIT Press.
- Schoelkopf, B and Smola, AJ. 2002. *Learning with Kernels*. Cambridge, MA, MIT Press.
- Boersma, P and Weenink, D. 2006. Praat: doing phonetics by computer (Version 4.4.04) [Computer program]. Retrieved from <http://www.praat.org/>.

Prosody, syntax, macrosyntax

Philippe Martin

UFRL Université Paris 7 Denis Diderot, France

Abstract

Most of research work on French intonation has been conducted on prepared speech data (i.e. read speech), with various approaches ranging to purely syntactic (Rossi), Autosegmental-Metrical (Sun and Fougeron) or Phonosyntactic (Martin). This paper examines the bases of the phonosyntactic approach extended to spontaneous (non-prepared) speech data, described syntactically from a macrosyntactic point of view (Blanche-Benveniste, Deulofeu).

Experimental intonation phonology

Prosodic words

There is a general agreement to look on or around the accented (stressed) syllable for prosodic phenomena. Minimal prosodic units – prosodic words – contain one (lexical) stress and one optional initial stress. A minimum prosodic unit contains one or more content word (open class word), and optional grammatical words, constrained by a 7 unstressed syllable rule.

One content word forms a group with grammatical words through dependency relations with one stressable last syllable in French. Depending on the speech rate, stressable syllables are effectively stressed (with a final – primary - stress). If two groups have few syllables (in the order of 2 or 3) they can form a larger group with its final syllable stressed. If the group has a large number of syllables (say > 7), it will receive a secondary stress.

Prosodic structure

The prosodic structure organizes hierarchically the prosodic words and is not level limited. Prosodic words have no pre-established standard pattern, as their melodic characteristics depend on the application of 2 rules:

IMS: Inversion of Melodic Slope rule

AMV: Amplitude of Melodic Variation rule.

The description of the final accent of a prosodic word usually uses phonetic features such as Length (i.e. syllable duration), melodic Rise or Fall, Amplitude of melodic variation, etc. Initial (secondary) accents do not play a role in the marking of the prosodic structure, and are therefore normally described with a melodic rise. Their role is only to ensure the presence of at least one stress in sequences of 7 consecutive syllables.

Prepared speech intonation in French

In a phonosyntactic approach, the relationship between intonation and syntax in prepared speech is envisioned as follows: First a prosodic structure PS is assumed to exist in the sentence, independent but associated to the syntactic structure SS. In general, more than one PS can be associated to a given SS, the final choice being governed either by syntactic congruence or eurhythmicity, depending on the emphasis given to a) the syntactic hierarchy by the prosody, or b) by the balancing of the number of syllables at each level of the prosodic structure. More specifically:

1. The prosodic structure organizes hierarchically minimal prosodic words (stress groups);
2. Prosodic markers indicate the prosodic structure of the sentence;
3. Grammars of prosodic markers are specific to every language;
4. Specific realizations of prosodic markers characterize various dialects

The association between the syntactic and the prosodic structures is not straightforward, even in prepared speech. The constraints of this association can be summarized as follows:

- Planarity (no tangled structures);
- Connexity (no floating segments);
- Stress clash (no consecutive stressed syllables if the implied syntactic units are dominated by the same syntactic node);
- Syntactic clash (no prosodic grouping of stress groups – so at the lowest level in the structure - which are not themselves grouped in the syntactic tree by the same node);
- Stress group maximum number of syllables (a sequence of 7 syllables has a least one stress – either emphatic (narrow focus) or lexical, the number 7 depending on speech rate);
- Eurhythmicity (balancing the number of syllables in the prosodic structure, generally at the expense of congruence with syntax);
- Neutralization (phonological features not necessary to encode a given prosodic structure are not necessarily realized).

A 2 PW prosodic structure, instantiated by words of a sufficient number of syllables to involve a mandatory stress, would reveal a stress syllable whose phonetic realization has simply to be different from the above mentioned contours that could appear in its place. The following example shows this: *les hippopotames s'étaient étonnés*. The Subject NP and the VP contain each 5 syllables, forcing the realisation of a stress on the final syllable of *les hippopotames*. All the following examples use 5 syllables stress groups to achieve a rhythmically balanced prosodic structure.

Figure 1 (left). Pitch curve of the example *Les hippopotames s'étaient étonnés* with stressable syllables highlighted.

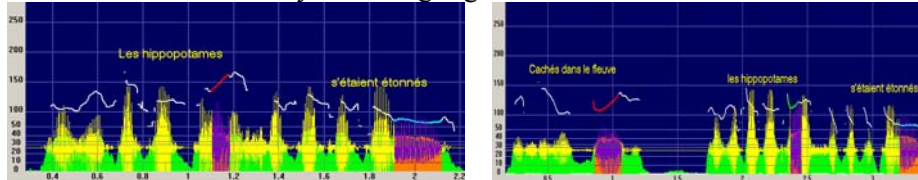


Figure 2 (right). Example *Cachés dans le fleuve les hippopotames s'étaient étonnés* with stressable syllables highlighted. In case of congruence between prosody and syntax, the accent on *fleuve* must be stronger than the one on *hippopotames*, which appears here as a larger rise of fundamental frequency and a longer syllable duration.

Figure 3 (left). Example *Les hippopotames étaient étonnés ils étaient cachés* with stressable syllables highlighted. In case of congruence between prosody and syntax, the accent on *étonnés* must be stronger than the one on *hippopotames*, which appears here as a contrast in melodic slope of fundamental frequency and a longer syllable duration.

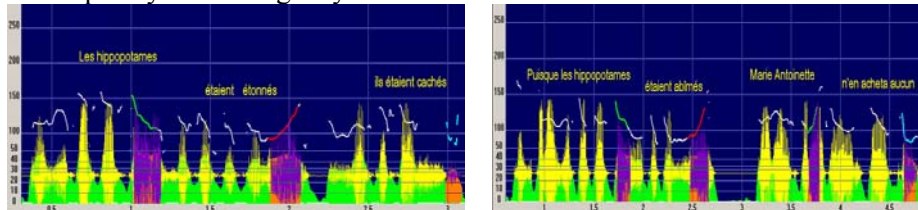


Figure 4 (right). Example *Puisque les hippopotames étaient abîmés Marie Antoinette n'en acheta aucun* with stressable syllables highlighted. In case of congruence between prosody and syntax, the accent on *abîmés* must be stronger than the ones on *hippopotames* and on *Marie Antoinette*, which is realized on the left by a contrast in melodic slope, and on the right by shorter syllable duration.

Spontaneous speech intonation in French

The basic idea, derived from the work of C. Blanche-Benveniste (1990, 2002), J. Deulofeu and collaborators from the GARS group in Aix-en-Provence, envisions the sentence in spontaneous speech as a sequence of macrosegments, syntactically well formed in the classical sense, and in relations of parataxis or rection with each other (in glossematic terms, in dependency relations of combination and selection).

One of these macrosegments has a special function and is called the Noyau: the Noyau contains information on the modality of the sentence,

constitutes a complete sentence by itself, its modality can be changed without affecting other macrosegments, as the change of modality (positive to negative, declarative into interrogative, etc.). In the sentence macrosegments placed before the Noyau are called prefixes, inside the Noyau Incises (imbedded), and after the Noyau Suffixes or Postfixes, depending on the syntactic or prosodic nature of their relationship with the Noyau (see below).

According to this view, prosodic structure indicates a hierarchical organization within the sentence, by defining the relationships between macrosegments.

Figure 5 (left). Prefix + Noyau structure. The prefix *le lendemain* is integrated in the sentence by the prosodic structure, which assembles it with the Noyau *grande surprise*. The prefix bears a final rising melodic contour, contrasting with the falling final declarative contour on the Noyau.

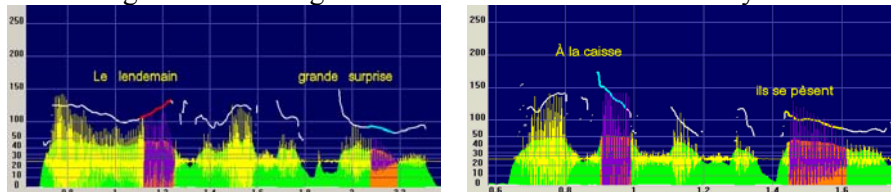


Figure 6 (right). Noyau + Postfix structure (= broad focus). The Noyau ends with a sharply falling melodic contour, whereas the Postfix ends with a falling declarative contour.

References

- Blanche-Benveniste, C., 2002, *Approches de la langue parlée en français*, Ophrys, Paris.
- Boulakia, G., Deulofeu, J. and Martin, Ph., 2001, Prosodic features finish off ill-formed utterances, don't they?, Proc. Congreso de Fonetica Experimental, Universidad de Sevilla, España, 5-7 mars 2001.
- Deulofeu, J., 2003, L'approche macrosyntaxique en syntaxe : un nouveau modèle de rasoir d'Occam contre les notions inutiles, *Scolia*, n° 16, Publications de l'Université de Strasbourg.
- Jun, S-A, and Fougeron, C., 2002. Realizations of Accentual Phrase in French Intonation, *Probus* 14, 147-172.
- Martin, Ph., 1987. Prosodic and Rhythmic Structures in French, *Linguistics*, 25-5, 925-949.
- Martin, Ph., 2004. L'intonation de la phrase dans les langues romanes : l'exception du français, *Langue française*, mars 2004, 36-55.
- Rossi, M. 1999. *L'intonation le Système du Français: description et modélisation*, Ophrys, Paris.

Effects of structural prominence on anaphora: The case of relative clauses

Eleni Miltsakaki, Paschalia Patsala
School of English, Aristotle University of Thessaloniki, Greece

Abstract

In this paper we present a corpus study and a sentence completion experiment designed to evaluate the discourse prominence of entities evoked in relative clauses. The corpus study shows a preference for referring expressions after a sentence final relative clause to select a matrix clause entity as their antecedents. In the sentence completion experiment, we evaluated the potential effect of head type (restrictive relative clauses are contrasted with non-restrictives and restrictives with an indefinite head). The experimental data show that the matrix clause subject referent is strongly preferred as an antecedent, thus strengthening the conclusion that entities evoked in relative clauses are less salient than their main clause counterparts. Some remaining issues are discussed.

Introduction

With the exception of a limited set of pronouns which are interpreted according to grammatical rules (e.g., Reinhart 1997), referential pronouns refer to contextually salient antecedents. Prior work on the relationship between discourse salience and the choice of referring expression has evaluated several factors. Most notably, structural focusing accounts such as Centering, a model of local coherence in discourse, argue that pronouns select antecedents which are highly accessible with discourse topics being the most prominent of all (Ariel 1990, Grosz et al 1995). At least for English, subjects rank high on salience. Semantic and pragmatic focusing accounts have examined the effect of thematic roles and the semantics of connectives in determining entity salience. Stevenson et al (2000), for example, argue that the focusing properties of action verbs make ‘patients’ more salient than ‘agents’ independently of grammatical role.

Note that most of the related work in this area has examined sequences of simple sentences. The aim of the present study is to advance our understanding of the factors determining the salience status of individual entities in discourse by examining entities in complex sentences. Specifically, we designed a corpus study and a sentence completion task to compare the salience status of entities evoked in main and relative clauses.

Previous work on Relative Clauses

The syntax and semantics of relative clauses have been the subject of a huge literature (e.g., McCawley (1981)). To-date the debate is still on regarding the appropriate syntactic analysis of relative clauses.

Relative clauses with resumptive pronouns have also been the source of several syntactic puzzles. Prince (1990) investigated the discourse functions of relative clauses containing a resumptive pronoun in English and Yiddish. Based on a corpus of naturally occurring relative clauses with resumptive pronouns, she argues that there is a set of data which cannot be explained based on previous accounts. Specifically, she finds that, for these data, resumptive pronouns are licensed in the case of non-restrictive and restrictive relative clauses with an indefinite head but not in the case of restrictives with a definite head. She argues that this phenomenon can be explained with Heim's file card metaphor. Resumptive pronouns are licensed when an entity has already been evoked in the discourse and is therefore available for pronominal reference.

Fox and Thompson (1990) avoided the distinction between restrictive and non-restrictive relative clauses. In their corpus analysis, they looked at discourse properties of relative clauses and argued that the attested discourse functions of relative clauses accounts for the grammatical properties of relative clauses.

Note that no claims have been made yet regarding the discourse salience of the entities evoked in relative clauses. Miltsakaki (2005) compared the salience of entities in main and relative clauses of the English and Greek language. Based on a centering analysis of the data, she concludes that in contrast with main clause subjects, subjects of relative clauses do not always warrant pronominal reference.

The Corpus Study

The dataset of our corpus study was constructed from a corpus of ten literary works available from the Project Gutenberg Literary Archive. We extracted 100 tokens of relative clauses according to the following criteria: a) the relative clause was in a sentence final-position, b) at least two animate entities were evoked in the main clause, and c) the sentence following the relative clause included reference to at least one entity evoked in the sentence containing the relative clause—either in the main or in the relative clause. For each token, we annotated the grammatical role of the relativised entity in the main clause, the relativised entity in the relative clause and the type of referring expression in the following clause.

The analysis of our data reveals the following patterns of reference. In 64% the antecedent was evoked in the matrix clause only, shown in (1) and (2). In an additional 31%, the antecedent was evoked in both clauses, and only in 5% the antecedent was evoked exclusively in the relative clause, shown in (3).

- (1) Has no letter been left here for me since we went out? said **she** to the footman who then entered with the parcels. **She** was answered in the negative.
- (2) Then **Huck** told his entire adventure in confidence to Tom, who had only heard of the Welshman's part of it before. "Well," said **Huck**...
- (3) The Queen used to ask me about the English noble who was always quarrelling with **the cabmen** about their fares. **They** made...

With respect to the 64% of our tokens in which the antecedent was evoked in the main clause only, in the 75% of cases, the antecedent of the referring expression was evoked in the subject position of the main clause and 9% the antecedent was evoked in the object position of the main clause. As for the grammatical role of the relativised entity in the relative clause itself, in the 85% of our tokens it was the subject of the relative clause, in the 8% the object, whereas in the 7% a PP complement.

The experiment

In this study, we tested the potential effect of the information status of the entity evoked in the head noun of the relative clause (see discussion of Prince (1990) in Section 2). To test the hypothesis that non-restrictives and restrictives with an indefinite head pattern alike and are processed as autonomous discourse units on a par with main clauses, we designed a sentence completion study with three conditions, sampled below:

1. Non-restrictive, Head=Proper noun (PN)

Samantha met Jennifer who played in Friends. She...

2. Restrictive, Head=Indefinite noun (IN)

Matthew adopted a boy who lost his family in civil war. He...

3. Restrictive, Head=Definite noun (DN)

The professor collaborated with the guy who was hired last month. He...

A total of 15 native speakers of English were asked to write a natural continuation for 12 critical items each (and 36 fillers). We counted how many times the ambiguous pronoun was interpreted as the main clause subject. An ANOVA analysis of the results did not show any significant effect of the head type, as in the majority of the data the pronoun was interpreted as the main clause subject (76% with PN head, 73% with IN head and 81% with a DN head). So, in the absence of a larger context, main clause subjects appear to be more salient than relative clause subjects, but looking closer at the data we see that IN restrictives pattern more closely with PN non-restrictives.

Conclusions

The results of both the corpus and the sentence completion study reveal that main clause referents make better antecedents for subsequent referring expressions, including pronouns. It is therefore clear that discourse salience is sensitive to structural prominence, i.e., main clause entities are more salient than relative clause entities. However, scrutinizing the data we observe that in some cases other discourse factors might be interacting with structural prominence. In the sentence completion study, we saw some variation in the three conditions which is not significant but gives some hints for further study. Also, when we look closer at the 31% of cases in the corpus study in which the antecedent of the referring expression was present in both the main and the relative clause, we observe that in most of these cases the antecedent was the object of the main clause and the subject of the relative clause, shown in (5).

(5) she carried me to **the king**, who was then retired to his cabinet. **His majesty**, a prince of much gravity and austere countenance, not well observing my shape at first view, asked the queen,...

Further study is clearly required to understand what the conditions are under which the otherwise strong effect of structural prominence is overridden. We suspect that a promising avenue of research would take into account the effects of the hierarchical organization of the discourse.

References

- Ariel, M. 1990. *Accessing NP antecedents*. London, Routledge.
- Fox B.A. and Thompson S.A. 1990. A Discourse Explanation of the Grammar of Relative Clauses in English Conversation. *Language* vol. 66, 2, 297-316.
- Grosz, B.J., A.K. Joshi and S. Weinstein. 1995. Centering: A Framework for Modelling the Local Coherence of Discourse. *Computational Linguistics* 21, 2, 203-225.
- McCawley, J.D. 1981. The syntax and semantics of English relative clauses. *Lingua* 53, 99-149.
- Miltsakaki, E. 2005. A Centering Analysis of Relative Clauses in English and Greek. Proceedings of the 28th Penn Linguistics Colloquium, University of Pennsylvania, Philadelphia.
- Prince, E.F. 1990. Syntax and Discourse: A Look at Resumptive Pronouns. In Hall, K. et al., eds. Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society, 482-497.
- Reinhart, T. 1997. Quantifier-Scope: How labor is divided between QR and choice functions. *Linguistics and Philosophy*, 20, 335-397.
- Stevenson, R., A. Knott, J. Oberlander and S. McDonald. 2000. Interpreting Pronouns and Connectives: Interactions among Focusing, Thematic Roles and Coherence Relations. *Language and Cognitive Processes*, 15, 3, 225-262.

Speaker based segmentation on broadcast news- on the use of ISI technique

S. Ouamour¹, M. Guerti² and H. Sayoud¹

¹ USTHB, Electronics Institute, BP 32 Bab Ezzouar, Alger, Algeria

² ENP, Hacem Badi, El-Harrach Alger, Algeria

Abstract

In this paper we propose a new segmentation technique called ISI or “Interlaced Speech Indexing”, developed and implemented for the task of broadcast news indexing. It consists in finding the identity of a well-defined speaker and the moments of his interventions inside an audio document, in order to access rapidly, directly and easily to his speech and then to his talk. Our segmentation procedure is based on an interlaced equidistant segmentation (IES) associated with our new ISI algorithm. This approach uses a speaker identification method based on Second Order Statistical Measures. As SOSM measures, we choose the “ μGc ” one, which is based on the covariance matrix. However, experiments showed that this method needs, at least, a speech length of 2 seconds, which means that the segmentation resolution will be 2 seconds. By combining the SOSM with the new Indexing technique (ISI), we demonstrate that the average segmentation error is reduced to only 0.5 second, which is more accurate and more interesting for real-time applications. Results indicate that this association provides a high resolution and a high tracking performance: the indexing score (percentage of correctly labelled segments) is 95% on TIMIT database and 92.4% on Hub4 Broadcast news 96 database.

Introduction

Speaker tracking consists in finding, in an audio document, all the occurrences of a particular speaker (target). But with the evolution of the information technology and the communications (broadcasting satellite, internet, etc), there are thousands of television and radio channels which transmit a huge quantity of information. Among this incredible number of information, finding the utterances and their corresponding moments of one particular speaker in an audio document requires that these documents must be properly archived and accessed, for this purpose many existing techniques are using different keys (keyword, key topic, etc), however these techniques can be not efficient enough for the task of speaker tracking in audio documents. A more suitable key for this task could be the speaker identity.

In that sense, the speaker is known a-priori by the system (i.e. a model of his features is available in the reference book of the system). Then, the task of indexing can be seen, herein, as a speaker verification task applied locally along a document containing multiple (and unknown) interventions of vari-

ous speakers: *Speaker Detection*. The Begin/End points of the tracked speaker interventions have to be found during the process. At the end of this process, the different utterances of the tracked speaker are gathered to obtain the global speech of this particular speaker in the whole audio document.

Thus, the research work presented in this paper is set in this context. So, we have developed for this task, a new system based on SOSM measures and a new interlaced speech indexing algorithm. This algorithm is easy to implement, simple and efficient since it significantly improves the results.

Speaker detection and tracking

Speaker tracking is the process of following who says what in an audio stream (Delacourt 2000, Bonastre 2000). Our speaker identification method is based on mono-Gaussian models and uses some measures of similarity called Second Order Statistical Measures (Gish 1990, Bimbot 1995). In our experiments we used the μGc measure (based on the covariance matrix).

A. Interlaced segmentation

In our application, we divide the speech signal into two groups of uniform segments, in which each segment has a length of 2 seconds. The second segment group is delayed from the first one by a delay of 1 second, i.e. the segments are overlapped by 50%. These two groups of segments, called respectively the odd sequence and the even sequence, form the interlaced segmentation.

B. Labeling

Once the covariance has been computed for each segment, some distance measures (μGc) are used in order to find the nearest reference for each segment (in a 24-dimensional space).

Once the minimal distance between the segment features and the reference features (e.g. corresponding to speaker L_j) is found, the segment is labeled by the identity of this reference (speaker L_j). Thus, this process continues until the last segment of the speech file. Finally, we obtain two labeling sequences corresponding to an even labeling and an odd labeling, as shown in figure 1.

C. Interlaced speech indexing (ISI)

The ISI algorithm is a new technique in which there are two segmentations (one displaced from the other) and a logical scheme is used to find the best speaker labels, by combining the two segmentation sequences.

Having two different indexing sequences, we try to give a reasonable labeling compromise between the two previous labeling sequences. Thus, we divide each segment into two other similar segments (of 1 second each), called sub-segments, so that we obtain “2n” even labels (denoted by $L_{\text{even}}^{1/2}$)

for the even sub-segments and “ $2n+2$ ” odd labels (denoted by $L^{1/2'}_{\text{odd}}$) for the odd sub-segments. Herein, $L^{1/2'}_{\text{even}}$ and $L^{1/2'}_{\text{odd}}$ are called sub-labels. Our intuition would be that the even sub-label and the odd sub-label at the same sub-segment should be the same, therefore we must compare $L^{1/2'}_{\text{even}}(j)$ with $L^{1/2'}_{\text{odd}}(j)$ for each sub-segment j . Herein, two cases are possible:

- **if** $L^{1/2'}_{\text{even}}(j) = L^{1/2'}_{\text{odd}}(j)$ **then** the label is correct:

$$\text{new label} = \text{correct label} = L^{1/2'}(j) = L^{1/2'}_{\text{even}}(j) = L^{1/2'}_{\text{odd}}(j) \quad (1)$$

- **if** $L^{1/2'}_{\text{even}}(j) \neq L^{1/2'}_{\text{odd}}(j)$ **then** the label is confused:

$$\text{new label} = L^{1/2'}(j) = \text{Cf} \quad (2)$$

where $L^{1/2'}$ represents a sub-label and Cf means a confusion.

In case of confusion, we derive a new algorithm called “ISI correction”.

Algorithm of ISI correction: In case of confusion, we divide the corresponding sub-segments (of 1 s) into two other sub-segments of 0.5 second each, called micro-segments. Theirs labels, called micro-labels, are denoted by $L^{1/4'}$. The correction algorithm is then given by:

- **if** $\{ L^{1/4'}(j) = \text{Cf} \text{ and } L^{1/4'}(j+1) = \text{Cf} \text{ and } L^{1/4'}(j-1) \neq \text{Cf} \}$
then $L^{1/4'}(j) = L^{1/4'}(j-1)$ (3)
 this is called a left correction (see the micro-segment j_0 in figure 1),

- **if** $\{ L^{1/4'}(j) = \text{Cf} \text{ and } L^{1/4'}(j-1) = \text{Cf} \text{ and } L^{1/4'}(j+1) \neq \text{Cf} \}$
then $L^{1/4'}(j) = L^{1/4'}(j+1)$ (4)
 this is called a right correction (see the micro-segment j_1 in figure 1).

Where, $L^{1/4'}$ denotes a micro-label for a micro-segment of 0.5 second.

Results and discussions

The first test database consists of several utterances from TIMIT uttered by different speakers and concatenated into speech files.

Table 1: Tracking error for discussions between several speakers.

		Tracking error (%) for discussions between:			
		2 speakers	3 speakers	5 speakers	10 speakers
Clean speech	With silence detection	7,2	8,1	7,9	10,3
	Without silence detection	5,3	7,3	5,9	8,0
Music + speech	Without silence detection	4,8	6,6	7,5	9,1
Corrupted speech at 12 dB	Background noise	26,0	55,7	53,7	67,2
	Office noise	19,9	24,3	57,6	66,1
	Human noise	9,1	7,9	23,0	19,9
Corrupted speech at 6 dB	Background noise	32,8	58,4	64,7	79,1
	Office noise	28,1	37,7	63,4	70,6
	Human noise	11,8	12,9	15,5	24,3

Each speech file contains several sequences of utterances from different speakers and with several speaker transitions per file. In order to investigate the robustness of our method, one part of the database is mixed with noise and music. In table 1, we note that the tracking error increases if the number of speakers increases too. For example, in case of clean speech, the error is only 5.3% for 2 speakers and it is 7.3% for 3 speakers. Concerning the different noises added in this experiment, we see that human noise do not disturb significantly the speaker tracking (degradation of 4% at 12dB) which implies that this type of noise may not disturb the tracking, considerably. The other speech data used in the experiments are extracted from the *HUB-4 1996-Broadcast-News* and consists of natural news.

Here we note that the tracking error obtained after ISI correction is lower than that obtained without ISI correction. For example, if the segment duration is 3 seconds, the error of tracking without ISI correction is about 9% but it decreases to 7.7% when an ISI correction with two iterations is applied and decreases to 7.6% when an ISI correction with four iterations is applied.

Moreover, we notice that the best tracking is got for segments duration of 3s.

Conclusion

Experiments done on corrupted speech and on *Hub4 Broadcast News* indicate that the ISI technique improve both the indexing precision and the segmentation resolution. Furthermore, they show that the best segment duration for speech segmentation is 3 seconds.

In general, compared to previous works, this method gives interesting results. Although it is difficult to compare objectively the performances of all the existing methods, we believe that this technique represents a good speaker indexing approach, since it is easy to implement, inexpensive in computation and provides good performances.

References

- Bimbot F. et al. 1995. Second-Order Statistical measures for text-independent Broadcaster Identification. *Speech Communication*, 17, 177-192.
- Bonastre J.F. et al. 2000. A speaker tracking system based on speaker turn detection for NIST evaluation. *IEEE ICASSP*, Istanbul, june 2000.
- Delacourt P. et al. 2000. DISTBIC: a speaker-based segmentation for audio data indexing, *Speech Communication*, 32, Issue 1-2.
- Gish H. 1990. Robust discrimination in automatic speaker identification. *IEEE Inter. Conference on Acoustics Speech and Signal Processing*. April 90, New Mexico, 289-292.
- Liu D., and Kubala F. 1999, "Fast speaker change detection for broadcast news transcription and indexing". *Eurospeech*, 1999. Vol. 3, 1031-1034.
- Reynolds D.A. et al. 1998, "Blind clustering of speech utterances based on speaker and language characteristics". *ICSLP*, 1998. Vol. 7, 3193-3196.

The residence in the country of the target language and its influence to the writings of Greek learners of French

Zafeiroula Papadopoulou

Didactique des langues et des cultures, Université de la Sorbonne Nouvelle - Paris 3, France

Abstract

The study of linguistic acquisition implies the study of the contexts where this process evolves. In order to analyze the role of the residence in the country of the target language, we formed two groups of informants (one in Greece, one in France). In order to make this comparison, we chose to analyze the reference to the entities and the temporal reference in the texts collected, and we hoped to show the influence of a residence in France.

Rationale

The aim of this research is to examine the relation between the references to the people and to the time and the textual cohesion in texts produced by various groups of script writers, by wondering about the independent variable “residence in the country of the target language”.

Theoretical base

This research reposes to the model of the *quaestio* of Stutterheim and Klein (1991). **The *quaestio* is the general/abstract question to which any text answers.** The speaker, each time he constructs a narration, has to answer an implicit question having the general form “What happens to P then?” or “What does it occur to P at time T?” P representing the protagonists.

The *quaestio* exerts a constraint on the local level of the enunciation, into two parts: the topic (T) and the focus (F). The **topic** is that about which the speaker speaks, the support reference frame, and ensures the continuity of the text according to the rule of the repetition, while the **focus** is what one says of the topic, the contribution of information which brings new information, and ensures its progression

Recueil of the corpus

Our corpus is composed of narrations requested using extracts of two films. The first montage consists of eight photographs of the film “American pie”.

At the beginning a young man is looking at a girl, who is playing the flute, in front of whom a group of girls, in row, also play the flute. Once the concert is over, and the girls leave the scene, our hero decides to go and speak to the girl. When she notices him, she seems surprised.

The second series of photographs is an extract of the film “The pink panther”. The scene takes place on the island Saint Louis in Paris, where a man exchanges something with a woman. Two police cars arrive, and the heroes start running. After entering a hotel, the woman enters the elevator surveyed by a monk and the receptionist. A police officer waits outside the elevator and the two other take the staircases. At the same time in the elevator, the woman changes her appearance. After that disguise, the policemen fail to stop the suspect woman.

The productions of two groups were collected, resulting 20 texts coming from: 10 Greek learners of French who never lived within a French-speaking community (G1) and 10 Greek learners of French who live at the time of the collection of the data in France (G2). It is necessary to add that French is not their second language but their third. All the informants speak already English.

Analysis

Let us come now to the results of our research: concerning the reference to the entities, we noticed entities to a position T or F during their introduction. Their position constitutes an indicator of the hierarchy of these entities in the two stories.

In the film *American Pie*, for the G1 the entities occupy a position T or F according to their appearance to the film. For the film *Pink Panther*, the police force is, most of the time in T position. Eight people treat the history as if the monk and the receptionist did not exist. An explanation could be the ignorance of the words *monk* and *receptionist*. Concerning the G2 the role of the protagonists is shared for two films.

As for the maintenance of the entities, a great percentage of substitution of the entities is made by a pronoun. The pronouns indicate mutually known entities and are used to avoid the repetition. In French, where the subject is obligatory, the maintenance of the reference is marked by a pronoun. In our study, we distinguish that the pronouns the most often used are the personal pronouns

Most of our informants (16/20) choose the possessive in order to maintain an entity. The possessive expresses various semantic relationships such as: property (« *sa robe* »), characteristic (« *son apparence* »), semantic roles associated to a process (« *son attention* ») as well as possession itself.

The studies relating to the maintenance of the reference also show that the languages tend to mark the degree of accessibility to the entities by more or less explicit forms. In general, there is pluralism concerning the full forms which refer to the protagonists. By order of frequency the full forms used for the maintenance 'the young man', 'the girl' and 'girls' for the first history and 'the man', 'the woman', and 'the police officers', for the second one, are used massively to reintroduce a distant referent and to promote it to a T position.

Finally the reinforcement of cohesion is made by the co-presence of 2 or 3 entities within the same sentence or by the forms of recovery 'ça' or 'là'.

According to Klinger (2003) the cohesion of a text is ensured by the reference to the acting entities and by the temporal reference. Regarding the temporal reference in our corpus we observe the almost total dominance of present. And it is essential to mention that the subjects also express temporality through the grammatical and lexical aspect. Finally we have observed in the corpus temporal markers (connectors, prepositions) used to reinforce cohesion.

Discussion

Our study relates to advanced learners. Similar researches exist, for example, that of InterFra. As for the results of the project of InterFra, we could notice that our informers, like those of InterFra, acquire the rules of the verbal agreement. Another project is that of ESF (Klein W. & Perdue C., 1997) where the learners manage to build phrases rich in progression and temporal returns. Our informers, according to their productions show a good use of verbal morphology; but it should not be forgotten that our informers follow or have already followed French courses in Greece.

Following Dabène (1990), we distinguish our groups according to their residence. The first group is the case of the *exolingue* situation, in other words they live in a country where another language than the TL, whereas the second group is the case of the *endolingue* situation. Our last objective consisted of making a comparative analysis between the two groups. We seize that the differences between the two groups relate to the sociolinguistic competences, the lexical level and the length of the accounts.

Research relating to advanced learners during a stay in the native community shows that as a whole, there is no spectacular development at the structural level. We have nevertheless, the impression that they have a benefit of the stay in the foreign country, in particular in the development of various aspects of the sociolinguistic capacities. We observe for example in our corpus, that six people out of ten, of the G2, identify the bridge of the island Saint Louis which is not the case for the G1.

Moreover, we find enunciations which do not include characters. These enunciations reinforce the narrative, and they are used by the learners of the G2 who are more analytical in the description.

Another very interesting remark consists of the number of the temporal markers. The percentage of the G1 is much higher than that of the second group. The informers who live in Greece used in the 179 propositions 104 markers. However, in the 296 propositions of the second group, 82 markers were used. We can explain this phenomenon by the fact that the informers in Greece use more marks so as to explicate the temporal relations, while those in Paris, express these relations by verbal morphology and by the aspect (lexical and grammatical) At the lexical level also, there are differences between the two groups. If we observe their productions attentively, we understand that the group of Paris, use words which are not 'institutionally' taught like "*le flic*" (the cop), "*la gendarmerie*", "*merde*" (shit) and which are normally learned by listening to them.

Conclusion

In conclusion, the result of this research is that the residence in a native community is essential for the improvement of sociolinguistic and lexical competences, but only the exposure to the TL in the natural environment does not improve morphosyntactic competence.

References

- Clerc, St. 2003. L'acquisition des conduites narratives en français langue étrangère www.marges-linguistiques.com, 1-18.
- Dabène, L. and Cicurel, F. and Lauga-Amid, M.-C. and Foerster, C. 1990. Variations et rituels en classe de langue, Paris, coll. LAL, Hatier Credif.
- Juli, S. 2003. Comment se développe la morphologie verbale en français L2 chez les sinophones ? www.marges-linguistiques.com, 1-13.
- Klein, W. and Von Stutterheim, Ch. 1991. Text structure and referential movement. Arbeitsberichte des Forschungsprogramms Sprache und Pragmatik. Lund University.
- Klinger, D. 2003. Raconter dans deux langues : traitement et fonction des connecteurs dans le cadre expositif et le déclenchement de la narration. Étude menée sur la production française (L2) et japonaise (L1) de deux locutrices japonophones, www.marges-linguistiques.com, pp. 1-14.
- Lambert, M. 2003. Cohésion et connexité dans des récits d'enfant et d'apprenants polonophones du français, www.marges-linguistiques.com, 106-121.

The Acquisition of Epistemic Modality

Anna Papafragou¹ and Ozge Isik Ozturk²

¹Department of Psychology, University of Delaware, USA

²Department of Linguistics, University of Delaware, USA

Abstract

In this paper we try to contribute to the body of knowledge about the acquisition of English epistemic modal verbs (e.g. Mary may/has to be at school). Semantically, these verbs encode possibility or necessity with respect to available evidence. Pragmatically, the use of epistemic modals often gives rise to scalar conversational inferences (Mary may be at school -> Mary doesn't have to be at school). The acquisition of epistemic modals is challenging for children on both these levels. In this paper, we present findings from two studies which were conducted with 5-year-old children and adults. Our findings, unlike previous work, show that 5-yr-olds have mastered epistemic modal semantics, including the notions of necessity and possibility. However, they are still in the process of acquiring epistemic modal pragmatics.

Introduction

This paper is concerned with the acquisition of semantics and pragmatics of epistemic modals. Epistemic modal verbs encode the speaker's certainty towards the probability or predictability status of the proposition embedded under the modal verb. The sentences in (1) are examples of epistemic modality:

- (1) a. It has to rain in the afternoon.
- b. It may rain afternoon.

On the semantic level epistemic modal operators encode modal force (necessity or possibility) and get interpreted against a conversational background which is a function from possible worlds into sets of propositions. Necessity in a given world encodes truth in all alternative possible worlds, whereas possibility encodes truth in at least one alternative possible world (Hintikka, 1969).

On the pragmatic level, epistemic modal verbs typically give rise to conversational implicatures of the following sort:

- (1) a. It has to rain in the afternoon.
- b. It may rain in the afternoon.
- (2) It does not have to rain in the afternoon.

Logically, (1b) is compatible with (1a). However, in conversation, (1b) excludes (1a) – hence it implicates (2).

In order for the child to acquire epistemic modality, he/she needs to acquire both the semantic aspects of modal meaning (including the notions of possibility and necessity) and the pragmatic inferences associated with modal expressions. Our goal in this paper is to shed light on the processes underlying the acquisition of epistemic modality.

Experiment

Method

Participants

A total of 40 native English-speaking-children (mean age: 5;8 mo) and 40 native English-speaking adults participated in this study.

Stimuli and Procedure

Participants were presented with eight short animated stories on a computer screen. The experimenter told the participant that they would play a game together with two puppets (Minnie and Daisy) which were seated across the computer and several animals which are computer-animated. The test phase of the experiment involved a stage on the screen whose curtains could be lowered and two containers (identical in shape and size but different in color). The experimenter told the participant that each of the animals would hide in one of the boxes on the screen while the stage curtains were lowered, after the curtains were lifted up again, Minnie and Daisy would take turns and guess in which box the animal had hidden. Participants had to say whether they agreed with each puppet or not.

Stories and statements were identical in both conditions except for the modal verb used in the puppets' guesses (*may* in the Possibility and *have to* in the Necessity condition). For instance, in one of the stories a mouse hid in one of the two boxes (a yellow or a pink one) while the curtains were lowered. Each story gave the puppets two opportunities to guess. In the first guessing phase (closed boxes phase) the puppets made the following guesses right after the animal was hidden but before any of the boxes were opened:

Possibility Condition

(3) Minnie: "The mouse may be in the yellow box." (True)

(4) Daisy: "The mouse may be in the pink box." (True)

Necessity Condition

(5) Minnie: "The mouse has to be in the yellow box." (False)

(6) Daisy: "The mouse has to be in the pink box." (False)

After each statement the experimenter asked the participant whether or not the puppet was right. Because of the design of the task, both puppets are correct in the Possibility condition and incorrect in the Necessity condition during this first phase of each story.

In the next guessing phase, one of the boxes was opened. In four of the stories, it revealed that it had the animal inside. For instance, in our earlier story, the yellow box was opened to reveal the animal. The experimenter again asked each of the puppets where the animal was hidden. Depending on the condition the child was assigned to, the puppets offered the answers (3-6). Again, the experimenter asked the participant whether or not the puppet was right.

In the remaining four stories, there was no animal hidden in the opened box. For instance, a cow hid in one of the two boxes (an orange and a blue one). After the first guessing round, the blue box was opened and was found empty. The puppets offered the answers given below:

Possibility Condition

(7) Minnie: "The cow may be in the orange box." (False)

(8) Daisy: "The cow may be in the blue box." (True but under-informative)

Necessity Condition

(9) Minnie: "The cow has to be in the orange box." (False)

(10) Daisy: "The cow has to be in the blue box." (True but under-informative)

Adults were tested individually in the same way as the children. They were expected to accept the true statements and to reject the false statements in both the Possibility and the Necessity conditions; we also expected them to accept the under-informative statements in both conditions on semantic grounds, even though pragmatic responses were also acceptable. We were interested in examining whether five-year-old children lack the correct semantics and pragmatics for epistemic modals, as previous work has suggested (Noveck, 2001) – hence, whether their responses would be different from adults'.

Results

Overall, we found that 5-year-olds have acquired the semantics of the modal of Possibility *may* and of the modal of Necessity *have to*, since they successfully accept true modal statements and reject false ones most of the time (for fuller discussion, see Papafragou & Ozturk, 2006). As expected, we found that adults' performance was better in both the Possibility and Necessity conditions. However, the question of whether 5-year-olds treat a relatively weaker term logically or pragmatically remains open as both the children

and our adult participants treated these items semantically and not pragmatically (i.e., they did not reject true but underinformative statements).

We have planned future research to examine further questions, specifically, whether children are aware of epistemic modal scales, whether they would be able to treat an informationally weak modal verb pragmatically if it were explicitly contrasted with a stronger modal, and whether they would prefer true modal statements over true but under-informative ones.

Conclusion

In this paper we investigated 5-year-old children's acquisition of epistemic modality. Interestingly, unlike earlier findings, our data show success on the part of young children with the modal concepts of necessity and possibility. Specifically, we have shown that 5-year-olds have acquired the concepts of possibility and necessity and they are able to reason about statements based on these concepts. These findings may provide an opportunity for linguistic developmental data to throw light on theories of conceptual development and the acquisition of the semantics-pragmatics interface.

References

- Hintikka, J. 1969. *Models for Modalities*, Reidel, Dordrecht.
- Noveck, I. A. 2001. When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, 78(2), 165-188.
- Papafragou, A., & Ozturk, O.I. 2006. On the acquisition of modality. To appear in *Penn Working Papers in Linguistics*. Dept. of Linguistics, UPenn.

Towards empirical dimensions for the classification of aphasic performance

Athanassios Protopapas¹, Spyridoula Varlokosta², Alexandra Economou³ and Maria Kakavoulia⁴

¹Institute for Language & Speech Processing, Maroussi, Greece

²Department of Mediterranean Studies, University of the Aegean, Greece

³Department of Psychology, University of Athens, Greece

⁴Department of Communication, Media, and Culture, Panteion University, Greece

Abstract

We present a study of 13 patients with aphasia, not screened by presumed subtype, showing strong correlations among disparate measures of fluency and measures of receptive and expressive grammatical ability related to verb functional categories. The findings are consistent with a single underlying dimension of severity. We suggest that subtyping be re-examined in light of performance patterns and only accepted when patient clustering is empirically derived and theoretically meaningful.

Introduction

Patterns of breakdown in aphasia can be informative about the human cognitive system of language. Classical neurological and aphasiological taxonomy use localization and clinical criteria to distinguish among subtypes; for example, fluent vs. nonfluent, expressive vs. receptive, or structural vs. semantic. These distinctions have important implications for the conceptualization of language ability, implying that distinct dimensions of skill underlie observed performance variance. However, clinical practice suggests that up to 80% of patients with aphasia cannot be clearly classified, depending on the classification scheme and diagnostic instrument (Spreen & Risser 2003).

Furthermore, cross-linguistic evidence has led to re-evaluation of certain assumptions on which subtyping is typically based, and has highlighted the role of language-specific properties (Bates et al. 2001). The different opportunities for linguistic analysis and performance breakdown patterns offered by different languages have made cross-linguistic research indispensable in aphasia. In the case of Greek, the rich verbal morphology allows the study of functional categories in situations of controlled structural complexity and in relation to more global assessments such as fluency and severity.

An inclusive approach to participant selection permits objective comparisons on the basis of performance patterns rather than a-priori categorization potentially leading to selection bias. If there is a valid categorization of patient per-

formance patterns into clinically useful subtypes, then this should emerge empirically as a result of clustering and dissociation analyses. In this paper we extend the study of Varlokosta et al. (in press) with measures of speech production and a new group of patients, and we suggest an experimental methodology for the study of aphasia based on patterns of covariance among measures of expressive and receptive language performance.

Method

Participants

Seven Greek-speaking men 42–81 years old diagnosed with aphasia formed patient group A. The details for this group and a control group matched on age, sex, and years of education can be found in Varlokosta et al. (in press). In addition, patient group B included 4 men and 2 women 42–72 years old diagnosed with aphasia. Patients were not screened for aphasia (sub)type.

Test materials and procedure

A grammaticality judgment test included 80 correct and 80 corresponding incorrect active-voice sentences manipulating verbal aspect, tense, and agreement with subject in number and person. A sentence completion test, using the same 80 sentence beginnings as cues and corresponding baseline sentences, was used to measure expressive performance. Verbs were controlled for phonological properties, regularity (in aspectual formation), and frequency (estimated via subjective familiarity). Details about these materials are reported in Varlokosta et al. (in press).

Patient group A and the corresponding control group were administered a brief interview, the sentence completion task, 2 standard picture description tasks (Cookie Theft and the store scene from Wechsler Memory Scale III), and the grammaticality judgment task, in this order. Patient group B was only administered the interview and picture description.

Results

Performance in verb production and reception revealed that aspect was most vulnerable whereas subject-verb agreement was most resistant (Varlokosta et al. in press). There was no dissociation between impairment in production and reception (Figure 1, left). Moreover, analysis of lexical errors separately from grammatical (morphological) errors showed that there was little basis for a dissociation among structural vs. semantic dimensions.

Here we have analyzed production performance (from the picture descriptions) with two quantitative indices: “fluency” and “mean length of utterance”

(MLU). As shown in Figure 1 (right), patient fluency was strongly correlated with MLU, and also with measures of grammatical performance (Table 1), suggesting a common underlying dimension of severity.

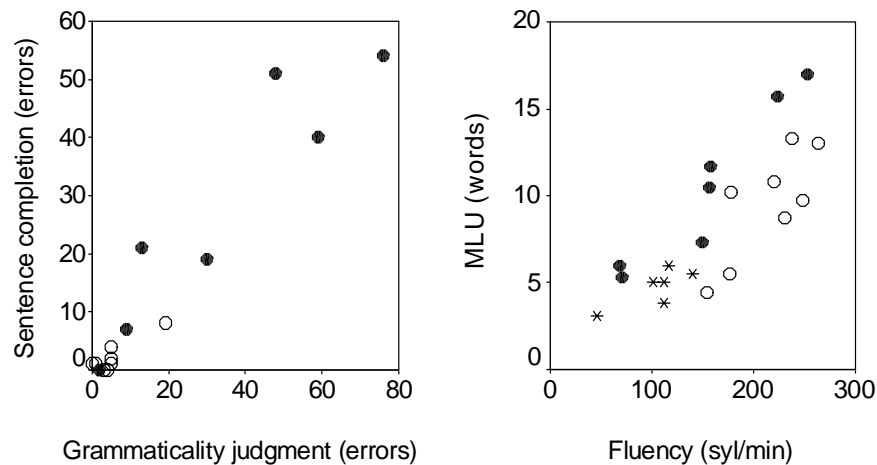


Figure 1. Interrelations among measures of grammatical performance (left) and production volume/rate (right). Filled circles: Patient group A; Open circles: Control group; Asterisks: Patient group B.

Table 1. Correlation coefficients (Pearson’s *r*) among measures. Above the diagonal, for Patient group A only (*N*=7). Below the diagonal, for all participants as available (*N*=15 except between MLU and fluency, where *N*=21).

(* <i>p</i> <0.05 ** <i>p</i> <0.005)	1	2	3	4	5
1 Fluency		-0.80*	-0.77*	0.96**	-0.41
2 Grammaticality judgment	-0.73**		0.93**	-0.80*	0.39
3 Sentence completion	-0.74**	0.95**		-0.86*	0.69
4 Mean length of utterance	0.80**	-0.37	-0.38		-0.65
5 Number of utterances	-0.40	0.24	0.42	-0.64*	

Discussion

Our findings do not support a dissociation between severity and fluency. Instead, the pattern of correlation among disparate measures of speech production and grammatical performance, both expressive and receptive, calls for a reconsideration of traditional groupings and highlights the need for additional cross-linguistic research, if confirmed with more patients and tests.

Dick et al. (2001) have demonstrated that language processing deficits can be revealed in unimpaired participants if tasks are sufficiently demanding. Extending this line of thinking, one might expect participants with aphasia to lie on

one side of a single language performance continuum instead of forming a qualitatively distinct group. Our data (Figure 1) offer partial support for this prediction in that control performance was largely overlapping with aphasic performance and apparently lying along a single line.

Structural accounts of language breakdown (e.g., Friedmann & Grodzinsky 1997), aiming to explain dissociations based on linguistic type differences, might have difficulty with unidimensional patterns of impairment. In contrast, processing accounts (e.g., Kolk & Hartsuiker 2000) may be at an advantage to the extent that patient performance can be reliably related to independent indices of severity of aphasia, task difficulty and cognitive capacity. Differences in the dynamics of lexical activation is one such attempt to explain the observed co-occurrence of language impairments (Blumstein & Milberg 2000).

We suggest that a wide range of measures, spanning distinct domains of performance, be administered in future studies of aphasia and that patients not be pre-selected by presumed subtype. Instead, aphasia subtyping should be re-examined in light of actual performance patterns, and only accepted when patient clustering is empirically derived and theoretically meaningful.

Acknowledgements

We thank N. Valeonti, M. Lazaridou, and F. Lykou for administering the assessment tests, and M. Diamanti, A. Xofillis, and M. Moudouris for referring their patients.

References

- Bates, E., Devescovi, A., and Wulfeck, B. 2000. Psycholinguistics: A cross-language perspective. *Annual Review of Psychology* 52, 369–396.
- Blumstein, S. E., and Milberg, W. P. 2000. Neural systems and language processing: Toward a synthetic approach. *Brain and Language* 71, 26–29.
- Dick, F., Bates, E., Wulfeck, B., Utman, J. A., Dronkers, N., and Gernsbacher, M. A. 2001. *Psychological Review* 108, 759–788.
- Friedman, N., and Grodzinsky, Y. 1997. Tense and agreement in agrammatic production. Pruning the syntactic tree. *Brain and Language* 56, 71–90.
- Kolk, H. J., and Hartsuiker, R. J. 2000. Agrammatic sentence processing: Severity, complexity, and priming. *Behavioral and Brain Sciences* 23, 39–40.
- Spree, O., and Risser, A. 2003. *Assessment of Aphasia*. Oxford Univ. Press.
- Varlokosta, S., Valeonti, N., Kakavoulia, M., Lazaridou, M., Economou, A., and Protopapas, A. in press. The breakdown of functional categories in Greek aphasia: Evidence from agreement, tense and aspect. *Aphasiology*.

Analysis of intonation in news presentation on television

Emma Rodero
Faculty of Communication, Pontificia University, Spain

Abstract

The effectiveness on the television communication is sustained on the handling of voice, which is the base of fundamental expression of the audio-visual contents. Consequently, the audience attention to the television message as well as understanding and assimilating of it, will depend on a correct and expressive use of voice. These reasons guarantee the necessity to make an empirical investigation about the prosody form in which the news broadcasters on television use voice. Therefore, the actual communication is based on the results of a study in order to determine the intonation in broadcasters on television.

Overview

There are not many empirical investigations about TV broadcaster's intonation on television. Nevertheless, there is a wide bibliography in Spain which continuously refers to the lack of effectiveness in the use of prosodic elements of journalists. Both experiments that have been carried out, and bibliographical references agree in signaling that there is a specific way of intonation of TV news which is, paradoxically, not the most appropriated, neither from a linguistic nor from an expressive point of view. In the presentation of news, journalists reproduce one singsong delivery pattern that represents the continuous repetition of a same structure: a circumflex intonation.

Method

According to this, the actual investigation tries to define the local movements of intonation which are more used in TV news to verify if this type of circumflex intonation in broadcaster's presentations does really take place. To prove this theory, the superficial phonological level of speech will be analyzed following the Aix-in Provence pattern.

The corpus of this research is composed by ninety utterances corresponding to the interventions of six broadcasters on television, fifteen utterances by each speaker, in the different news programs emitted by three main Spanish stations (two programs by a TV station). When selecting the definitive utterances for the analysis, two facts have been taken into account: utterances had to be declaratives and with similar extension.

Once the corpus has been obtained, the analysis of the local movements is made by using the Praat program together with the Momel pattern of stylization, and the Intsint transcription system. In order to analyze the local movements, in the first place, the stylization and the transcription of each one of the utterances that compose the corpus has been made. Once the errors of detection of the system have been corrected by hand, and being the target points labeled by means of the Intsint tones, movements are computed by building groups of three targets, since it is our aim to characterize local manifestations. Target points have been entered considering if they were at the beginning (three first tones), at the middle or at the end of the analyzed utterance (three last tones), to be able to establish similar comparisons between the presentations of all broadcasters.

Results

Once collected the data, they have been analyzed with the software of statistical analysis SPSS. With this program, the calculation of the relative frequency of appearance of each one of the analyzed sequences has been made, and the percentage has been extracted.

In the analysis of the corresponding data at the beginning of the utterances, 25 combinations for the 90 cases have been obtained. The results reveal that most of the speakers start their utterances with the combination of MTD (almost a 27 percent), BHL (almost a 17 percent), BHD (a 12 percent) and MUS (an 11 percent). The rest of the groups are away from it yet. In the graphic, combinations which obtained just one percent of the total have been eliminated (18 combinations of 25)

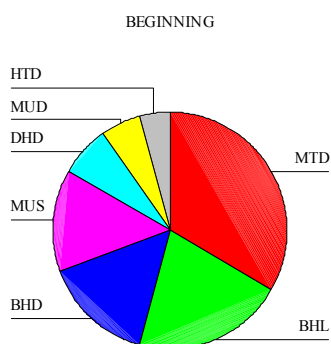


Figure 1: Percentage of frequency of appearance of pitch movements at the beginning of the utterance.

In the analysis of the segments corresponding to the middle of the utterances, 64 combinations for 1927 cases have been entered. The results

indicate that most broadcasters use sequences DUD (28 percent), UDU (19 percent), ULU (17 percent) and LUL (16 percent). The rest of the groups are further from them. In the graph, combinations which obtained less than one percent do not appear (57 combinations of 64).

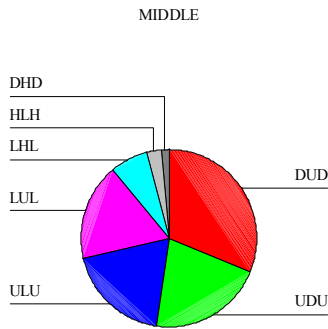


Figure 2: Percentage of frequency of appearance of pitch movements in the middle of the utterance.

Finally, the pitch movements at the end of utterances, in which 24 combinations for the 90 cases were obtained, have been analyzed. In this case, it is observed that the commonest grouping between speakers has been LTB (34 percent) followed by DTB (17 percent), DUD (14 percent) and LUL (11 percent). This descendent pattern takes place because the utterances are declaratives. The rest of them do not present significant differences. One percent is excluded from the graph (19 combinations of 24).

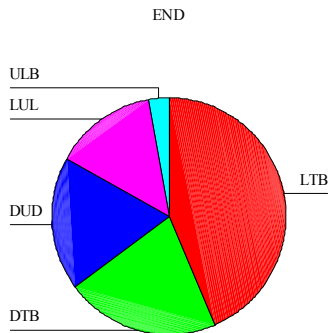


Figure 3: Percentage of frequency of appearance of pitch movements at the end of the utterance.

Discussion

As we have verified in the results, every pitch combination that news broadcasters use on television, at the beginning, in the middle and at the end of their utterances, tend to present a F0 circumflex contour, although it varies the pitch level.

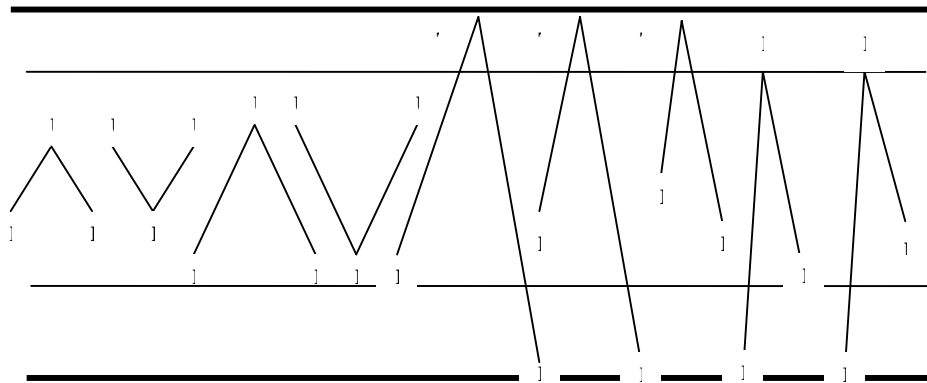


Figure 4: More frequent local movements.

It's certain that the results of the most frequent local movements do not throw too high percentages and that by just grouping the first options it is possible to collect data that are over fifty percent of the cases. Nevertheless, we must consider, firstly, that the number of possible combinations is very high and that most of them obtain pretty little frequencies of appearance. This fact contributes to show that the results are not so evident, since there is a great amount of groupings that appear only once.

Secondly, we must consider that, although there are different combinations which agglutinate the greater percentages, many of them are inverse or similar movements regarding acoustic impression, and they differ only in the ascendance or decline of pitch quantification. In conclusion, it is now considered that, at least in this analysis, the results gathered in the bibliography are according to the extracted empirical data in this study.

References

- Hirst, D., Di Cristo, A. and Espesser, R. 2000. Level of representation and levels of analysis for the description of intonation systems. In Horne, M. (ed.), *Prosody: Theory and Experiment*, Dordrecht, Kluwer Academic Press, 51-87.
- Louw, J.A. and Barnard E. 2001. *Automatic intonation modelling with Intsint*, Pretoria, Human Language Technologies Research Group, University of Pretoria.
- Prieto, P. 2003. *Teorías de la entonación*, Barcelona, Ariel.

Templates from syntax to morphology: affix ordering in Qafar

Pierre Rucart
Université Paris 7 Denis Diderot, France

Abstract

The functional head suggests that verbs acquire their inflectional properties by moving from one head position to the next in the syntactic derivation. A problem arises as affixes' ordering is not sensitive to syntactic properties, as it is the case in Qafar. This Cushitic language exhibits two verbal classes depending on whether verbs can have prefixes. I argue that the hierarchical structure of template corresponds to the syntactic structure. Phonological constraints on templates formation activate adequate syntactic operations. If we assume that templatic domains lie at the interface between syntax and phonology, we account for some issues of affix ordering, that involve no syntactic property.

Introduction

In this paper, I suggest an approach of the interface between morphology and syntax in order to explain affixes ordering that are not sensitive to syntactical properties.

First, I present the verbal morphology in Qafar and the templatic representation of verbs. Then I will argue that phonological constraints interact with the syntactical derivation in order to account for the order of affixes.

Verbal classes in Qafar

The inflectional morphology of verbs in Qafar exhibits two verbal classes: on the one hand weak verbs that have only suffixes and strong verbs that have prefixes and suffixes (cf. Hayward 1978, Parker & Hayward 1985).

The position of inflectional markers around weak and strong verbs can be represented as follow in table 1:

Table 1. Positions of inflectional markers

		weak stem	person	aspect/mode	number
person	aspect	strong stem		mode	number

The examples of the weak verb *dunqe* 'push' and the strong verb *uktube* 'write' can illustrate these positions. The marker /t/ of second person is a prefix in *t-uktube* 'you wrote' and a suffix in *dunuq-t-e* 'you pushed'.

Imperfect is marked by the suffix /a/ in *dunqa* ‘you push’ versus *dunqe* ‘you pushed’. In strong verbs, this marker /a/ is at the left edge of the stem as in *a-ktube* ‘I write’ and *a-ɕfide* ‘I learn’. This /a/ alternates with the initial vowel that is always a copy of the vowel stem in perfect: *u-ktube* ‘I wrote’ and *i-ɕfide* ‘I learnt’. I assume that the suffix /e/ in the perfect and imperfect forms of strong verbs cannot appear in the imperfect forms of weak verbs because hiatus are prohibited in Qafar. If both markers /e/ and /a/ are in successive position, only one is phonetically realized (cf. Rucart 2000).

The suffix /e/ alternates with the suffix /o/ to express a modal distinction with purposive that is *dunq-o* ‘that I push’ and *aktub-o* ‘that I write’. Again, in weak verbs, the marker /o/ cannot be realized at the same place as the aspectual marker /a/.

The plural marker /n/ is the last suffix in both classes as in *dunqten* ‘you pushed-plural’ and *tuktuben* ‘you wrote-plural’.

Verbal template

The examination of strong verbs shows that they have trilateral roots. In case of the lack of one consonant, one can observe a vocalic lengthening. Facing the trilateral verb *uktube* ‘write’, there is verbs like *egeere* ‘bail’ or *eexge* ‘know’, with a long vowel in the stem. Additional consonant are always derivational onto express passive or causative for example.

These can be prefix consonant like in *u-s-kutube* ‘make s.o. write’, internal geminate like *usku-tt-ube* ‘be written’ or suffix like *uktubute* ‘write for oneself’.

Then I propose that strong verbs share a unique template with three root consonant positions and three derivational positions (in brackets below) within the Government Phonology - CVCV option theory (cf. Rucart 2002). The following template allows us to represent every strong verb stem:

(CV) CV (CV) CV CV (CV)

I assume that weak verbs share the same template. Most of them are trilateral and can have an internal germination or a consonant suffix like in *dunnuqime* ‘be pushed’. Moreover the passive of weak verbs with two root consonants (like *fake* ‘open’) have a final geminate in association with an unexpected following long vowel like in *fakkiime* ‘be opened’. If no underlying can be invoked to explain this vocalic lengthening, we can assume that it is an effect of the underlying template. This template with three root consonant positions has to be identified.

Initial CV and Proper Government

Lowenstamm (1999) propose that template of all major category have an initial CV at their left edge. Like every other nucleus, the V position of the ini-

come at the left edge of the template and can identify the initial CV. This position doesn't need to be properly governed any longer.

Conclusion

If we assume that templatic domains lie at the interface between syntax and phonology, we account for some issues of affix ordering, that involve no syntactic property.

References

- Hayward, R.J. 1978, The prefix conjugation in Afar, in *Atti del secondo congresso internazionale di linguistica Camito-Semitica*, Fonzaroli, P. (eds.), Firenze, Università di Firenze.
- Kaye, J., Lowenstamm, J., Vergnault, J.R. 1990, Constituent structure and Government Phonology, in *Phonology Yearbook 7*.
- Lowenstamm, J. 1996, CV as the only syllable type, in Durand, J. and Laks, B. (eds.) 1996, *Current trends in Phonology*, Salford, Manchester, E.S.R.
- Lowenstamm, J. 1999, The beginning of the word, in *Phonologica 1996*, Rennison & Kühnhammer (eds.), The Hague, Holland Academiv Graphics.
- Parker, E.M., Hayward, R.J. 1985, *An Afar-English-French dictionary*, London, S.O.A.S.
- Pollock, J.Y. 1989, Verb movement, universal grammar and structure of the IP, in *Linguistic Inquiry* 20:3.
- Rucart, P. 2000, *Aspect de la morphologie verbale de l'Afar*, D.E.A. dissertation, Paris, Université Paris VII
- Rucart, P. 2002, The vocalism of stong verbs in Afar, in *BLS 27S Afroasiatic Linguistics*, USA, Berkeley.
- Stump, G.T. 1998, Inflection, in *The handbook of morphology*, Spencer & Zwicky (eds.), London, Blackwell.

Processing causal and diagnostic uses of *so*

Sharmaine Seneviratne

Research Centre for English and Applied Linguistics, University of
Cambridge, UK

Abstract

This study compares the processing of causal and diagnostic uses of the connective *so*. In an experiment measuring reading times, it is shown that diagnostic *so* constructions, (e.g. *Carl has been limping lately. So he was injured in the accident.*) take longer to read than causal ones (e.g. *Carl didn't wear his seatbelt. So he was injured in the accident.*). This difference is reduced in parallel constructions where *so* is absent. It is proposed that readers are strongly biased towards a default causal reading of *so* and therefore anticipate a consequence when they encounter *so*. In diagnostic constructions, since *so* is followed by an antecedent, processing is more effortful and complex because the default expectation has to be reversed.

Introduction

Connectives play an important role in discourse production and comprehension as they are used to signal how discourse segments are linked with their context. It is generally agreed that the connectives *so* and *because* aid utterance interpretation by cueing that a causal link is intended between the discourse segments that they link. However, the types of processes that are triggered by these linguistic cues are conceptualised differently in various theoretical frameworks. In this paper we consider two contrasting theoretical accounts that describe the function of *so*. We compare the processing of two types of statements, causal and diagnostic, and show that these accounts do not adequately explain the processing difference observed between these two uses. An alternative account, which posits a default setting for *so*, is proposed. Causal and diagnostic uses of *so* are illustrated in (1) and (2) below:

- (1) Carl didn't wear his seatbelt. So he was injured in the accident.
- (2) Carl has been limping lately. So he was injured in the accident.

In both these examples the discourse segments are causally linked. However, in (1), which is a basic causal use of *so*, the event mentioned first is given as the cause for the second event, which follows the connective *so*. In (2), Carl's limping is not the cause for Carl being injured. The relation in (2) is better understood as a diagnostic one, in which the first segment provides evidence for the deduction of the conclusion stated in the second segment.

Theoretical accounts of *so*

The meaning of *so* and the way in which it functions in integrating events have been described within the framework of Relevance theory (Blakemore 1988) and in Coherence theory (Sanders et al. 1992). According to the Relevance account, connective expressions have procedural meaning: they encode instructions on how propositional representations are manipulated in arriving at the intended meaning. The connective *so*, specifically, instructs the hearer to process the proposition it introduces as an implication or conclusion. In this account, hearers follow the same inferential route in understanding both causal and diagnostic utterances. Therefore it would not predict that interpreting one type of statement would involve more processing complexity and effort than the other. Furthermore, we are not told how and whether it is important to identify the cause/evidence and consequence/conclusion elements in interpreting these statements.

The Coherence theorists propose that coherence relations such as the ones cued by *so* can be decomposed to combinations of cognitive primitives which each allow binary options: basic operation (causal or additive), source of coherence (semantic or pragmatic), order (basic or non basic) and polarity (positive or negative). In the causal use in (1), according to the criteria listed, *so* signals the relation *cause-consequence*, which is decomposed as causal-semantic-basic-positive. The diagnostic use in (2), however, signals the relation *consequence-cause*, which has the configuration causal-semantic-non basic-positive. Since *so* can be used to cue both combinations it is unclear how hearers initially determine the option for the order element in order to arrive at the correct coherence relation via the combination of cognitively salient primitives.

Both these accounts do not explain how, given a single linguistic cue *so*, hearers recognise that the inference to be made is one that leads from evidence to a conclusion or one that links a cause with a consequence. We hold that it is crucial for hearers to make this distinction in arriving at the intended interpretation. We suggest that, in general, *so* is strongly associated with a basic causal reading and that hearers normally assume, upon hearing *so*, that a consequence will follow. A reading time experiment was designed in order to test this assumption. Since nothing in the constructions unambiguously signal which interpretation should be adopted, it was predicted that in reading diagnostic statements, this default assumption would have to be withdrawn, thus causing a delay in the reading time.

Experiment

A self-paced, clause-by-clause reading experiment was used to measure reading times for identical second clauses in parallel causal and diagnostic

constructions (see Table 1). The two clauses were linked by the connective *so*, but they also appeared without the explicit marking of the connective.

Table 1: Sample items for experiment

Item Type	Sample Item
Marked Causal	Carl didn't wear his seatbelt. So he was injured in the accident.
Marked Diag.	Carl has been limping lately. So he was injured in the accident.
Unmarked Causal	Carl didn't wear his seatbelt. He was injured in the accident.
Unmarked Diag.	Carl has been limping lately. He was injured in the accident.

Forty native speakers of (British) English, undergraduate and graduate students at the University of Cambridge, took part in this experiment. The experimental items were 24 pairs of sentences similar to those in Table 1. Thirty-six filler items - pairs of sentences connected by *but* and *after all* and those that were not explicitly linked by a connective - were used. Twenty comprehension questions requiring yes/no responses were also included in order to ensure that readers continuously attempted to fully understand the sentences. There were two within-subjects variables: type of construction (causal or diagnostic), and presence or absence of connective. The items were rotated among the four conditions, so that each item was seen in a causal marked and unmarked condition and in a diagnostic marked and unmarked condition. Each participant saw a particular item under only one of the four conditions.

A self-paced reading task was run on a computer using Superlab Pro software. Participants clicked on the mouse button to initiate a trial, and viewed the first sentence of a pair on the screen. Once they read and understood the sentence, they clicked on the mouse to view the second sentence and again, to initiate the next trial. Time between each successive mouse click was recorded, and served as the dependent measure. Comprehension questions sometimes appeared after a pair of sentences, to which readers responded by clicking the left or right mouse button to answer either 'yes' or 'no'. These questions appeared at random intervals among the test items.

Results and discussion

A 2×2 ANOVA was conducted to test the effect of item type and presence of marker. There was a strong interaction of item type and the presence of *so* in the subjects analysis $F(1,40)=6.77$, $p<0.01$. As shown in Figure 1, reading times for marked diagnostic sentences was on average 366 milliseconds more than for the marked causal ones, $T_1(40)=0.74$, $p<0.0001$; $T_2(24)=4.26$, $p<0.01$. In the unmarked conditions the difference between reading times for the causal and diagnostic sentences was not significant $T_1(40)=0.60$, $p>0.1$; $T_2(24)=0.59$, $p>0.1$. In diagnostic sentences, explicit marking of *so* brought about slower reading times, $T_1(40)=0.51$, $p<0.0001$; $T_2(24)=3.35$, $p<0.002$.

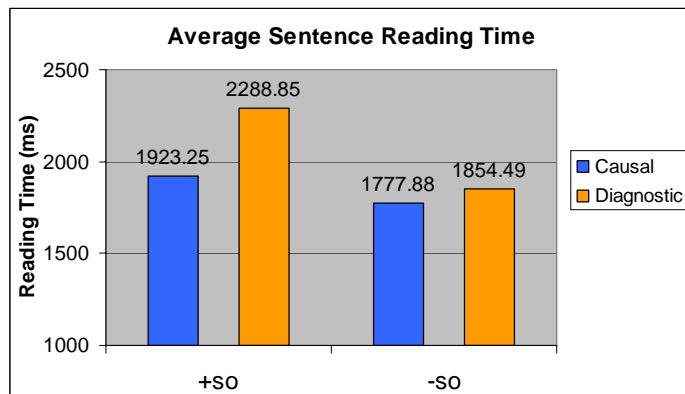


Figure 1: Average reading times for target clauses

The reading time differences show that causal and diagnostic uses of *so* have differential levels of processing difficulty. Furthermore, because diagnostic statements were easier to read when they were unmarked, it can be observed that the difficulty in processing diagnostic statements was at least partly due to an effect of *so*. This supports our hypothesis that the preferred reading of *so* is a basic causal one, in which *so* is expected to precede a consequence.

Conclusions and Future work

This study shows that diagnostic and causal uses of *so* involve different levels of processing effort. While *so* may signal a single basic procedure or relation this information is not sufficient to interpret causally linked sequences. It appears that hearers attempt to identify cause and consequence elements which would then prompt different sets of generalisations and inferences in recovering causal and diagnostic interpretations. In work on first language acquisition that is currently in progress, we explore further evidence for a default causal reading of *so* and whether diagnostic reasoning follows a different and more complex inferential path.

References

- Blakemore, D. 1988. 'So' as a constraint on relevance. In Kempson R. (ed.) 1988, *Mental Representations: The interface between language and reality*, 183-95. Cambridge, Cambridge University Press.
- Sanders, T., Spooren, W. and Noordman, L. 1992. Towards a taxonomy of discourse relations. *Discourse Processes* 15, 1-35.

Acoustics of speech and environmental Sounds

Susana M. Capitão Silva¹, Luis M. T. Jesus² and Mário A. L. Alves³

¹Secção Autónoma de Ciências da Saúde, Universidade de Aveiro and Direcção Regional da Educação do Norte, Ministério da Educação, Portugal

²Escola Superior de Saúde da Universidade de Aveiro and Instituto de Engenharia Electrónica e Telemática de Aveiro, Universidade de Aveiro, Portugal

³Hospital Senhora da Oliveira, Guimarães, Portugal

Abstract

In this study we present a preliminary acoustic analysis of environmental sound stimuli, based on Gaver's (1993) classification. Results showed similarities between sounds produced by objects with the same type of material and interaction. We also designed an experiment, where the subjects had to propose onomatopoeic representations for environmental sounds. The onomatopoeic representations used by the subjects shared common speech features (manner and place of articulation, and vowels used).

Introduction

In our everyday life we hear different kinds of sounds. The most common and to which we are exposed since birth are the environmental sounds. It is difficult to describe an environmental sound, but when we do, onomatopoeic representations are often used, and thereby speech sounds are used to report environmental sounds.

Environmental sounds acoustic features have been studied by various authors (Gygi, Kidd, and Watson 2004). According to Gaver (1993), musical listening and everyday listening are quite different processes: when listening to music we analyse features of the sound itself; when we try to perceive environmental sounds we listen to events and not sounds, we pay attention to what is generating the sound and not the emotional sensations or acoustic features conveyed by the sound. Therefore, research methods should consider the materials involved (solids, liquids or gases), and the type of interaction (solids: impacts, scrapping, rolling, deformation; liquids: drip, pour, splash, ripple; gases: explosion, gusts, wind).

Onomatopoeic representations of environmental sounds have been recently analysed by Takada, Tanaka and Iwamiya (2006), for a set of commercially available stimuli. The authors asked their subjects to describe the sounds according to quality rating scales and using onomatopoeic represen-

tations. Results showed similar acoustic properties in the stimuli expressed by onomatopoeic representations classified into the same clusters.

Environmental Sounds Stimuli

We selected 26 stimuli from audio recordings commonly used in clinical practice by Portuguese speech and language therapists. The sample contained a total of 42 environmental sounds that included sounds of animal vocalizations, sounds produced by the human body, sounds of nature, sounds of objects and sounds of means transportation. The selected stimuli were classified according to the type of interaction occurring between materials (Gaver 1993), which can be solid, liquid or gases. Most of the sounds consist of a single event, i.e., involved only one type of interaction between materials, but there were some stimuli that involve two types of interaction. There were 14 stimuli generated by solids, 6 by liquids and 6 by gases.

Method

Time waveforms and spectrograms of all stimuli were analysed using *Praat*. The following acoustic measures were extracted: duration, and F0, F1, F2 and F3 for periodic signals. Most of the sounds were noise signals, so multi-taper spectra (PSD Thomson estimates) were calculated with 11 ms windows left aligned to the start of the samples using *Matlab*, and peaks and broad peaks in the spectra were analysed.

Ten subjects (5 male and 5 female) with normal hearing abilities participated in this study. Subjects were asked to describe the selected stimuli using an onomatopoeic representation. Participants listened to the stimuli through Senheiser eH 1430 headphones. The stimuli were presented in the randomized order, and the subjects were able to listen to the sound stimuli for as many times as they felt necessary. Onomatopoeic representations produced by the subjects were recorded using a Philips SBC ME 400 unidirectional condenser microphone connected to a PC through a Sound Blaster Live! 24 bit soundcard.

The onomatopoeic representations were coded using 15 phonetic parameters: 5 places of articulation (bilabial, labio-dental, alveolar, palatal and velar), 6 manners of articulation (plosive, fricative, liquid), voiced, voiceless, nasal, and 4 vowel features (Group 1 - /ə, a, ε, e/; Group 2 - /i/; Group 3 - /u, o, ɔ/). These groups were based on the results presented by Takada, Tanaka and Iwamiya (2006), i.e., they were the most relevant features characterizing onomatopoeic representations.

Results

Temporal and spectral analysis of the selected sounds revealed some acoustic characteristics that allowed us to establish relations between speech and environmental sounds. The acoustic effects of source attributes proposed by Gaver (1993) were used as a reference, and new frequency and temporal domain characteristics of sounds generated by solids, liquids and gases with specific properties, were defined.

Duration was analysed according to the type of material producing the acoustic event. Results revealed that sounds produced by solids are shorter (specially the ones caused by a deformation) than sounds produced by liquids and gases (continuous sounds). The acoustic properties of sounds generated by an impact between solids are related to the acoustic properties of stops. In both cases we have an interaction between two solids that involves some sort of deformation. Aerodynamic sounds can be generated by an explosion (similar to the stops' plosion) or by more continuous sources (e.g. wind) which are similar to those used to produce fricatives. Liquid sounds, like those generated by water dripping, are produced by resonance cavities, with slight variations of pitch. When the interaction is a splash the sound is continuous like in fricatives, and when it is generated by waves or by water that is poured, those sounds share characteristics with laterals.

Sounds resulting from impacts of solids presented an acoustic aperiodic signal with high average amplitude which decayed over time. Most of the frequency components of the noise signal produced by a river were located above 1 kHz and by the wind below 1 kHz, as shown in Figure 1.

We observed that most onomatopoeic representations of solids used unvoiced alveolar stops and vowels from Group 3, as shown in Figure 2. An example of an onomatopoeic representation would be [tək tək].

Liquids were also predominantly represented by an alveolar place of articulation, but palatal consonants were also used. Voiceless fricatives are often used, indicating the absence of a periodic source.

Sounds produced by gases interactions were represented by alveolar and palatal places of articulation (mostly voiceless fricatives). Nasal consonants were used more often than for any other group. Vowels from Group 3 were also used. One example of onomatopoeia would be /fu:/.

Conclusions

This paper proposes novel ways of understanding how speech perception relates to environmental sounds perception, and presents a preliminary acoustic description of different categories of environmental sounds. Results showed that sounds produced by basic level events generated by similar materials or interactions shared common acoustic properties.

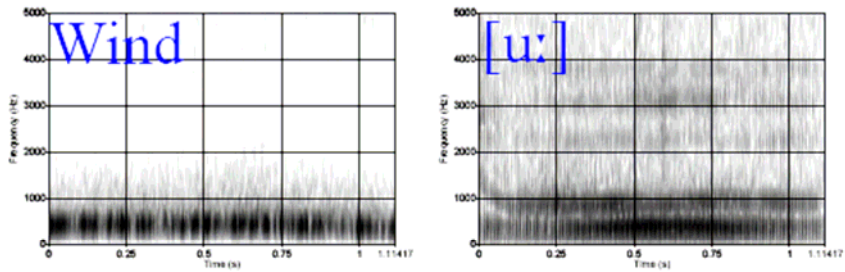


Figure 1. Spectrogram of stimulus “wind” and its onomatopoeia [u:].

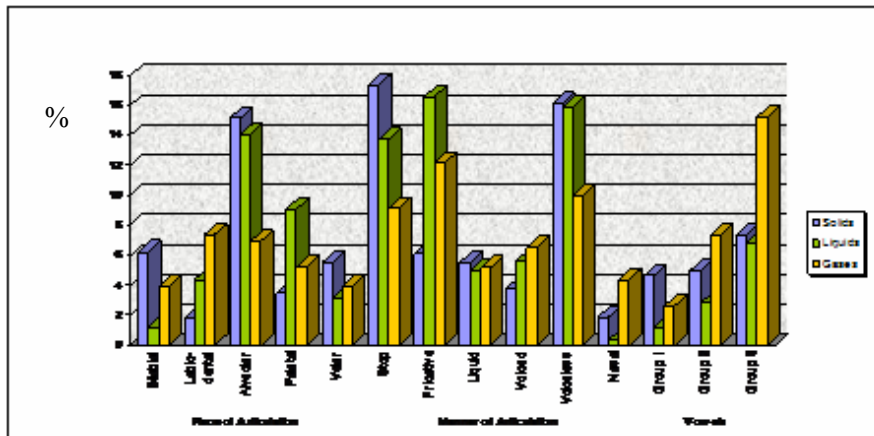


Figure 2. Phonetic features used to classify onomatopoeic representations of the different types of material: solids, liquids and gases.

References

- Gaver, W. 1993. What in the world do we hear?: an ecological approach to auditory event perception. *Ecological Psychology* 5(1), 1-29.
- Gygi, B., G. Kidd, C. Watson 2004. Spectral-temporal factors in the identification of environmental sounds. *Journal of the Acoustic Society of America*, 115 (3), 1252-1265.
- Takada, M., K. Tanaka, S. Iwamiya 2006. Relationships between auditory impressions and onomatopoeic features for environmental sounds. *Acoustic Science Technology* 27(2), 67-79.

Romanian palatalized consonants¹: A perceptual study

Laura Spinu

Department of Linguistics, University of Delaware, USA

Abstract

Most consonants in the Romanian inventory are claimed to have a surface palatalized counterpart carrying functional meaning; it is not clear, however, whether the set of palatalized consonants is represented underlyingly. The present study investigated the perception of palatalized consonants based on two experiments. It was found that native speakers can discriminate between plain and palatalized segments, but sensitivity to palatalization decreases when another morphological cue is present. It was further confirmed that the primary place of articulation affects perceptibility. This situation is intriguing from a representational perspective, since neither phonemic nor allophonic status in the sense of a classical analysis can fully describe the behavior of these consonants.

Introduction

Restricted to word-final position, the occurrence of palatalized consonants in Romanian is generally associated with the presence of two homonymous suffixes: the plural for nouns and adjectives (1a), and the 2nd person singular in the present indicative of verbs (1b); hence, palatalized consonants are considered morphologically predictable².

- (1) a. [lup^ɨ] ‘wolves’
b. [sar^ɨ] ‘you jump’

Contrasts in meaning are thought to arise between the inflected form, ending in a palatalized consonant, and the uninflected one³, which ends in a plain consonant, thus resulting in minimal pairs (1a-2a, 1b-2b).

- (2) a. [lup] ‘wolf’
b. [sar] ‘I jump’

Previous accounts. Two different phonetic descriptions prevail regarding these surface realizations: (a) palatalized consonants, in the sense of a secondary palatal feature overlapping with the primary place of articulation, as can be found in Slavic languages; (b) sequences of consonant-glide. More disagreement is found with respect to their underlying representation: palatalized consonants are considered part of the phonemic inventory by Petrovici (1956), while Schane (1971) does not take them to be underlying, but grants the native speaker a special level of awareness at which surface contrasts can be phonemic. Other linguists take palatalization or the final glide to be underlyingly represented as a glide or semi-vowel (Agard 1958, Avram 1991),

or an archiphoneme, sharing features of both /j/ and /ɟ/ (Vasiliu 1990). Finally, there is the view according to which surface palatalized consonants or word-final, post-consonantal glides correspond to an underlying /i/ (Stan 1973, Ruhlen 1973, Steriade 1984).

A Perceptual study

With so many differing views on the theoretical level, the present study approaches the phenomenon of palatalization in a different way. Emphasis is laid on the facts, in the belief that a thorough analysis of the measurable phenomena may serve to settle the disagreements on the phonetic/phonological status of these structures. The current study is aimed at determining whether native speakers can reliably distinguish between plain and palatalized consonants. A secondary goal is to find out whether the primary place of articulation (POA) of palatalized consonants plays any role in perception, as previously demonstrated by Kochetov (2002).

Experiment 1

This experiment tested the sensitivity of native speakers to palatalized segments. Two hypotheses were formulated:

Hypothesis 1. Native speakers are sensitive to (i.e. can reliably identify) both plain and palatalized consonants.

Hypothesis 2. Differences in perceptibility are expected for different POAs.

The stimuli consisted of 36 target words ending in either a plain or a palatalized consonant from one of three different POAs: labial, dental, and post-alveolar. The targets were placed inside a carrier sentence of neutral meaning (“I will choose the word ___ tomorrow.”), and a native speaker of Romanian recorded these sentences and additional fillers.

12 subjects listened to these sentences. They were asked to write down the words they heard before ‘tomorrow’. This task was chosen because Romanian orthography marks the plural of nouns and adjectives, as well as the 2nd person of verbs with a word-final -i.

Results. The correct identification rate for plain consonants was 89.8%, and for palatalized consonants 65.7%. While a noticeable decrease could be noted for the palatalized group, the identification rate was still significantly above chance; Hypothesis 1 was thus confirmed.

Figure 1 shows the listeners’ perceptual sensitivity, known as d' (d' prime), to each POA. The d' value obtained for post-alveolar segments is significantly lower than the others, in support of Hypothesis 2.

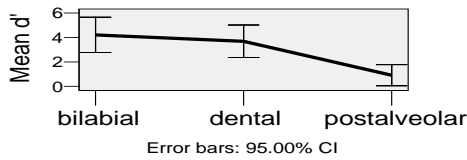


Figure 1. Perceptual sensitivity (d') to palatalization for each POA.

Experiment 2

In Experiment 1, palatalization was the only cue to the morphological status (sg./pl.) of the targets. Given that normally there is rich agreement in Romanian, the question arises whether the presence of additional morphological information makes listeners rely less on palatalization. Experiment 2 investigated the perceptibility of palatalization in the presence of other cues.

Hypothesis 3. In the presence of other morphological cues, subjects will pay less attention to palatalization (as compared to Experiment 1).

Hypothesis 4. Even in this case, POA effects are expected.

144 targets were selected from the three POAs previously tested. They were inserted in carrier sentences containing only one other cue to their morphological status (singular or plural). Each sentence was in one of two conditions: the matched context, in which the target word and the cue to its status were consistent (e.g. *un lup* 'one wolf'), and the mismatched context, in which the target word and the cue were conflicting (e.g. **doi lup* 'two wolf'). The task consisted of a forced choice between "acceptable" and "not acceptable". There were 20 subjects for this experiment.

Results. Figure 2 shows the perceptual sensitivity to palatalization for Experiment 1 as compared to Experiment 2. A tendency for perception to decrease in the presence of another morphological cue can be noticed, as predicted by Hypothesis 3.

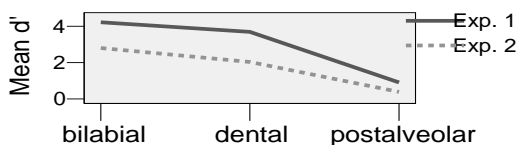


Figure 2. A comparison between the two experiments (d').

As for POA effects, decreased sensitivity was found once more with post-alveolars. Hypothesis 4 was confirmed by a one-way analysis of vari-

ance revealing that the d' value for post-alveolars was significantly lower than for the other two places ($F(2,57) = 18.95, p < .001$).

Conclusion

Native speakers of Romanian were able to discriminate between plain and palatalized consonants, but a decrease in perceptibility was found in the presence of another morphological cue, as well as depending on the primary POA of the palatalized consonant. The ability of these segments to contrast meaning is compatible with the assumption that palatalization corresponds to an underlying /i/. As for the perceptual differences noted for the three places of articulation, recent work explains phenomena such as the one under discussion by allowing more freedom to the surface phonemic representation, instead of positing many different rules to account for phonetic variation (Ladd 2006). This view is in line with modern findings emphasizing the quantitative nature of phonetic manifestations.

Notes

- 1 To my knowledge, no acoustic/articulatory study has established if these structures are palatalized consonants (with both a primary and a secondary place of articulation) or consonant-glide sequences. The term 'palatalization' was chosen because it has been more recurrent with respect to this topic in recent years.
- 2 Closer examination reveals the existence of monomorphemic words exhibiting the same pattern and, with few exceptions, a general lack of word-final C-unstressed [i] sequences in the language. It may be the case then that palatalized consonants are also phonologically predictable in word-final position.
- 3 Steriade (1984) shows that masculine/neuter nouns end in an underlying theme vowel /u/, and the 1st person singular affix for the present indicative of verbs (in some conjugations) is /u/ as well, even though these vowels rarely surface due to a rule of high vowel desyllabification.

References

- Avram, A. 1991. Semivocalele, semiconsoanele și pseudovocalele în română. *Studii și cercetări lingvistice* 5/6, 213-229.
- Chitoran, I. 2001. *The Phonology of Romanian: A Constraint-based Approach*. New York, Mouton de Gruyter.
- Kochetov, A. 2002. *Production, Perception, and Emergent Phonotactic Patterns: A Case of Contrastive Palatalization*. New York, Routledge.
- Ladd, D. R. 2002. "Distinctive phones" in surface representation. In Goldstein, L., Whalen, D. H. and Best, C. T. (eds.) *Papers in Laboratory Phonology 8: Varieties of Phonological Competence (2006)*, 100-120. Mouton de Gruyter.
- Ruhlen, M. 1973. *Rumanian Phonology*. PhD dissertation. Stanford University.
- Schane, S. 1971. The phoneme revisited. *Language* 47, 503-521.
- Steriade, D. 1984. Glides and vowels in Romanian. In *Proceedings of the tenth annual meeting of the Berkeley Linguistics Society*, 47-64.

Formal expressive indiscernibility underlying a prosodic deformation model

Ioana Suciu¹, Ioannis Kanellos² and Thierry Moudenc¹

¹TECH/SSTP/VMI, France Telecom R&D, Lannion, France

²Computer Science Department, ENST Bretagne, Brest, France

Abstract

We are here concerned by the setting up of a model and a formalism for expressive speech synthesis under the paradigm of a corpus-based approach. Our objective is to apply prosodic expressive forms, acquired from natural human-reading recordings, on a new textual matter. We outline a general model for speech expressiveness. Then we deal with some formal aspects of expressive representation. We point out the core transformational aspects and the indiscernibility criteria allowing comparisons between forms. We finish by some interpretational issues of such an approach.

Introduction

In speech synthesis one cannot for long avoid the prosodic deformation model requirement, especially when expressiveness becomes a central research theme for natural speech (Keller 2002). The arguments are important and massive (epistemological, theoretical, applicative...) and increase their intelligibility once one opts for a corpus-based synthesis approach. Indeed, acquired corpora, even extended, cannot give high quality acceptability results for all expressive forms one encounters in everyday life linguistic uses. Thus, the only issue is the generative-transformational one: on the basis of a kernel of forms one has to generate a relevant range of the expressive universe one typically encounters in a practice. This last defines a research program in which the model architecture and the choice of the formal representation condition an important part of both the expected results and their validation protocols. The model gives the framework in which the conception of the speech expressivity is thoroughly thought out, while the formality gives the scope of its transformational potentiality likely to be implemented. Such remarks give evidence to the structure we subsequently follow for our talk.

A model for speech synthesis expressivity

Without any exigency for exhaustiveness, we outline here a general model for the expressive synthesised speech, as shown in (Kanellos & *al.* 2006). It is set up through five consecutive steps:

Firstly, one has to start by positioning a discursive form in a multidimensional space defined by three global characteristics, extrinsic to the textual matter: the textual genre (*tg*), the discursive situation (*ds*) and the reader's profile (*rp*). Implicit for a human speaker and essential for the speech production, perception and interpretation, they are required as entries for an expressive speech handling.

Once situated, the text to synthesize is analysed through complementary points of view giving rise to lexical, morpho-syntactical, typographical, semantic, punctuation etc. treatments.

Then, the levels of the textual analysis are chosen. For complexity efficiency and adequacy to current techniques, three of them seem sufficient: the syllable (*syl*), the syntagm (*syn*) and the phrase group (*phg*) level. The expressive transposition may concern the deformation of a unit of any of these levels.

The next step deals with prosodic representation that concerns precisely the units of these three levels. It allows the description of the melodic (*F*), temporal (*T*) and intensity (*I*) movements (local and global), to be associated with them.

Finally, an expressive discursive form is defined as a choice strategy among the values of these three prosodic parameters.

Formal representation of expressiveness

Let us now see a possible formal description for the expressive phenomena.

One naturally starts by representing the space vector of the extrinsic characteristics: $S =_{df} \langle tg, ds, rp \rangle$. Therefore, any expressive form type *E* has to be envisaged as a situated complex vector in such a space: $E_{sit} =_{df} \langle E, S \rangle =_{df} \langle E, \langle tg, ds, rp \rangle \rangle$ (cf. step 1 above).

For a correct synthesis realisation of a textual unit *U*, one generally needs to know its phonological level of analysis *l* (i.e. *syl*, *syn* or *phg*; cf. step 3), its compositional structure *C*, as well as some linguistic aspects *D* describing it. We represent these by the linguistic unit description vector: $U =_{df} \langle l, C, D \rangle$. *C* informs about the number *n* of the units of *l*-1 level composing *U* (number of syllables for a syntagm, number of syntagms for a phrase group) and the (ordered) list of their identifiers: $C =_{df} \langle n, id_1 \dots id_n \rangle$; for reliability reasons, these identifiers are supposed to be unique. Finally, *D* gives information on the syntagm nature (noun, verb etc.), the focus localisation(s), some punctuation or typographical specifications etc. (cf. step 2).

Local or global, each expressive unit carries pieces of information corresponding to the prosodic movements in speech (cf. step 4). We represent them by a three-dimensional prosodic vector: $P =_{df} \langle F, T, I \rangle$. *F*, *T*, *I* are typed data; their type is decided by the *l* level (for instance T_{syl} , T_{syn} and T_{phg}).

We define an expressive form type as a triplet: $E =_{df} \langle id, U, P \rangle$, where, *id* is its unique identifier. Thus, $E =_{df} \langle id, \langle l, C, D \rangle, \langle F, T, I \rangle \rangle$ (1) and clearly: $E_{sit} =_{df} \langle \langle id, \langle l, C, D \rangle, \langle F, T, I \rangle \rangle, \langle tg, ds, rp \rangle \rangle$ (2).

Indiscernibility and transformation scenarios

The expression (2) encapsulates the expressive transformational potentiality allowed by the model above (Zaldivar-Carrillo 1995). Indeed, any possible transformation modifies some (or all) of the components of the E_{sit} vector at different ranks of occurrence: l , C , D and F , T , I as well as tg , ds , rp at the first rank, U and P at the second, E and S at the third one. These transformations constitute an inescapable formal ingredient of two typical applicative orientations: the first corresponds to productive objectives and concerns the generation of expressive speech on the basis of existing forms; the second is interested in acquisition and learning purposes and deals mainly with the comparison of a new form with an existing one in the expressive data base.

Two types of formal operators are defined with respect to these applications: O (unary) and R (binary). The O-type operators generate new forms (of the same level l) following a given transformation scenario, such as compression, stretching, scaling, transgression, reversion, inversion, translation etc. These scenarios are different for different E_{sit} components (for instance, the stretching in time is not the same as the stretching in melody). An R-type operator searches to compare two expressive forms under some indiscernibility conditions, informing about their eventual complete or partial equivalence. Here again, compared forms must have the same level l . Partial equivalence renders the expressive indiscernibility over one or more of the U , P or S dimensions (thus $e_1 =_{F,T} e_2$ means that the two forms are indiscernible over both the melodic and temporal aspects; $e_1 =_C e_2$ that they are indiscernible over the C component etc.). These indiscernibility criteria are on the basis of any prosodic transformation procedure. They also found similarity and tolerance relationships (Ferret 1998) between expressive forms.

Clearly, one can easily envisage compositions of O-type operators. On the other hand, the R-type operators act as condition for the O transformations (they qualify the relationship between the original and the generated form by an O operator). Moreover, if e_1 and e_2 are expressive forms of the same level, it is possible to find at least one O-type transformational sequence O_1, \dots, O_k such that $O_k \circ \dots \circ O_1 (e_1) = e_2$. In other words, once the level of analysis l is fixed, the universe \mathcal{E}_l of the expressive forms becomes complete under the O-type transformations. The corollary of this is that starting from any form e_1 , it is possible to generate an e_2 such that $R(e_1, e_2)$ satisfies a given indiscernibility criterion (e.g. $e_1 =_{F,T,I} e_2$). In both cases, it is possible to set up different criteria determining transformational costs.

Interpretational cues for indiscernible forms

We illustrate here some of the ideas above (*cf.* also (Suciu & *al.* 2006)).

The prosodic particularities of two speakers (Max vs. Tom) are studied in function of *F*, *T* and *I* variations, but supposing that the implied transformations preserve indiscernibility over *C*, *D*, *tg* and *ds* dimensions (=C, D, *tg*, *ds*). Detecting recurrences in these particularities may furthermore be a clue for a stylistic *rp* profile description and give rise to profile expressive universes.

Different reading manners (drunken vs. hysterical) are obtained by variations in *T*, *F* and *I*, while *C*, *D*, *tg* and *rp* remain indiscernible (=C, D, *tg*, *rp*).

An altered effect for the listener, going from strangeness to parody (a love letter read as a political discourse) may be produced by the same prosodic realisations for two different textual genres *tg*. Generally, it necessitates indiscernibility over *U*, *P*, *ds* and *rp* (=C, D, *F*, *T*, *I*, *ds*, *rp*).

On the other hand, for the same *ds* and *tg* values, the *U* and *P* indiscernibility scenario becomes a speaker imitation one. It is the case where the *rp*₁ speaker appropriates the prosodic profile of the *rp*₂ (Max speaking as some famous president, for example).

Conclusions

As semantic extension of a written text, expressiveness formulates an indelible problem insofar as it calls for extrinsic supplementary pragmatic information and concerns mainly the reception which strongly depends on the interpretation strategies of the listener. It seems however legitimate and even highly promising to extrapolate the corpus-based paradigm for expressive forms. In such a case, the set of transformations and the indiscernibility relationships give the essentials of the approach that can be implemented as an extension on a traditional corpus-based speech synthesis system.

References

- Ferret, S., 1998. *L'identité*, Flammarion, Paris.
- Keller, E. (ed), 2002. *Improvements in Speech Synthesis. COST 258: The Naturalness of Synthetic Speech*, John Wiley & Sons Ltd., England.
- Kanellos, I., Suciú I., Moudenc, Th., 2006. *Émotions et genres de locution. La reconstitution du pathos en synthèse vocale*. In Rinn, M. (ed.) *Le Pathos en action*, Presses Universitaires de Rennes, France (to appear).
- Suciú I., Kanellos, I., Moudenc, Th., 2006. *What about the text? Modelling global expressiveness in speech synthesis*. Proceedings of the ICTTA Conference, Damascus, Syria, 177-178 (extended version in the DVD of the Proceedings).
- Zaldivar-Carrillo, V.-H., 1995. *Contributions à la formalisation de la notion de contexte. Le concept de « théorie » dans la représentation des connaissances*. Ph.D, University of Montpellier 2, France.

What is said and what is implicated: A study with reference to communication in English and Russian

Anna Sysoeva
Department of Linguistics, University of Cambridge, UK

Abstract

The study of cross-cultural differences in the degree of reliance on different types of inferences shows that pragmatic inference contributing to the additional implicated proposition is the only kind of inference that can be preferred or dispreferred for cultural reasons. Defaults and pragmatic inferences contributing to the truth-conditional representation, on the other hand, are not a matter of preference/dispreference. This observation signifies that different types of inferences differ in their functions and processing.

Introduction

Since the difference between *the said* and *the implicated* was revealed, there has been a lot of debate as to what categories should be distinguished in the domain between what is said and what is implicated (Carston 2002, Grice 1989, Levinson 2000, Recanati 2004). I suggest that studying cross-cultural differences and similarities in the degree of reliance on the sources of information leading to the expression of the overall meaning can shed light on (i) the functions different types of inferences have, (ii) the degree of universality in the reliance on different types of inferences and (iii) the necessity of distinguishing between different types of inferences.

Theoretical framework and terminological remarks

Russian and English languages allow performing most types of speech acts through both what is said and implicature. The difference lies in the degree to which people of different cultures are willing to communicate by inviting the hearer to infer certain information in order to get to the overall meaning of the utterance. I will use the notion of *preference* to denote this willingness.

Cross-cultural differences in the degree of reliance on the sources of information contributing to the meaning of an act of communication are studied in the framework of Default Semantics (Jaszczolt 2005). Jaszczolt (2005) identifies the following sources: combination of word meaning and sentence

structure; defaults; conscious pragmatic inference₁ (CPI1) and conscious pragmatic inference₂ (CPI2). The first three sources contribute to the truth-conditional representation of an act of communication. CPI2 lies outside this representation. My aim is to investigate which of the sources are preferred with varying degrees in different cultures. The results are illustrated by a study of 300 randomly selected American, British and Russian online job advertisements.

Cross-cultural differences in the degree of reliance on different types of inferences

Conscious pragmatic inferences₂

In 84 out of 100 Russian online job advertisements no essential part of information is left to CPI2. By contrast, only 8 out of 100 British and 17 out of 100 American advertisements do not require CPI2 of the reader. This considerable difference in the preference for CPI2 is explained by a difference in cultural values. Russians report that implicating is a hindrance on truthful and sincere relations between people, which is one of the most important traditional values in Russian culture (Wierzbicka 1992, Zaliznjak et. al. 2005). In British and American cultures, on the other hand, implicating is perceived as a sign of politeness.

Because of the difference in cultural values, CPI2s differ in the social functions they perform in different cultures. In Russian CPI2s are used out of necessity rather than out of preference: to disguise illegal content, as in (1), or to avoid direct comparison with competitors working illegally, as in (2).

(1) Trebuetsja massazhist (možno bez opyta raboty – obučenie predostvljaetsja) – devuška, možno inogorodnjaja, prijatnoj vnešnosti, bez kompleksov, s razumnymi ambicijami, gotovaja zarabatyvat’.

“A masseur is needed (work experience not required – training provided) – a girl, may be Moscow non-resident, good-looking, without complexes, with reasonable ambitions, ready to earn.”

+> A prostitute is required.

(2) Oformlenie po trudovoj knižke, sobljudenie KZOT.

“Registering in the work book, following the labour code.”

+> Unlike many Russian companies, we work legally.

In British and American advertisements communicating meaning by inviting CPI2 contributes to persuasiveness by emphasising politeness and attention to the needs of a particular reader. For example, (3) is a milder way of communicating requirements than the bald on record form ‘The candidate must...’ that is preferred in Russian. In (4) personal reader addressing through the use of personal pronouns, imperatives, questions and elements of

spoken language implicitly communicates that the writer cares about the reader's individual interests.

(3) We would be happy to receive applications from candidates who can offer expertise and experience in at least four of the following areas...

+> Candidates must have expertise and experience in at least four of the following areas...

(4) Italy, Singapore, Australia, Hong Kong, Hawaii, Philippines – where would you most like to show off your creative talents?

+> We offer you a choice of location depending on your individual preferences.

Pragmatic inferences₁

In Default Semantics pragmatic inference contributing to truth conditions may be either developments of the logical form, as in (5), or completely different propositions. For example, the truth-conditional representation for (6a) may be (6c) rather than (6b)

(5) If you do not change [for the better] something in your life now [in the nearest future], you will always [during your lifetime] have things you already have.

(6a) You are not going to die.

(6b) You are not going to die from this wound.

(6c) You should not worry.

The degree of reliance on CPIs of both types does not seem to differ cross-culturally. The reason is that in case of CPI1 the speaker is not perceived as saying one thing and implying something different. Therefore, there is no dispreference for CPI1 in Russian culture.

Defaults

Though the content of defaults stemming from cultural stereotypes is different cross-culturally, there is no cross-cultural difference in the preference for communicating meaning by inviting default inference from the hearer. The reason seems to be that speakers are not aware of performing default inferences. The universal reliance on defaults is explained by the natural tendency of human beings to search for the most economical ways of expressing thought (Levinson 2000).

Conclusions

The study has shown that there is a different degree of universality in the reliance on different types of inferences in communication. Cultures are different in the preference for CPI2s and similar in the preference for defaults

and CPI1s. People choose to invite CPI2s not because it is more effective for cognitive reasons, but because it is more effective for social reasons. This is corroborated by the fact that CPI2s perform different social functions in cultures with different values. Defaults and CPI1s, on the other hand, are required by cognitive factors.

Difference in the psychological preference for inviting CPI1s, CPI2s and defaults from the hearer shows that there is a difference in the process of arriving at these types of inferences. Conscious pragmatic choice is present in case of CPI2. Defaults and CPI1s, on the other hand, seem to be arrived at by some facilitated inference of which the hearer is unaware. Experimental evidence is needed to be able to judge if it is a low-cost spontaneous conscious inference or an unconscious inference. Differences in the process of arriving at CPI1s and CPI2s show that it is justified to distinguish between these types of inferences from the point of view of processing as well as from the point of view of a theory. It should be recognised that functionally independent propositions may act as primary meanings.

All the components that contribute to the truth-conditional representation in Default Semantics are similar in the sense that a person cannot prefer or disprefer to make use of these sources. The conscious choice takes place between the merger proposition and the post-merger layer.

References

- Carston, R. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford, Blackwell.
- Grice, P. 1989. *Studies in the Way of Words*. Cambridge, Mass, Harvard University Press.
- Jaszczolt, K.M. 2005. *Default Semantics: Foundations of a Compositional Theory of Acts of Communication*. Oxford, OUP.
- Levinson, S.C. 2000. *Presumptive Meaning: The Theory of Generalized Conversational Implicature*. Cambridge, Mass, MIT Press.
- Recanati, F. 2004. *Literal Meaning*. Cambridge, CUP.
- Wierzbicka, A. 1992. *Semantics, Culture, and Cognition: Universal Human Concepts in Culture-specific Configurations*. Oxford, OUP.
- Zaliznjak, A.A., Levontina, I.B. and Shmelev, A.D. 2005. *Klyuvhevye idei russkojazykovoj kartiny mira*. Moscow, Yazyki slavyanskoy kul'tury.

Animacy Effects on Discourse Prominence in Greek Complex NPs

Stella Tsaklidou, Eleni Miltsakaki
School of English, Aristotle University of Thessaloniki, Greece

Abstract

This paper is concerned with the factors determining the relative salience of entities evoked in Complex NPs. The salience of entities evoked in complex NPs cannot be predicted by current theories of salience which attribute salience to grammatical role (subjects are more salient than non-subjects) or thematic role (agent are more salient than non-agents). A plausible hypothesis might be that, in complex NPs, head nouns are more salient than non-head nouns. Based on a sizable corpus of Greek, we analyze 484 instances of complex NPs. The results of the analysis reveal a semantic hierarchy of salience which predicts the full range of data independently of headedness: Animate Human>Inanimate Concrete Object>Inanimate Abstract Object.

Introduction

The perceived prominence of entities in discourse has been analyzed extensively in a large body of the linguistic, psycholinguistic and computational literature. Discourse prominence is often correlated with the interpretation of referential expressions, especially pronouns and other referentially underspecified forms. Extensive work in this area has identified several factors responsible for making some discourse entities more prominent than others. Notably, syntactic and semantic properties of entities have repeatedly been found to correlate with discourse salience. Many researchers have observed that grammatical role is important with subjects being more salient than non-subjects (e.g., Brennan et al 1987, Grosz et al 1995). Others have observed that thematic roles are important and argued that the semantics of verbs may be responsible for bringing to focus entities instantiating specific thematic roles (e.g., Stevenson et al 2000).

The work presented here is also concerned with discourse prominence. Specifically, this paper presents a corpus-based analysis of prominence in complex NPs in Greek. Complex NPs (henceforth CNPs) are especially interesting because they evoke more than one discourse entities whose salience cannot be predicted by grammatical or thematic role. To give an example, 'John's mother' is a CNP which evokes two entities, 'John' and 'John's mother'. CNPs may evoke multiple entities which do not participate in a possession relation, for example 'garden table', 'abortion law', etc. The whole NP can be subject or object and the head referent can be an agent or a patient but

the non-head is harder to characterize in terms of grammatical and thematic role. Our aim here is to test empirically which entity (the structural head or some other entity) is perceived as more salient. The rest of this paper is organized as follows: In Section 2, we give a brief overview of prior research in CNPs. In Section 3, we present the methodology, data, results and conclusions from our corpus study of Greek CNPs.

Related work on CNPs

Prior work in entity salience in complex NPs is very limited and in most cases not tested on empirical data. Specifically, Walker and Prince (1996) proposed, but did not test, the complex NP assumption which states that in English the Cf ranking within a Complex NP is from left to right. For example, in [*Heri mother*]j knows *Queen Elizabeth*, the salience order of the entities is i>j>k. Di Eugenio (1998) proposed (but again did not test) that in possessive NPs animate possessors rank higher. If both entities in the possessive NP are animate then the head noun ranks higher. Gordon and Hendrick (1997, 1998) through a series of experiments found that the collective entity evoked by the CNP is more accessible and prominent than its component entities, noting, that this happens when the CNP is in subject position. Note that in their data, consisting of possessive NPs in English, both entities were animate.

Corpus analysis of complex NPs in Greek

The dataset for this study was constructed from a corpus of approximately 182,000 words which we collected from the on-line publication of the Greek newspaper 'Eleftherotypia'. We restricted our dataset to only those occurrences of CNP constructions which were followed by a sentence containing a reference to at least one of the entities evoked in the CNPs. We did this under the assumption that when more than two entities are evoked in the discourse, subsequent reference to one of them indicates that the referenced entity is more salient. There were no cases with subsequent reference to more than one of the entities evoked in the CNP.

The final version of our dataset consists of 484 tokens of CNPs extracted according to the following criteria: a) each CNP evokes two or more entities, and b) one of the evoked entities is referenced in the following sentence. We excluded CNPs with coordinated nouns (see Gordon & Hendrick 1997 for a related study), cases of intra-sentential anaphora (e.g., reference to one of the entities in a relative clause construction) or reference in a parenthetical sentence. In what follows we report the results of the analysis of 402 complex NPs. These are NPs evoking two entities, the head noun followed by a genitive noun. Only 82 complex NPs evoked more than two entities and we

will not discuss them further. For each entity evoked in the CNP we coded the following semantic types: animate human, (there were no animate non-human), inanimate concrete objects, (e.g., table, book, etc.), and inanimate abstract objects (e.g., freedom, honesty, etc.). Table 1 shows the results of the coding for all attested combinations. The column ‘Ref. to H’ shows how many times there was reference to the head noun of the CNP and the column ‘Ref. to GN’ shows the number of times that the referenced entity of the CNP was the entity evoked in the genitive noun.

Table 1

Semantic labels	Ref. to H	Ref. to GN	Tokens total
AH-AH	27	17	44
AH-IC	32	10	42
AH-IA	8	0	8
IC-AH	13	36	49
IC-IA	12	1	13
IC-IC	18	17	35
IA-IA	20	32	52
IA-AH	15	99	114
IA-IC	5	40	45

Table 1 shows a strong preference for reference to the head of the CNP when the head is AH, shown in (1) for AH-IC. However, the headedness effect is lost when the head is IC and the genitive noun is AH, which suggests that AH is more salient than IC. The same pattern is observed when the head is IA and the genitive noun is AH. When the head is IA but the genitive is IC we see a strong preference for reference to IC, which suggests that IC is more salient than IA. But what happens when both the head and non-head entities are of the same semantic type? In the case of AH-AH, there is a headedness effect with heads being more frequently referenced than non-heads. There is no headedness effect, though, in the IC-IC case, shown in (2), and a rather puzzling preference for reference to the non-head IA in the IA-IA cases, shown in (3). Note, however, that in (3) the referring expression is with a full noun phrase.

(1) *oi megaliteri epihirimaties, tis horas_j itan sto eleos tis dikastikis tis krisis. Orismeni apo aftous_i ine ke simera desmii tis siopis tis.*

‘The most important businessmen_i in the country_j were in the mercy of her judgment. Some of them_i are still captive to her silence’

(2) *I plioktitria eteria ihe prosthesi ki allous orofous sto plio me apotelesma na epireasti I statherotita_i tou_j. O proedros tis Egiptou dietakse ti dieksagogi epigousas erevnas gia tin eksakrivosi ton etion vithisis tou pliou_j.*

'The shipowner aimed to add some more floors on the ship with result for its_j stability_i to be affected. The President of Egypt ordered the carrying out of urgent investigation for the identification of the causes of the sinking of the ship_j'.

(3) ...sintelestike ena eglima me tous heirismous tis kivernisis ke **tin prospathia_j sigkalipsis_i. I sigalipsi_i** ine sinenohi.

A crime was committed with the manipulations of the government and the attempt_j of covering-up_i. The covering-up_i is complicity.

The analysis of the data reveals a semantic hierarchy that strongly predicts subsequent reference: **AH>IC>IA**. The results of our corpus analysis are also supported by a Centering-based study that Poesio and Nissim (2001) designed to evaluate the relative salience of entities evoked in possessive NPs in English. They compared the complex NP assumption with the left-to-right ranking with Gordon and Hendrick's finding that the head of the complex NP is more salient and concluded that actually ranking animate entities higher than ranking heads higher yields fewer Centering violations. Further studies are required to evaluate if the animacy effect that has been empirically observed in Greek and English is specific to the entities evoked in complex NPs or a more general factor for salience ranking that has been missed by the most widely accepted accounts of discourse salience.

References

- Brennan, S. Walker Friedman, M. and Pollard, C. 1987. A centering approach to pronouns. In Proceedings, 25th Annual Meeting of the Association for Computational Linguistics, pages 155--162, Stanford.
- Di Eugenio, B. 1998 Centering in Italian. In "Centering in Discourse", Prince, E. Joshi, A. and Walkers, L. editors. Oxford University Press.
- Gordon, P. C. Hendrick R. Ledoux K. and Yang C.L. 1999. University of North Carolina at Chapel Hill, USA Processing of Reference and the Structure of Language: An Analysis of Complex Noun Phrases .
- Grosz, B.J. Joshi, A.K. and Weinstein, S. 1995. Centering: A framework for modeling the local coherence of discourse. Computational Linguistics, 21(2):202-225.
- Poesio, M. and Nissim, M. 2001. Architectures and Mechanisms for Language Processing Conference, Saarbrücken, September 2001
- Stevenson, R. Knott, A. Oberlander, J. and McDonald, S. 2000. Interpreting pronouns and connectives: Interactions among focusing, thematic roles and coherence relations. Language and Cognitive Processes 15(3):225-262.
- Walker, M. and Prince, E. 1996. "A bilateral approach to givenness: A hearer-status algorithm and a centering algorithm." In T. Fretheim and J. Gundel, editors, Reference and Referent Accessibility. John Benjamins, Amsterdam, pages 291-306.

Formality and informality in electronic communication

Edmund Turney, Carmen Pérez Sabater, Begoña Montero Fleta
Departamento de Lingüística Aplicada, Universidad Politécnica de Valencia,
Spain

Abstract

Electronic mails have nowadays become the most usual support to exchange information in professional and academic environments. A lot of research on this topic to date has focused on the linguistic characteristics of electronic communication and on the formal and informal features and the orality involved in this form of communication. Most of the studies have referred to group-based asynchronous communication. But the increasing use of e-mails today, even for the most important, confidential and formal purposes is tending to form a new sub-genre of letter-writing. This paper studies the formulae of etiquette and protocol used in e-mails for salutation, opening, pre-closing and closing, and other elements related to formality and provides new insights on these features. Our research is based on the analysis of a corpus of formal and informal messages in an academic environment.

Introduction

In this paper we compare different linguistic features of e-mails in English on the basis of their mode of communication (one-to-one or one-to-many) and the sender's mother tongue (native or non-native).

The linguistic features analysed are:

- i) overall register of the message measured by the level of formality or informality of its opening and closing;
- ii) the use of contractions;
- iii) the number of politeness indicators per message;
- iv) the number of non-standard linguistic features per message.

Our initial hypotheses, based on previous research, were that: computer mediated communication (CMC) reflects the informalization of discourse (Fairclough, 1995) and that CMC is not homogeneous but is made up of a number of genres and sub-genres that carry over distinctive linguistic features of traditional of-line genres. The aim of the study is to corroborate the hypothesis and to determine whether the writer's first language impinges upon the register of the message.

Methodology

In order to study the degree of formality of e-mails an analysis was made of a corpus of e-mail messages exchanged by members of academic institutions on the topic of Erasmus exchange programs. 100 e-mail messages were analysed: 25 one-to-many native messages, 25 one-to-one native messages, 25 one-to-many non-native messages and 25 one-to-one non-native messages.

Results and Discussion

The overall register of the message was measured by assigning to its salutation and farewell values for formality along a continuum of 0 to 1 and by examining the number of steps involved in the farewell: that is if there a one step closing or a two step closing with a pre-closing of the type "I look forward to hearing from you". The results for the overall register of formality are shown in Table 1.

Table 1.

Overall register of formality			
	Salutation	Farewell	Steps
1-many native	1,0	0.41	1,12
1-1 native	0.51	0.53	1.08
1-many non-native	0.93	0.51	1.18
1-1 non-native	0.74	0.61	1.69

These results largely conform to our initial hypotheses, but with interesting variations. It is clear that, in one-to-many messages, the greetings are very formal (1, the highest possible score, for natives and 0.93 for non-natives). It would seem that here there is clear carry over from the traditional business letter and memorandum as Yates and Orlikowski (1992) argued. As regards one-to-one communication both native and non-native salutations are more informal: 0.51 for natives and 0.74 for non-natives. In one-to-one communication, non-native writers are more formal for all categories. The sharp asymmetry between the formality of salutations and farewells of native one-to-many e-mails (1.0 vs. 0.41) is striking. Although more research is needed in this area, a tentative explanation is that the formality of the sign-off is being transferred to the electronic signature.

The use of contractions is a clear marker of informality (Biber, 1988). Table 2 shows the results for contractions in the corpus analysed:

Table 2.

	Possible contractions	Full forms	Contractions
1-to-many native	116	115 (99.13%)	1 (0.87%)
1-to-1 native	111	109 (98.19%)	2 (1.81%)
1-to-many non native	47	42 (89.36%)	5 (10.64%)
1-to-1 non-native	79	72 (91.13%)	7 (8.87%)

The analysis of the corpus surprisingly revealed a very low percentage of contractions in native e-mails (0.87% and 1.81%). Contractions were more frequent in non-native e-mails (10.64% and 8.87%). The greater use of contractions by non-native participants may reflect real stylistic differences for this formality marker.

Measures of politeness indicators have been obtained by counting the number of expressions of gratitude and pragmatic, routine formulae used in the mails:

Table 3.

Politeness indicators per message	
1-to-many native	3.22
1-to-1 native	2.28
1-to-many non native	1.09
1-to-1 non-native	1.31

As shown in the table native e-mails contain the highest number of politeness indicators per message. Again native speakers write considerably more formally than non-native speakers.

The results of the number of non-standard linguistic features per message are as follows:

Table 4.

Non-standard linguistic features per message			
	Misspellings	Non standard grammar/ spelling	Paralinguistic cues/ emoticons
1-to-many native	0.11	0.06	0.17
1-to-1 native	0.28	0.04	0.04
1-to-many non-native	0.32	0.23	0.55
1-to-1 non-native	0.08	0.12	0.50

The low number of errors per message is striking; it is probably because writers are aware that they represent their institutions. The lowest number is in non-native speakers. Non-native speakers may be more concerned about the idea of showing their accuracy in English.

The scores for non-standard grammar and spelling are very low. In these subgenres of CMC, the grammatical norms of formal letters seem to be firmly in place. Non-native speakers use paralinguistic cues and emoticons more, probably because it is easier for them to use these resources to be creative.

Although these mails show a very formal style of writing, we can observe a slight move towards the use of the new CMC linguistic features to communicate more expressively.

Conclusions

In conclusion, the results tend to suggest that there are significant stylistic and pragmatic differences between e-mails that can be established on the basis of their mode of communication, with one-to-many emails tending to be more formal and one-one emails incorporating more informal features. In addition, the results of the corpus analysed seem to indicate that, within International Standard English (McArthur 1998), stylistic and pragmatic features may be a significant parameter delimiting native and non-native varieties.

References

- Baron, N. B. 2000. *Alphabet to e-mail: How Written English Evolved and Where it is Heading*. London/New York, Routledge.
- Bunz, U.. Accomodating politeness indicators in personal electronic mail messages. <http://www.scils.rutgers.edu/~bunz/AoIR2002politeness.pdf> [22.09.2003].
- Crystal, D. 2001. *Language and the Internet*. Cambridge, Cambridge University Press.
- Fairclough, N. 1995. *Critical Discourse Analysis*. London, Longman.
- McArthur, T. 1998. *The English Languages*. Cambridge, Cambridge University Press.
- Rentel, N. 2005. Interlingual varieties in written business communication- intercultural differences in German and French business letters. <http://www.businesscommunication.org/conventions/Proceedings/2005/PDFs/08ABCEurope05.pdf>.
- Yates, J., Orlikowski, W.J., & Rennecker, J. 1997. Collaborative genres for collaboration: Genre systems in digital media. In *Proceedings of the 30 Annual Hawaii International Conference on System Sciences: Digital Documents 6*, 50-59. Los Alamitos CA, IEEE Computer Society Press.

All roads lead to advertising: Use of proverbs in slogans

Helena Margarida Vaz Duarte, Rosa Lúcia Coimbra and Lurdes de Castro Moutinho
Department of Languages and Cultures, University of Aveiro, Portugal

Abstract

This paper presents a research on the use of proverbs in written advertising texts from Portuguese press. The analysis focus on the presence of the proverb in the text both changed and unaltered. The different strategies of transformation are presented.

Introduction

Proverbs are wise sayings, short definitions of wisdom (Mieder, 1999). When we hear or read the sentence “All roads lead to Rome” we will easily recognize that we have encountered a proverb. It will, for sure, sound familiar. This characteristic is used in advertisement slogans in order to attract the reader’s attention. Being familiar with the sentence, the reader will feel more involved and the product will be presented as something close to the consumer.

The corpus of this research includes forty three written advertisements published in Portuguese press and also in outdoors, all of them including a proverb in the slogan.

In our corpus, these proverbs are sometimes left unaltered. But most of the times some kind of alteration is made to the sentence. In this paper, we show the type of changes performed in these slogans and their importance to the persuading power of the message. Several strategies are used to modify the proverb such as: lexical exchanges, syntactic alterations, elisions.

The use of proverbs and these strategies are also present in other text types, namely literary (Duarte Mendes, 2000; Nunes; Duarte, 2004) and journalistic (Coimbra, 1999).

Altered and unaltered proverbs

In our corpus, we noticed that in the great majority of the cases there is an alteration of the fixed form of the proverb. As can be observed in Figure 1, the difference is notorious. In fact, there are 13 cases in which there is no transformation on the linguistic form and the meaning of the proverb is also preserved.

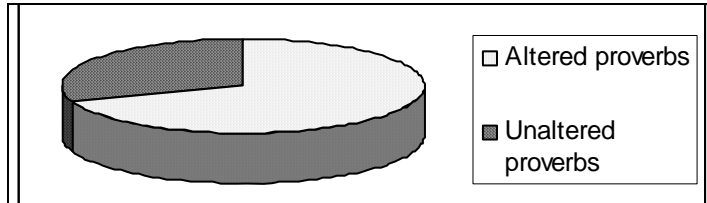


Figure 1. Altered and unaltered proverbs

Concerning the 30 altered forms, we may observe several different processes of transformation. The strategies may be grouped into three categories.

Lexical replacement

Among the 30 sentences that were altered, 22 were by lexical replacement, which means that this is the most preferred strategy. The replacement is accomplished by substituting one or more words of the original proverb by one or more words concerning the characteristics of the product advertised.

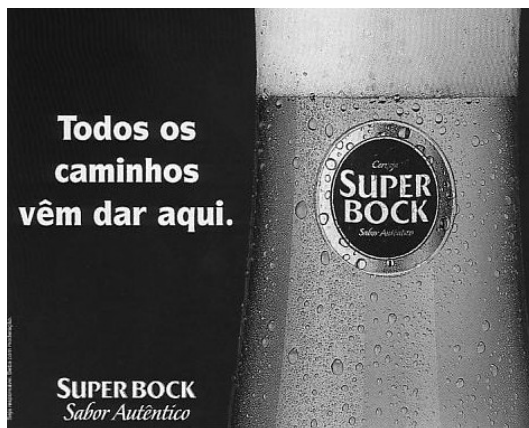


Figure 2. Example of lexical replacement

For example, as can be seen in figure 2, the proverb *Todos os caminhos vão dar a Roma* (*All roads lead to Rome*) is changed by substituting the word *vão* (*go*) by *vêm* (*come*), and the word *Roma* (*Rome*) by *aqui* (*here*). This place adverb means the advertised product, a beer.

Syntactical changes

With 7 occurrences, we found cases where the alteration of the fixed form was made on the syntactical level. The main process is the change of the

sentence form or type. Declarative is changed to interrogative; negative is changed into affirmative and vice-versa.

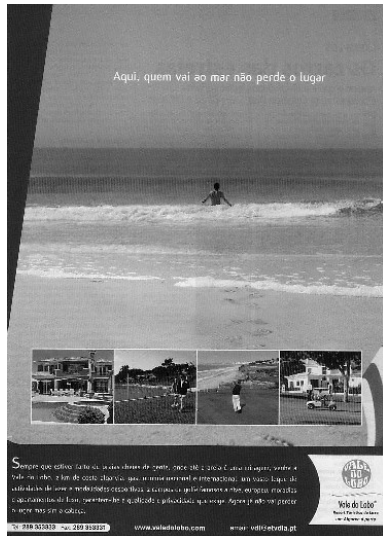


Figure 3. Example of syntactical change

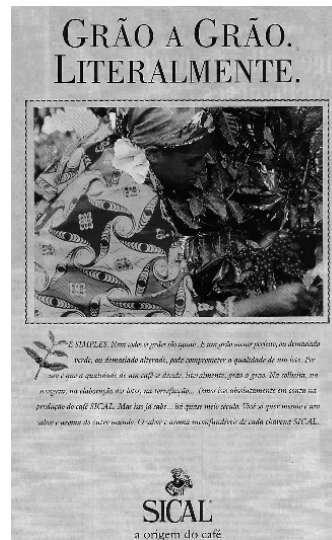


Figure 4. Example of lexical suppression

Figure 3 shows an example of a syntactical change in the proverb *Quem vai ao mar perde o lugar* (meaning that if you go away someone may take you place). The sentence, which originally was affirmative, is, on the slogan, negative. Thus, there is an emphasis on the quality and privacy of the touristic resort advertised.

Lexical suppression

Finally, we found an example in which the alteration consists on suppressing the final part of the proverb.

The fixed form of the proverb, *Grão a grão enche a galinha o papo* (*Slowly, slowly catches monkey*) is reduced to *Grão a grão*. Furthermore, the slogan adds the adverb *Literalmente* (*literally*) directing the reader's interpretation to the preoccupation on the careful selection of the coffee grains of the advertised brand.

Conclusion

We conclude that the majority of the slogans containing proverbs presents an alteration on its form, mainly due to lexical replacements.

In future, oral advertisements (radio and television) can also be studied in order to see if these or other strategies (prosodic clues, e.g.) are also used.

Another possibility of expanding this study is to verify the reader's skill to identify the original proverb as well as the meaning and intention of the altered form.

References

- Coimbra, R. L. 1999. Estudo Linguístico dos Títulos de Imprensa em Portugal: A Linguagem Metafórica. (thesis), Universidade de Aveiro.
- Duarte Mendes, H. M., 2000. Estudo da Recorrência Proverbial – de Levantado do Chão a Todos os Nomes de José Saramago (thesis). Universidade Nova de Lisboa.
- Mieder, W. 1999. Popular Views of the Proverb, vol. 5, n. 2. <<http://www.deproverbio.com>>
- Nunes, A. M.; Duarte Mendes, H. M. 2004. Alguns aspectos da reformulação parafrástica e não parafrástica em José Saramago e Mia Couto. In Actas do XIX Encontro Nacional da Associação Portuguesa de Linguística. Lisboa: APL, 623-630.

Perception of complex coda clusters and the role of the SSP

Irene Vogel¹ and Robin Aronow-Meredith²

¹Department of Linguistics, University of Delaware, U.S.A.

²Departments of French, Italian, German, and Slavic and Communication Science, Temple University, U.S.A.

Abstract

Modern Persian permits coda clusters, many of which violate the Sonority Sequencing Principle. In a syllable counting task, Persian speakers consistently perceived clusters in CVCC target items as monosyllabic, whereas English speakers generally perceived clusters existing in English as monosyllabic but those not existing in English as bi-syllabic. Moreover, the latter were perceived as monosyllabic more frequently if they adhered to the SSP than if they did not. In a follow-up experiment, French speakers performed a similar task, related to the clusters of that language. It is anticipated that the French speakers will exhibit similar perceptual behavior demonstrating the influence of the native language when the cluster exists in French, and the influence of the SSP if it does not.

Introduction

Cross-linguistically, it is generally observed that sequences of consonants in syllable onsets and codas are restricted by the Sonority Sequencing Principle (SSP), such that the sonority of the segments decreases from the nucleus out towards the margins of the syllable. The general sonority hierarchy is as follows:

(1) Sonority Hierarchy: Vowel > Glide > Liquid > Nasal > Obstruent

Despite the general tendency for languages to observe the SSP, a number of languages contain clusters that violate it, such as modern Persian, which permits numerous clusters in word final position, many of which violate the SSP (Alamolhoda 2000, Mahootian 1997).

(2) a. fekr 'thought'

b. hosn 'beauty'

Recent research has demonstrated that speakers of languages with relatively simple syllable structures have difficulty in accurately perceiving the number of syllables in words with complex syllable structures. For example, speakers of Japanese, a language with simple syllable structures, were unable to accurately identify the number of syllables in test items such as *ebzo* vs. *ebuzo*, perceiving three syllables in both cases (Dupoux et al. 1999). Such findings have been interpreted as an indication that listeners impose

their native language syllable structure on the strings they hear. Similar findings have also been reported by Kabak and Idsardi (2003) for Koreans listening to English stimuli. It should be noted, however, that the CV structure which is perceived, is the universally least marked syllable type. Thus these findings might also be interpreted as showing that when faced with a complex structure, Japanese (and Korean) listeners rely on universal principles, and favor the unmarked CV syllable structure.

In the present research, we first examined the perception of English speakers listening to CVCC Persian words in which several of the coda clusters also exist in English while others do not. Furthermore, some of the clusters observed the SSP and some did not. This allowed us to evaluate the relative contributions of the existence of a particular syllable type in one's native language and the role of universal principles, in particular the SSP, in perception behaviour. A second experiment with French listeners is in progress to assess the generalizability of the original English findings.

Perceptual study - English

In order to evaluate the relative roles of native language influence and the SSP, a perceptual experiment was conducted. Specifically, we tested the following hypotheses:

Hypothesis 1: In a CVCC structure, English speakers will perceive 1 syllable if the cluster is found in English; 2 syllables if not.

Hypothesis 2: In a CVCC structure, English speakers will perceive 1 syllable if the cluster observes the SSP; 2 syllables if not.

The subjects in the experiment were 22 native English speakers and 4 Persian speakers. The participants listened to pre-recorded Persian words and indicated whether they heard 1 or 2 syllables. The stimuli were 97 target words consisting of a CVCC syllable, as well as 20 CVC and 20 CVCVC words that served as distractors. The stimuli were all real words in Persian, and contained only consonantal segments also found in English, so as not to add any unnecessary complications for the English listeners. The set of targets was randomized twice and both lists were presented to each subject. Thus there were 194 targets per subject, which yielded a total 4,268 responses. Figure 1 shows the responses which were evaluated using an ANOVA.

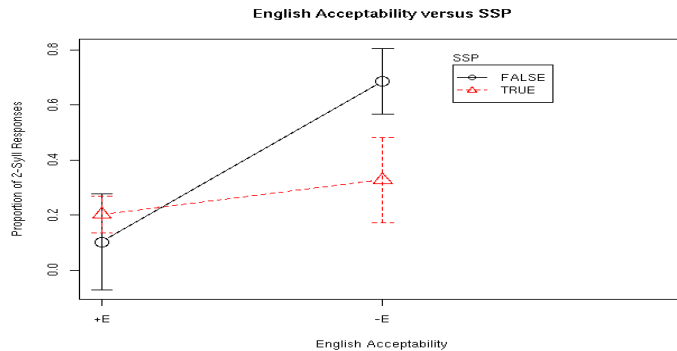


Figure 1. Perception of clusters based on Language and SSP.

These results revealed that clusters existing in English (+E) were perceived as monosyllabic significantly more frequently than clusters not found in English (-E) (e.g. [sk] vs. [šk]). Furthermore, it was found that if the cluster exists in English, the choice of one or two syllables was not affected by whether the cluster adhered to or violated the SSP. Among the clusters not existing in English, however, our results showed a significant effect of the SSP. Clusters that adhere to the SSP (True) were most often perceived as monosyllabic and clusters violating the SSP (False) were most often perceived as bi-syllabic (e.g. [šk] vs. [kr]). Thus both Hypotheses 1 and 2 were supported.

Perceptual study – French

To further investigate the relative influence of L1 and the SSP in the perception of coda clusters, a comparison study is underway with French native speakers. French permits a variety of word final clusters (which surface when word final schwa is deleted), some of which are closer to those of Persian than are the English coda clusters. While most clusters in French observe the SSP (e.g. [rk]), a number do not (e.g. [bl]).

In this experiment, the stimuli were nonce words, which consisted of 84 CVCC targets as well as 10 CVC and 10 CVCVC distractors. All items are possible words of Persian, and were recorded by a native speaker of Persian. Of the target items, 44 are words with clusters that are found in French, while 40 have clusters that are not. Furthermore, 44 targets conform to the SSP, while 40 do not. As in the English study, the set of stimuli is randomized twice, and each subject hears both sets. Again, the subjects' task is to indicate whether they perceive one or two syllables.

It is hypothesized that those clusters present in French will be perceived as monosyllabic more frequently than those not present in French. Furthermore, we expect that the perception of the clusters not present in French will

show sensitivity to the SSP. That is, it is predicted that the clusters that conform to the SSP will tend to be perceived as monosyllabic, while those that do not conform to the SSP will be perceived as bi-syllabic.

Conclusions

It has been shown that English speakers' perception of Persian coda clusters appears to be determined by the presence or absence of the cluster in English as well as by the cluster's adherence to or violation of the SSP. One syllable was perceived if the cluster was acceptable in English, while two syllables were perceived when it was not. Furthermore, one syllable was perceived if the cluster conformed to the SSP, while two syllables were perceived if it did not, in particular when the cluster was not found in English. Similar findings are anticipated for French listeners. Thus, we propose that while there is no doubt that one's native language, or L1, affects a listener's perception of another language, in some cases the perceptual behaviour might, in fact, also be due to more universal properties of phonology, ones that give rise to the patterns of the L1 in question in the first place.

References

- Alamolhoda, S. M. 2000. Phonostatistics and Phonotactics of the Syllable in Modern Persian. Helsinki, Studia Orientalia. The Finnish Oriental Society.
- Dupoux, E., K. Kakehi, Y. Hirose, C. Pallier, & J. Mehler. 1999. Epenthetic Vowels in Japanese: A Perceptual Illusion? *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1568-1578.
- Kabak, B., Idsardi, W. 2003. Syllabically Conditioned Perceptual Epenthesis. *Parasession on Phonetic Sources of Phonological Patterns: Synchronic and Diachronic Explanations*, 233-244
- Mahootian, S. 1997. *Persian*. London & New York, Routledge.

Factors influencing ratios of filled pauses at clause boundaries in Japanese

Michiko Watanabe¹, Keikichi Hirose², Yasuharu Den³, Shusaku Miwa² and Nobuaki Minematsu¹

¹ Graduate School of Frontier Sciences, University of Tokyo, Japan

² Graduate School of Information Science and Technology, University of Tokyo, Japan

³ Faculty of Letters, Chiba University, Japan

Abstract

Speech disfluencies have been studied as clues to human speech production mechanisms. Major constituents are assumed to be principal units of planning and disfluencies are claimed to occur when speakers have some trouble in planning such units. We tested two hypotheses about the probability of disfluencies by examining the ratios of filled pauses (fillers) at sentence and clause boundaries: 1) the deeper the boundary, the higher the ratio of filled pauses (the boundary hypothesis); 2) the more complex the upcoming constituent, the higher the ratio of filled pauses (the complexity hypothesis). The both hypotheses were supported by filler ratios at clause boundaries, but not by those at sentence boundaries. The results are discussed in light of speech production models.

Introduction

Disfluencies such as filled pauses (fillers) and repetitions are ubiquitous in spontaneous speech, but rare in speech read from written texts. Therefore, they are believed to be relevant to on-line speech production: When speakers have some trouble in speech planning, they tend to be disfluent.

It has been claimed that disfluencies are more frequent at deeper syntactic and discourse boundaries in speech. In early studies it was argued that disfluencies are frequent at the points at which transition probabilities of linguistic events are low and as a consequence the information value is high. Major syntactic boundaries are assumed to be a type of such locations (Maclay & Osgood, 1959). More recent studies have shown that disfluencies tend to cluster at the point at which what the speaker talks about widely shifts (Chafe, 1980), or near the point at which many listeners recognise a boundary (Swerts, 1998). We call the claim that the deeper the boundary, the higher the disfluency ratio *the boundary hypothesis*.

There have been arguments whether speech planning is incremental or hierarchical. Holmes (1995) examined filler ratios at the beginning of basic and surface clauses in English and French, and found no significant differ-

ence in ratios between the two types of locations. Holmes argued that speakers plan one basic clause at one time and that there is no added difficulty even if the upcoming clause contains other clauses. Clark and Wasow (1989), on the other hand, examined repetition rates of articles and pronouns and found that the more complex the following constituents, the higher the ratios. They claimed that the complexity of the following constituents does affect speakers' planning load and consequently the ratio of disfluencies. We call Clark and Wasow's view *the complexity hypothesis*, following their naming. We tested the two hypotheses by examining filler ratios at sentence and clause boundaries in a Japanese speech corpus.

Japanese adverbial clauses are marked by connective particles or certain conjugations of verbs, adjectives or copula markers at the end of the clauses. They always precede main clauses. The clause order of Japanese complex sentences is shown below.

((adverbial clause <connective particle>) (main clause))

Adverbial clauses are classified into three groups according to the degree of dependency on the main clauses (Minami, 1974). *Type A* clauses are the most dependent on the main clauses. Grammatically they can have neither their own topics nor subjects. *Type B* clauses can contain their own subjects, but not their own topics. *Type C* clauses can have both their own topics and subjects. Therefore, *Type C* clauses are the most independent of the main clauses. Consequently, it is assumed that boundaries between *Type C* and the main clauses are deeper than the boundaries between *Type A* or *Type B* and the main clauses and that boundaries between *Type B* and the main clauses are deeper than the boundaries between *Type A* and the main clauses. Therefore, it is predicted from the boundary hypothesis that filler ratios are highest at *Type C* boundaries and lowest at *Type A* boundaries. We considered sentence boundaries as well. As sentence boundaries are assumed to be deeper than clause boundaries, we predict that filler ratios at sentence boundaries are even higher than those at *Type C* clause boundaries.

We predict from the complexity hypothesis that the more complex the upcoming clause, the higher the filler ratio at the boundary. We employed the number of words in the clause as an index of complexity. Details of the experiment are described in the remaining sections.

Method

We analysed 174 presentations (69 academic and 105 casual presentations) in *the Corpus of Spontaneous Japanese* (CSJ) (Maekawa, 2004). The classification of three clause types employed in the present study is described in Table 1. Minami (1974)'s classification was partly modified based on the relevant studies such as Takanashi et al. (2004).

First, we marked *A*, *B* or *C* at each adverbial clause boundary and *D* at sentence boundaries. Then, the number of words in each clause between the boundaries was counted. The clauses were grouped into three according to the number of words in the clause: short (1- 8 words), medium (9-16 words), and long (more than 16 words). Filler rate for each length group of the following clauses at each boundary type was computed for each presentation, and the mean values of the conditions were compared.

Table 1: Classification of adverbial clauses

Type	connective	meaning, usage
A	~nagara, ~tutu	expresses accompanying actions
	~mama	expresses continuous accompanying actions
	~tari, dari	Lists actions or situations
B	~to, ba, tara, nara	if ~
	~te, te kara, te mo	~ and, after ~, even if ~, respectively
	~yoo ni	so that ~
	adverb forms	~and
C	~kara, node	as ~ (reason)
	~noni, ke(re)do	though ~,
	~ga	although ~, ~ but
	~si	~ and (lists similar actions or features)
	~de	~ and
	~masite, ~desite	~ polite auxiliary verb + and

Results and discussions

Type A clauses were excluded from analysis because of low frequency in each presentation and treated as adverbial phrases. Table 2 illustrates mean filler ratios in nine conditions. Because of the space limitation, we describe only the main results. Repeated measures ANOVA showed main effects of the boundary type and the length factor. An interaction between the two factors was significant, $F(4, 680) = 4.02$, $p < .005$. We first compared the ratios by boundary type. For *Type B* boundaries, the filler ratio was higher, the longer the following clause, long vs. short: $t(170) = 5.18$, $p < .001$; long vs. medium: $t(170) = 2.95$, $p < .05$; medium vs. short: $t(170) = 3.08$, $p < .007$. For *Type C* boundaries, only the filler ratio before long clauses was significantly higher than that before short clauses, $t(170) = 2.75$, $p < .05$. For *Type D* boundaries, there was no significant difference among length factors, $F(2, 169) = .25$, $p = .78$. When we compared the ratio by length group, there were significant differences among boundary types in all the length groups. Paired comparisons showed that in all the length groups, the ratios of fillers at *Type C* and *Type D* boundaries were significantly higher than those at

Type B boundaries, but that there were no significant differences between the ratios at *Type C* and *Type D* boundaries.

The complexity hypothesis was supported by the results for *Type B* boundaries, and also for *Type C* boundaries with less degree, but not supported by the results for *Type D* boundaries. The boundary hypothesis was supported by the difference between *Type B* and *Type C* boundaries and the difference between *Type B* and *Type D* boundaries, but there was no significant difference between *Type C* and *Type D* boundaries.

We speculate that these results derive from difference in the most influential factors at different types of boundaries. At deeper boundaries such as *Type C* and *Type D* boundaries, most of speakers' attention and time tend to be devoted to message conceptualisation. As a consequence, speakers cannot plan linguistic units far ahead at those points and the complexity effects are relatively small. In contrast, at shallower boundaries at which cognitive loads for conceptualisation are not so heavy, the complexity effects seem to play a significant role.

Table 2: Rate of clause boundaries with fillers (%) before short (1-8 words), medium (9-16 words) and long (over 16 words) clauses.

Boundary type	1-8 words	9-16 words	17- words
B	26	29	33
C	39	40	43
D	40	39	40

References

- Chafe, W. 1980. The deployment of consciousness in the production of a narrative, in Chafe, W. (ed.) *The Pear Stories, Cognitive, Cultural, and Linguistic Aspects of Narrative Production*, 9-50. New Jersey, ALEX Publishing Corporation.
- Clark, H. H. & Wasow, T. 1998. Repeating words in spontaneous speech. *Cognitive Psychology* 37, 201-242.
- Holmes, V. M. 1995. A crosslinguistic comparison of the production of utterances in discourse. *Cognition*, 54, 169-207.
- Levelt, W. J. M. 1989. *Speaking*. The MIT Press, Cambridge, Massachusetts.
- Maclay, H., & Osgood, C. E. 1959. Hesitation phenomena in spontaneous English speech. *Word* 15, 19-44.
- Maekawa, K. 2004. Outline of the Corpus of Spontaneous Japanese. In Yoneyama, K. and Maekawa, K. (eds.) *Spontaneous Speech: Data and Analysis*. National Institute for Japanese Language.
- Minami, F. 1974. *Gendai nihongo no kouzou* (The structure of modern Japanese). Taisyukan syoten, Tokyo.
- Swerts, M. 1998. Filled pauses as markers of discourse structure. *Journal of Pragmatics* 30, 485-496.
- Takanashi, K., Uchimoto, K., & Maruyama, T. 2004. Identification of clause units in CSJ. In vol1, the Corpus of Spontaneous Japanese.

Assessing aspectual asymmetries in human language processing

Foong Ha Yap, Stella Wing Man Kwan, Emily Sze Man Yiu, Patrick Chun Kau Chu and Stella Fat Wong
Department of Linguistics and Modern Languages, Chinese University of Hong Kong, China

Abstract

This paper reports reaction time studies on aspect processing, and highlights that aspectual asymmetries (perfective vs. imperfective facilitation) in terms of reaction time is dependent on verb types.

Introduction

It is generally believed that the human mind constructs mental models of events and situations that unfold in the world around us. Previous studies indicate that various cues contribute to the dynamic representation of these mental models. In particular, Madden and Zwaan (2003) have shown that, with respect to accomplishment verbs, perfective sentences (e.g. *He lit a fire*) are processed faster than imperfective sentences (e.g. *He was lighting a fire*). This perfective advantage was also found in a number of East Asian languages—e.g. Cantonese (Chan et al., 2004) and Japanese (Yap et al., in press). In this paper we report findings from a series of reaction time studies that investigate the effect of both *grammatical aspect* and *lexical aspect* on language processing (see Yap et al., 2006 and Wong, 2006 for detailed discussions). We further discuss issues of methodological interests that have implications for our understanding of cognitive processing.

Definition of aspect

Grammatical aspect allows us to view a situation as temporally bounded or unbounded (i.e. with or without endpoint focus). More specifically, perfective aspect allows us to view the event as a whole ('bounded' perspective), while imperfective aspect allows us to focus on the internal stages of an event ('unbounded' perspective) (Comrie, 1976). Lexical aspect refers to the situation type denoted by the verb (predicate). Each situation type is distinguished on the basis of temporal features such as dynamism, durativity and telicity. Vendler (1967) identifies four basic situation types: states (e.g. know), activities (e.g. run), accomplishments (e.g. run a mile) and achievements (e.g. break). Smith (1991) includes a fifth category: semelfactives (e.g. cough, iteratively). The present study compares the processing times of two types of grammatical aspect markers (perfectives vs. imperfectives) on two situation types or lexical aspect categories (accomplishments vs. activities).

Methodology

Forced-choice utterance-and-picture matching tasks were used. For each test item, participants first heard a Cantonese utterance with a perfective aspect marker (*zo2*) or an imperfective aspect marker (*gan2*). They then immediately were shown a pair of pictures, one picture depicting a completed event and the other depicting an ongoing event. The participants then had to match which picture best describes the utterance they had just heard by pressing the corresponding key on the keyboard (the letter A for the picture on the left and the numeral 5 for the picture on the right). The participants' reaction times were recorded using millisecond INQUISIT software. The ISI between the onset of stimulus and target was 2200ms. Each picture remained on the screen for a maximum of 3 seconds. Only correctly matched responses completed within 3 seconds were analyzed. A perfective utterance and completed picture pairing constitutes a matched perfective response; an imperfective utterance and ongoing picture pairing constitutes a matched imperfective response. The reaction times for matched perfectives and matched imperfectives were compared using a ANOVAs.

The first part of this experiment involved a simple pair-wise design. This design was used to compare the reaction times of perfective vs. imperfective utterances in contexts involving *only one situation type* (accomplishments only or activities only). There were 20 test items; plus 8 trial items for the practice session at the beginning of the experiment. All stimuli were counterbalanced. The subjects (N=18) were native Cantonese speakers (mean age approximately 18).

The second part of the experiment involved a more complex 2x2 design. This design examined the reaction times of perfective vs. imperfective utterances *across two situation types* (accomplishments and activities). Hence it tested for potential interaction effects between grammatical aspect *and* lexical aspect. For this more complex design, there were 24 test items and 8 trial items. The subjects (N=32) were native speakers of Cantonese (also mean age approximately 18).

Results

Pair-wise design (accomplishment verbs vs. activity verbs)

In the pair-wise design, there was evidence of perfective advantage with accomplishment verbs, consistent with earlier findings. The effect of grammatical aspect was significant ($p = .001$). Perfective *zo2* utterances (mean=941ms, SD=242) were processed significantly faster than imperfective *gan2* utterances (mean=1032ms, SD=289). However, with activity verbs, the direction of aspectual asymmetry was reversed, with results showing imperfective facilitation instead. Imperfective *gan2* utterances (mean=1125ms, SD=367) were processed significantly faster than perfective *zo2* utterances (mean=1211ms, SD=379). The effect of grammatical aspect was significant ($p = .025$).

Crucially, the combined results indicate that the perfective advantage observed for accomplishment verbs in earlier studies is not generalizable to all verb types. In particular, there is evidence of strong imperfective facilitation for activity verbs. Table 1 highlights the observed aspectual asymmetries.

Table 1. Aspectual asymmetries across verb types (based on Yap et al., 2006)

Experiment	Verb class	Perfective zo2	Imperfective gan2	p- value	Aspectual facilitation
Pair-wise 1	Accomplishment	941ms (SD=242)	1032ms (SD=289)	p = .001	Perfective
Pair-wise 2	Activity	1211ms (SD=379)	1125ms (SD=367)	p = .025	Imperfective

In the case of accomplishment verbs, Madden and Zwaan (2003) earlier suggested that perfective utterances are processed faster because their inherent telicity (i.e. endpoint focus) allows the human mind to more rapidly converge on a mental representation of the event. However, what are the implications of imperfective facilitation for activity verbs? Yap et al. (2006) suggest that the tendency for imperfective constructions to focus on the internal stages of events perfectly matches the atelic nature of activity verbs. This then contributes to more rapid construction of mental models related to the event.

Complex design (accomplishment verbs + activity verbs)

Results from the more complex design, which simultaneously involved both activity verbs and accomplishment verbs, indicate that aspectual asymmetry is often sensitive to context. The results showed that, for activity verbs, imperfective *gan2* utterances (mean=1096ms, SD=326) were processed significantly faster than perfective *zo2* utterances (mean=1239ms, SD=445). The main effect of grammatical aspect was significant ($p = .011$), and the interaction effect of lexical and grammatical aspect was also significant ($p = .017$). A follow-up *t*-test showed that imperfective facilitation for activity verbs was statistically significant ($p < .001$). However, for *accomplishment verbs*, there was no significant difference between the reaction times of perfective and imperfective utterances ($p = .884$). Thus, whereas imperfective facilitation remained robust in complex environments involving activity and accomplishment verbs, perfective facilitation turned out to be rather fragile. Future studies will need to investigate the degree of robustness/fragility of perfective and imperfective facilitations in other types of complex environments (e.g. accomplishment and achievement verbs, with and without the presence of activity verbs).

Why use both pair-wise and complex designs?

The above results already justify the inclusion of both pair-wise and complex designs. With a pair-wise design, we can tease out the effect of grammatical aspect within a single verb

class. For example, perfective facilitation was found with accomplishment verbs, while imperfective facilitation was found with activity verbs. The stability of these two types of facilitation (perfective and imperfective) can further be tested in complex environments that more closely resemble natural discourse processing in real time. Our more complex design reveals that when both accomplishment and activity verbs are used, perfective facilitation is fragile while imperfective facilitation remains robust. Yap et al. (2006) suggest that neighborhood density is a factor. A squishing effect is found when accomplishment verbs compete for mental resources with activity verbs. Arguably, a greater concentration of [+durative] features from both activities and accomplishments, as opposed to the previously balanced concentration of [+durative] and [+telic] features in an accomplishment only environment, undermines an inherent perfective advantage.

Significance of reaction time studies for aspect studies

Reaction time studies allow us to examine how aspectual asymmetries work in real-time. More specifically, they provide us with a means of empirically examining how grammatical aspect and lexical aspect interact with each other and how such interaction contributes to the dynamic representation of events in the human mind. Subtle effects such as neighborhood density can also be assessed through reaction time studies. Equally important, reaction time studies provide us with baseline information before we proceed with more sophisticated and high-cost ERP and fMRI testing.

Acknowledgements

We gratefully acknowledge funding from Direct Grant 2004-06 (#2010255) from the Chinese University of Hong Kong and Competitive Earmarked Research Grant 2005-07 (#2110122) from the Research Grants Council of Hong Kong. We also thank Lai Chim Chow, Irene Lam, Calvin Chan, Kimmee Lo, Edson Miyamoto, Him Cheung and participating schools for their valuable help in various ways in the studies.

References

- Chan, Y.H., Yap, F.H., Shirai Y. and Matthews, S. 2004. A perfective-imperfective asymmetry in language processing: Evidence from Cantonese. *Proc. 9th ICLL*, 383-391. ASGIL, National Taiwan University, Taipei.
- Comrie, B. 1976. *Aspect*. Cambridge, UK: Cambridge University Press.
- Madden, C.J. and Zwaan, R.A. 2003. How does verb aspect constrain event representation? *Memory & Cognition*, 31, 663-672.
- Smith, C. 1991. *The parameter of aspect*. Dordrecht: Kluwer Academic Press.
- Vendler, Z. 1967. *Linguistics in Philosophy*. Ithaca, NY: Cornell University Press.
- Wong, F. 2006. Reaction time study on inherent lexical aspect asymmetry in Cantonese. Unpublished senior thesis in Linguistics. Department of Linguistics and Modern Languages, Chinese University of Hong Kong.
- Yap, F. H., Kwan, W.M., Yiu, S.M., Chu, C.K., Wong, F., Matthews, S. and Shirai, Y. 2006. Aspectual asymmetries in the mental representation of events: significance of lexical aspect. Paper presented at the 28th Annual Conference of the Cognitive Science Society, Vancouver, July 26-29.
- Yap, F. H., Inoue, Y., Shirai Y., Matthews, S., Wong, Y.W., and Chan, Y.H. (in press). Aspectual asymmetries in Japanese: Evidence from a reaction time study. *Japanese/Korean Linguistics*, vol. 14. Stanford, CSLI.

Toward a rich phonology

Robert Port

Departments of Linguistics and Cognitive Science, Indiana University, USA

Abstract

A radically new conception of linguistic representations is proposed. The claim is that language is stored in memory in the form of large distributions of specific utterances in a rich high-dimensional space, sometimes called exemplar memory. This is the form the brain uses for understanding and creating utterances in real time. In contrast, the abstract, speaker-independent description of language (as modelled by alphabetical orthographies and by linguistic descriptions using phonemes, etc.) exhibits many structures and patterns that comprise a social institution, maintained by speakers over time, and approximating a discrete system made from components. However, these phenomena, shaped by social as well as articulatory and auditory factors, play no clear role in real-time language processing.

The traditional view

For about a century, linguists have trusted their intuitions that speech presents itself to our consciousness in the form of *segments*, that is, consonant or vowel units (Saussure, 1916; Jones, 1918, p. 1; Ladefoged, 1972; IPA, 1999) but these powerful intuitions may be partially (or wholly) a result of the lifelong literacy training to which all readers of this paragraph have been subjected. The vast majority of experimental evidence supports a very rich memory for language (resembling our detailed and context-sensitive memory for everyday events and activities, Nosofsky, 1986; Shiffrin and Steyvers, 1997), not a memory that uses abstract, speaker-independent tokens in serial order. However, the abstract, phoneme-based view has prevailed in the field despite many kinds of evidence against it.

Counterevidence

The most powerful evidence that words are not represented in memory using a phonology-like code consisting of segments (represented by letters) and features is found in recognition memory experiments. Linguistic theory predicts that words will be remembered in "linguistic form", that is, they will be stored using abstract, serially ordered "spellings" in phonological units. So, if we hear someone say *tomato*, then what is supposedly stored and what should influence later cognitive tasks (such as recognizing the repetition of a word) should be some phonological spelling of, say, *tomato*. Details about the specific utterance, such as the identity and sex of the

speaker or the timing pattern of the pronunciation should not be stored (Halle, 1985). But there is persuasive evidence that speaker identity and timing patterns do influence performance on a recognition memory task (Goldinger, 1996; Palmeri, et al, 1993). For example, if a subject hears a list of spoken words and is asked to indicate when a word is repeated in the list, accuracy naturally declines the greater the amount of time between the first presentation and the second. But if this is done with a list pronounced by a variety of voices, then words that are repeated in the same voice as the first are recognized about 8% more accurately than when repeated in a different voice. The improvement lasts for up to a week. This implies that speakers store much richer and more detailed representations than linguists thought.

Many other well-known phenomena related to speech perception are compatible with the view that the form of word storage is highly concrete and detailed, that is, approximately an "exemplar memory." For example, the formidable problem of "coarticulation" – the problem of how speakers "hear" speech as though it consists of nonoverlapping, context-free segments when the auditory stimulus has a great deal of temporal overlap and context-sensitive variation (Kent and Minifie, 1977) disappears with rich memory. The answer is that memory does not extract an abstract, context-free invariant for each consonant and vowel. It is only the intuitions of someone trained to read and write using an alphabet that find a problem here (Morais, et al, 1979; Rayner, et al, 2001). Similarly, short-term memory or the so-called "phonological loop" (Baddeley, 1986), appears to be a motor store, something very concrete and sensitive to context and not an abstract, static code (Wilson, 2001). Further evidence comes from incomplete neutralization (Port and Leary, 2005) and the apparently continuous variation in sound change (Labov, 1963).

All these sources of evidence imply a speech memory that stores huge amounts of speech material coded by an extremely rich auditory code that is sure to differ in detail from person to person. It also implies that the notion of contrast plays no role whatever in linguistic memory. We seem to remember whatever we can about all details of specific utterances. Abstractions and generalizations can be easily extracted as needed from a memory containing concrete instances (Hintzman, 1986).

Rich phonology

If words are stored in this detailed and indeterminate way, then how can linguistic description be done? The answer proposed here is that phonological patterns have little to do with the memory structure of speech (despite our linguistic intuitions favoring alphabetical descriptions). Phonological patterns are found in the corpus of speech by some community. The linguist looks across a large set of utterances and, using whatever

descriptive tools may be available, describes the patterns found there. This corpus is, of course, the ambient language as it presents itself to the language learner. Some of these patterns (most of the traditional phonological phenomena involving phonemes, features, syllables, etc.) can be described using an alphabet of phonetic symbols. Other patterns require more careful measurements of time or frequency (e.g., voice-onset time, mora patterns, formant transitions, spectrum shapes, intonation contours, etc.). Both kinds of descriptions aim to capture (a) the properties that make one language different from another and that (b) represent distinctions that are exploited by speakers of the language to differentiate various sets of lexical items. Such descriptions will be essential tools for teaching a new language to adult speakers of another language and provide a practical basis for development of an orthography (in cases where one is needed). But neither description should be assumed to play much of a role in the realtime perception and production of speech.

If speech memory is so much richer than we thought, then why, one might ask, do languages have phonological systems at all? And why do small alphabets work as well as they do for recording them? The answer is not completely clear, but over time the language as institution is shaped by users toward a system with limited degrees of freedom. Presumably such a system is easier to learn and to understand and may facilitate creating new words. But this does not imply that the apparent "units" of this schematic phonology are units in human memory for language.

Conclusions

The story being told here represents a radical break with the past. It is asserted that the powerful intuitions we have relied upon for a century that language manifestly has a segmental structure in terms of discrete letter-like units are primarily the consequence of our literacy education. We were taught to listen to and think about language in segmental terms because these skills are what is demanded for skillful use of our orthography. Consequently we insisted the "real" structure of language is discrete and segmental too despite all the contrary evidence. The discipline of phonology cannot, as linguists from Saussure to Chomsky and Halle had hoped, be both a description of the psychological code for language in memory and also be a communicable description of our language suitable for writing and teaching to others.

Acknowledgements

Thanks to David Pisoni, Susannah Levi and Mark Van Dam for discussion.

References

- Baddeley, A. D. 1986. *Working Memory*. Oxford, U. K.: Oxford University Press.
- Chomsky, N., and Halle, M. 1968. *The Sound Pattern of English*. New York: Harper and Row.
- Goldinger, S. D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22, 1166-1183.
- Halle, M. 1985. Speculations about the representation of words in memory. In V. Fromkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* (pp. 101-114). Orlando, Florida: Academic Press.
- Hintzman, D. L. 1986. 'Schema abstraction' in a multiple-trace memory model. *Psychological Review*, 93, 411-428.
- IPA. 1999. *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge, England: Cambridge University Press.
- Jones, D. 1918. *An Outline of English Phonetics*. Leipzig, Germany: Teubner.
- Ladefoged, P. 1972. *A Course in Phonetics*. Orlando, Florida: Harcourt Brace Jovanovich.
- Nosofsky, R. 1986. Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115, 39-57.
- Kent, R. and Minifie, F. 1977. Coarticulation in recent speech production models. *Journal of Phonetics* 5, 115-135.
- Labov, W. 1963. The social motivation of a sound change. *Word*, 19, 273-309.
- Morais, J., Cary, L., Alegria, J., and Bertelson, P. 1979. Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 7, 323-331
- Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. 1993. Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology, Learning, Memory and Cognition*, 19, 309-328.
- Port, R. F., and Leary, A. 2005. Against formal phonology. *Language*, 81, 927-964.
- Port, Robert 2006 (under review). Words, symbols and rich memory. *New Ideas in Psychology*.
- Rayner, K., Foorman, B., Perfetti, C., Pesetsky, D., and Seidenberg, M. 2001. How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, 2, 31-74
- Saussure, F. d. 1916. *Course in General Linguistics* (W. Baskin, Trans.). New York: Philosophical Library.
- Shiffrin, R., and Steyvers, M. 1997. A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin and Review*, 4, 145-166.
- Wilson, M. 2001. The case for sensorimotor coding in working memory. *Psychonomic Bulletin and Review*, 8, 44-57.

Index of authors

- Abelin, Å., 61
Alexandris, C., 65
Alexiadou, A., 1
Alexopoulou, T., 69
Alves, M. A. L., 221
Aronow-Meredith, R., 249
Astruc, L., 73
Awad, Z., 77
Bader, M., 149
Bagshaw, P., 25
Bailly, G., 25, 141
Baltazani, M., 81
Barberia, I., 85
Bertrand, R., 121
Botinis, A., 89
Breton, G., 25
Capitão Silva, S. M., 221
Caplan, D. N., 9
Carson-Berndsen, J., 181
Chen, Y., 93
Chun Kau Chu, P., 257
Clemens, G., N, 17, 97
Coimbra, R., L., 101, 245
Cole, J., 165
Cummins, F., 105
D'Imperio, M., 121
Darani, L. H., 109
De Deyne, S., 113
De Moraes, J. A., 117
Den, Y., 253
Di Cristo, A., 121
Economou, A., 205
Efsthathopoulou, N. P., 125
Elisei, F., 141
Erlendsson, B., 137
Fat Wong, S., 257
Feth, L. L., 153
Fleta, B. M., 241
Flouraki, M., 129
Folli, R., 133
Fon, J., 173
Fotinea, S.-E., 65
Fourakis, M., 89
Fox, R. A., 153
Gawronska, B., 137
Geumann, A., 181
Gibert, G., 141
Govokhina, O., 25
Gryllia, S., 145
Guerti, M., 193
Ha Yap, F., 257
Häussler, J., 149
Hawks, J. W., 89
Hirose, K., 253
Hsin-Yi, L., 173
Jacewicz, E., 153
Jesus, L. M. T., 177, 221
Joffe, V., 157
Kakavoulia, M., 205
Kanellos, I., 229
Katsos, N., 161
Keller, F., 69
Kewley-Port, D., 33
Koo, H., 165
Lehtinen, M., 169
Lousada, M. L., 177
Macek, J., 181
Martin, Ph., 185
Miltsakaki, E., 189, 237
Minematsu, N., 253
Miwa, S., 253
Moudenc, T., 229
Moutinho de Castro, C., 101, 245
Nikolaenkova, O., 137
Orfanidou, I., 89
Ouamour, S., 193
Ozturk, O. I., 201
Papadopoulou, Z., 197
Papafragou, A., 41, 201
Patsala, P. 189
Payne, E., 133
Perkell, J. S., 45
Port, R. 261
Portes, C., 121
Prieto, P., 73
Protopapas, A., 205
Ridouane, R., 17, 97
Rodero, E., 209
Rucart, P., 213
Sabater Pérez, C., 241
Sayoud, H., 193
Schiller, N. O., 53
Seneviratne, S., 217
Spinu, L., 225
Storms, G., 113
Suciu, I., 229
Sysoeva, A., 233
Sze Man Yiu, E., 257
Tsaklidou, S., 237
Tse Kwok-Ping, J., 173
Turney, E., 241
Van Lommel, S., 113
Varlokosta, S., 157, 205
Vaz Duarte, H. M., 101, 245
Vogel, I., 249
Watanabe, M., 253
Wing Man Kwan, S. 257
-

ISBN: 960-6608-57-3