

EVALUATING METHODS FOR ELICITING DIALECTAL SPEECH*

EVIA KAINADA
MARY BALTAZANI
University of Ioannina

We compare four experimental methods specifically designed for eliciting dialectal speech suitable for quantitative linguistic analysis, ranging from conversational speech to tightly controlled tasks. We test each method's success in eliciting enough instances of the target linguistic variable spoken in authentic dialectal speech. The success of each method is measured by the frequency of occurrence of (i) three dialectal phenomena as markers of dialectal speech (unstressed high vowel deletions, unstressed mid vowel raising and stressed mid vowel diphthongization), and (ii) the linguistic phenomenon under investigation, in this instance the occurrence of pre-nuclear pitch accents in tandem with vocalic deletion. All methods were successful in eliciting dialectal speech, but only the most controlled task was successful in extracting both the appropriate intonational contour and most dialectal authenticity. We conclude that tightly controlled experiments can provide the setting for dialectal linguistic research.

1 Introduction

In a recent article by Post and Nolan (2012), evaluating different elicitation methods for prosodic research, they stress the need for balance between what they call 'ecological validity' (i.e., emulation of natural speech communication) and controlled elicitation (to secure sufficient comparable data). The term 'ecological validity', as we understand it, refers to the production of speech delivered in an informal register, as opposed to more formal, read speech. We would like

* This research has been funded by the State's Scholarship Foundation (IKY) as a post-doctoral grant to the first author, and by a NSRF funded THALIS project ("Vocalect"), as a research grant to the second author. We would like to thank Theo Marinis for allowing the use of pictures from his NWO research programme '*Cross-linguistic study of the production and processing of grammatical morphemes in L2 children compared to children with Specific Language Impairment*' with Susan Edwards, and Dimitris Papazachariou for his helpful comments. Special thanks go to Katerina Nicolaidis, Eleni Tsartsioni, Giannis Themelis and Theofano Christou for their help with material recordings. We are, of course, greatly indebted to all our participants.

Corresponding Author: ekainada@cc.uoi.gr (Evia Kainada), mbaltaz@cc.uoi.gr (Mary Baltazani)

to enrich the meaning of this term by adding another parameter to it, which is essential for dialectologists, that of dialectal authenticity. In designing the methodology for fieldwork on dialectal variation, fieldworkers aspire to not only approximate natural communication in register, but also capture dialectal authenticity, avoiding influences from the dominant linguistic norm of a more prestigious standard variety. As is well known in Speech Accommodation Theory (e.g. Bebee and Giles, 1984; Giles et al, 1977; Giles and Philip, 1979; Giles et al, 1991) individuals might de-accentuate ethnolinguistic characteristics and converge toward the dominant culture when they desire social approval from members of the dominant culture, since standard accents connote high socio-economic status and intellectual competence. This is exactly the situation informants are found in when they face a fieldworker who is viewed as a dominant figure in a powerful position (i.e., the observer's paradox (Labov, 1972: 209)). The fieldworker's task in such situations is to employ a method of data gathering which counteracts the tendencies mentioned above. Although these requirements seem contradictory they are actually reconcilable, as will be shown in the following sections.

Our review of the literature, in search for the best methodology to follow, yielded contradictory results. On one hand there are reports that non-scripted speech produces ecologically valid *and* token-rich results (e.g. Dilley et al, 1996; Ostendorf et al, 1995). On the other, that spontaneous speech corpora cannot provide enough tokens of the target phenomenon (e.g. Schötz et al, 2012). The recent increased interest in intonational dialectal research has generated a demand for methodology comparisons of this sort and consequently a number of papers have emerged comparing different methodological approaches (among others, see Warren 2005, Post and Nolan 2012; Prieto, 2012). Most of these comparisons have been conducted indirectly, that is, through drawing on secondary sources of papers which have employed one or another methodology. Such comparisons cannot provide reliable results because the objects of comparison are disparate, were not designed for a comparison and are thus non-quantifiable.

The novelty of our approach lies in the three main goals of the paper: First, we planned a direct comparison among a set of experiments, each testing a different method for eliciting linguistically relevant dialectal speech instead of relying on post-hoc comparisons of other papers. This was achieved by designing four different elicitation methods which varied in the amount of freedom allowed to the speaker. Second, our experimental design allowed for quantitative analyses to compare the different methods. Third we strived to ensure dialectal authenticity of our participants' speech, that is to avoid their convergence to the dominant standard accent (in our case Standard Modern Greek), by employing a native speaker of the dialect as the experimenter and delivering all aural prompts in the local dialect.

1.1 Elicitation methods in intonational dialectal research

Given the inherent difficulty of eliciting specific intonational contours as well as the lack of research specifically evaluating methodological approaches for tackling intonational questions, it comes as no surprise that a great number of studies investigating dialectal speech have relied on reading tasks involving sentences or paragraphs (e.g. Arvaniti, 1998; Arvaniti and Garding, 2007; Arvaniti, 2007; Atterer and Ladd, 2004; Avanzi et al, 2012; Ladd et al, 2009; Dalton and Ni Chasaide, 2005; Elordieta, 2005; Leemann et al, 2012; Schepman et al; 2006). A few studies have also used single word utterances (e.g. Papazachariou, 2004).

At the exact opposite end of the continuum in terms of speaker freedom lies the use of spontaneous speech materials. Studies using solely spontaneous speech are less common in

dialectal variation of intonation. For example, Archakis et al, (2009) elicited spontaneous speech from a dialectal speaker of Lesvos (Greece) who narrated stories from her life; Leemann and Siebenhaar (2008) examined tonal alignment in Swiss German dialects by conducting interviews that resulted in a corpus of spontaneous speech; Leemann (2009) also used spontaneous speech from interviews held with students answering questions regarding their future plans.

The prevalent difficulty in using a corpus of spontaneous speech is ensuring that enough instances of the linguistic variable under investigation arise to form a balanced experimental design. In an investigation of twelve Swedish dialects, Schötz et al (2012) initially analyzed pre-existing material from three databases of Swedish dialects whose material was recorded as telephone conversations, as well as read and spontaneous speech. These databases, although extremely large, did not contain sufficient material to form a balanced design, clearly evincing the difficulty in using spontaneous speech.

Several studies have used a combination of read material and conversational speech. This combination is also the path suggested by Himmelmann and Ladd (2008), who propose that experimenters start with conversational speech, identify the linguistic issues of interest and then move on to experimental approaches to data elicitation. For example, the IViE corpus for English dialects has relied on a combination of elicitation methods, i.e., read sentences and read passages, as well as story-telling from memory assisted with pictures (Grabe, 2004). Schötz et al (2012) used materials from three databases which comprised recordings of telephone conversations, as well as read and spontaneous speech (see Bruce et al, 1999); one of those databases was also used by Bruce et al (2007), indicating the importance of having available dialectal databases for comparable research, while Ulbrich (2002) again combined read speech with recordings of news broadcasts and fairytale descriptions and Grice et al (2005) combined read, spontaneous and semi-spontaneous speech to analyse four Italian varieties.

While reading tasks allow for greater control over the produced data, researchers often try to find more sophisticated ways of extracting their material to ensure ecological validity. One of the most prominent techniques used is that of a Map Task, whereby pairs of participants are provided with maps and in cooperation try to navigate from one point to another. The maps are not identical and the speakers are told this explicitly at the beginning of their first session. It is, however, up to them to discover how the two maps differ without being allowed to look at the other person's map. The ensuing discussion between the participants results in elicitation of a number of questions, negotiations, instruction giving and repeated use of the place names on the map (e.g. Anderson et al 1991 for Scottish English; Mayo et al 1997 for the intonation of Glasgow English; Ortega 2002 for yes/no questions in Glasgow English). The basic idea of a Map Task has been expanded in other game tasks, such as the one used by Warren et al (2003) and Schafer et al (2004) who asked participants to use a map for the designated Driver to direct the Slider to the right destination. Other possible methods include question answering, role playing, description (possibly to another participant) of videos or pictures the participant has just been exposed to (Himmelmann and Ladd 2008, see also Swerts and Collier 1992 on the use of spatial gridline networks, as well as Edlund et al 2010 on a multimodal database of Swedish). Finally, a few studies have conducted perception experiments on intonational variation (e.g. Peters et al 2002, Ulbrich 2002, Fournier et al 2006).

In this paper we focus on the elicitation of intonation from dialectal speech. Prieto (2012) and Post & Nolan (2012) provide overviews on the experimental methods used overall in prosody research. In these two papers again, however, it becomes clear that no research has specifically set out to provide a comparison of how successful different methods can be in eliciting dialectal

speech with sufficiently frequent instances of the relevant prosodic events to warrant a quantitative analysis of the dataset. The current paper presents such an effort, i.e., we focus on production and show how the combination of conversational and elicited speech manages to extract dialectal data, while at the same time tapping at the specific prosodic phenomenon under investigation. Importantly, by providing a quantitative comparison, we showcase that a combination of the two methods (i.e., conversational and elicited speech) is to be preferred to either one of the two in dialectal variation. We discuss methodological considerations on the elicitation of linguistically relevant materials from native speakers of any given dialect. While we focus on the elicitation of specific intonation and segmental patterns, we expect our results to be relevant for a variety of researchers from many different fields of linguistics.

1.2 Overview of past studies on intonational variation

The investigation of dialectal variation is of great importance in phonology, as exemplified by Foulkes and Doherty's (1999) quote (as cited in Grabe, 2004:9):

- (1) "Understanding the nature and role of variability would [...] appear to be a highly productive route towards constructing an adequate model of phonological knowledge".

With respect to intonation specifically, the last fifteen years have seen a great rise of interest in intonational variation, an area which, until recently, was neglected. As our overview will show, for many languages there is a deficit of research on trans-dialectal intonational differences, which to some extent is due to lack of good methodological tools.

This emerging interest in dialectal variation is clear cross-linguistically, providing insights into two distinct but inter-related directions; first, on a descriptive level to showcase differences in the phonological intonational architecture among different dialects/varieties of a language, and second, on a theoretical level, to use these differences as probes on issues concerning the typology and phonology of intonation. Studies on British English intonation, for example, have given rise to an online corpus called "English Intonation in the British Isles" (IViE) (<http://www.phon.ox.ac.uk/files/apps/IViE/>, Grabe and Post 2002 for a description of the corpus). A number of papers have arisen from the IViE project, some describing differences at the intonational level across urban dialects (Grabe, 2004; Grabe et al, 2007), while others tackling important theoretical issues by means of dialectal variation (Grabe, 2000; Grabe et al, 2004; Fletcher et al, 2005 on the "high rising tune" across four English varieties, Grabe et al 2000 on "truncation" vs. "compression" using four varieties of British English). In addition, dialectal variation of dialectal intonation has appeared in work by Arvaniti and Garding (2007) who investigated the intonational marking on high tunes in two dialects of American English, Arvaniti (2007) on final lowering across British and American English, as well as in work by Mayo et al (1997) who presented a tool for analyzing the intonation of Glasgow English. Theoretical was also the motivation of Ladd et al (2009) who used British English to investigate differences on the alignment of pitch accents across Scottish and British Standard English. Other varieties that have been the matter of investigation from a theoretical viewpoint are Mexican American (Goodwin et al, 2002), African American (Jun and Foreman, 1996), and Irish dialects (Dalton and Ni Chasaide 2005 on pitch accents alignment differences), among others.

Similarly to English, German has received a lot of attention, especially given the wealth of German varieties. Leemann (2009) tested the realization of pitch accents across two Swiss

German dialects, Barker (2005) provided an Autosegmental analysis of Tyrolean German, Leemann and Zuberbühler (2010) examined the intonation of declarative sentences across eight Swiss German varieties, a topic tackled also by Ulbrich (2002) on three different varieties of German, and Atterer and Ladd (2004) tested differences in the acoustic demarcation of pitch accents across northern and southern German. Moreover, Peters et al (2002) offer one of the few perception studies using discrimination tasks to distinguish between Hamburg and Berlin German.

Dutch and its intonational and tonal variation has also seen substantial research from the research team of Radboud University Nijmegen (http://www.ru.nl/gep/projects/intonation_in/), who have developed and presented results on two research projects, namely one on the intonation of *three dialects of Dutch (Zeelandic, Frisian and Low Saxon) as well as two urban dialects (spoken in Amsterdam and Rotterdam)* (e.g. Hanssen et al, 2008a, 2008b; Hanssen et al, 2007; Peters et al, 2007) and one on tone and intonation in Limburgian dialects (e.g. Peters, in press; Fournier et al 2006, 2004).

Various analyses can also be found in a variety of other languages, e.g. the analysis of six Northern Australian varieties by Bishop and Fletcher (2005), an overall description of four Italian dialects by Grice et al (2005), a comparison of pitch alignment across two Basque varieties (Elordieta and Calleja, 2005) and across three Romance varieties (Prieto et al, 2005), an overall description and modeling of three Swedish dialects (Schötz et al, 2012, Bruce et al, 2007), a classification of three French dialects using a variety of prosodic measures (Avanzi et al, 2012) among many others. The issue of pitch accent alignment, being very interesting theoretically, has also been examined by Ladd et al (2000) and Schepman et al (2006) on Dutch, however no cross-dialectal information was used in those papers.

From the above it is clear that some languages and varieties have attracted significant interest and have resulted in a good understanding of their intonational differences, while others still lack substantial research. Modern Greek dialectal intonation falls within the second category of languages for which we do not have a good description and understanding of their cross- and intra-dialectal intonational differences. Arvaniti (1998) compared Standard Modern Greek (SMG) and Cypriot Greek (CG) in terms of the phonological status of phrase accents. More recently, Themistocleous (2011 and 2012) investigated differences on the intonational contours expressing information structure between SMG and CG, focusing on the realization of nuclear pitch accents across the two varieties. The dialect of Goumenitsa (Northern Greece), especially the intonational marking of polar questions, has been the matter of detailed descriptive research by Papazachariou and Archakis (2001) and Papazachariou (2004). Even more recently, Tsiplakou et al (2011) and Gryllia et al (2011) examined the production and perception of polar questions and wh-questions intonation across seven dialects of Greek.

It is therefore apparent that intonational variation is an important under-investigated topic. With respect to the current paper, the lack of much dialectal research on intonation forms a gap also in methodological considerations for researchers. The elicitation of dialectal material is a problematic area, since prosodic effects can be influenced by a variety of factors, making it extremely hard to extract information on a specific linguistic variable (Himmelman and Ladd 2008). With respect to Modern Greek, elicitation of dialectally representative intonation is even harder since one has to create appropriate experimental settings to manage to disentangle influences from a prevailing SMG dialect (Archakis et al, 2009).

In what follows, we first present the theoretical research questions that the compared methods were designed to explore (section 2.1) as well as our motivation for the designed

comparison. Section 3 presents the four different methods in detail and section 4 presents the results. Section 5 concludes the paper.

2 Our project

This paper is part of a project examining the interaction between the segmental and suprasegmental levels in a dialect of SMG. In the phonological theory of intonation, according to the *segmental anchoring hypothesis* the targets of pitch accents are expected to align with specific targets in the segmental string. As can be seen in Figure 1, a pitch accent α comprising L and H targets is associated with a stressed syllable (in this instance the first syllable of a trisyllabic word). According to the segmental anchoring hypothesis, each of the separate tonal targets will manifest itself on specific landmarks with respect to the segmental string. For example, SMG marks pre-nuclear items in declarative sentences using the pitch accent L^*+H^1 . It is expected that the L tone appears near the onset of the stressed syllable's consonant, and the H is consistently aligned just after the onset of the first post-accentual vowel, as shown by the dotted lines in Figure 1 (Arvaniti et al, 1998).

However, it is open to debate whether segmental anchoring represents a *phonological* secondary association of each of the tones with edges of the segmental string (e.g. edges of syllables/segments), as proposed in some research (e.g. Ladd et al, 2000; Prieto et al, 2005), or whether it is *phonetic* in nature, specified at the phonetic level by dialect-specific rules (Arvaniti and Garding, 2007; Ladd et al, 2009). An interesting test-bed for this issue is what happens when the putative anchor point for the H target gets deleted, as shown in Figure 1. The dialect of Epeirus in Northern Greece is known to have the phonological phenomenon of unstressed high vowel deletion, providing the perfect conditions to test our linguistically motivated question. If we hypothesize a similar alignment of L^*+H targets in Epirus as in SMG, it becomes clear that high-vowel deletion can deprive the H target of its anchor point (i.e., /ma'loni/ → [ma'lon] 'scolds'; Topintzi and Baltazani, 2012).

In order to test this hypothesis, our experimental design needs to elicit instances of L^*+H pitch accents associated with words that have undergone high-vowel deletion, forming a rather demanding experimental target. Moreover, a number of experimental prerequisites are needed in order for the phenomenon to be analysed appropriately. First and foremost, the resulting database from dialectal recordings needs to include as many words as possible with high vowel deletion. Appropriate segmental make-up is also necessary in order to be able to "see" the F₀; fundamental frequency contours are only available in voiced segments, given that they have vocal fold vibration. In order to be able to analyse the pitch contour of the specific word with vocalic deletion, most (if not all) segments need to be voiced. Moreover, the position of the word in the utterance is of utmost importance, since the L^*+H pitch accent is found solely in pre-nuclear position. A variety of other considerations include the need for appropriate metrical structure to avoid tonal crowding (Arvaniti et al, 1998), as well as avoiding pauses, hesitations, too many continuation rises or list intonation, all of which exclude the possibility of eliciting the pitch accent under investigation.

¹ This analysis relies on the Autosegmental-Metrical Framework (see Ladd 2008 for an overview), and particularly on Arvaniti & Baltazani's (2005) analysis of SMG.

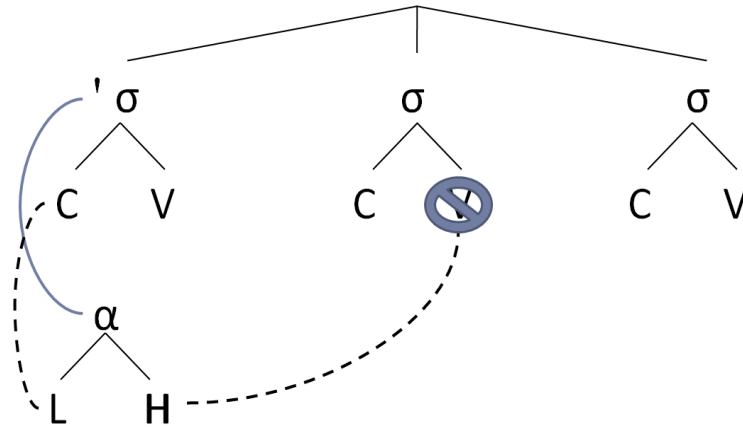


Figure 1: Representation of a possible analysis of phonological secondary association of tonal targets (taken from Ladd, 2008). The pitch accent is shown to be primarily associated as a whole to the stressed syllable, while the component tones are linked with the dotted lines to specific landmarks with respect to the segmental string.

It becomes very clear from the above discussion that the elicited material needs to fulfill a number of requirements that are unlikely to be met in spontaneous unscripted speech. In the next section we summarize the goals for the methodological design of our experiments before we proceed to the detailed presentation of each separate method and its results in sections 3 and 4.

2.1 Goal of the present paper

Combining an ecologically valid experiment with one that manages to elicit dialectal speech that does not resemble the standard dialect and one that elicits frequent samples of the investigated linguistic events is not an easy task. While this issue is commonly stated as a problem in prosodic and dialectal research, the question of how methodologies can be combined to work best in controlled experiments has not been tested yet. For the purposes of our greater research goal, therefore, we initially designed four different experimental methodologies that are directly comparable and can offer quantitative data on the success of each experiment.

Our main proposal is that a combination of conversational and controlled experiments offers the best insights into prosodic phenomena, in the sense that conversational, ecologically valid speech offers the baseline as to how the dialect really behaves. Even if only a few instances of the phenomenon arise, they still provide a qualitative description of what the experimenter is to expect, and form the basis on which subsequent tightly controlled speech tasks can be constructed. Then, by means of specific speech tasks the experimenter can elicit several instances of the particular prosodic phenomenon for quantitative analyses.

Importantly, by extension, the goal of this paper is to help researchers working on dialectal speech overcome their hesitation in using elicited speech, when that is the only solution for extracting relevant specific linguistic material. While our focus is on prosodic and segmental phenomena, we believe that the overall description of the methodology in eliciting the dataset will be helpful for a variety of researchers from different fields within linguistics.

3 Method

Four different elicitation methods were designed and tested. In order for the goals of the experiments to be met, and for the methods to be comparable, several requirements were taken into consideration.

The first goal of the experiment was to see how possible it is to elicit dialectal speech using speech directed tasks instead of free conversational speech. This was achieved by ranging the amount of freedom given to the participants in producing their own sentences. In this Section we present Methods 1 to 4, which range in terms of freedom from the one with the most to the one with the least freedom.

Second, following the notion of ecological validity and wanting to have a baseline as to what dialectal variation really looks like in the dialect of Epirus we made sure that the first method used was that of free conversational speech. However, to ensure elicitation of dialectally representative speech samples, and avoid samples that have undergone accommodation to SMG, the experiment was always run by a native speaker of the dialect. Therefore, across methods we will see that all conversations/questions are directed by a native speaker of the Epirus dialect. Moreover, all experiments were run at the participants' homes (except in some cases when elderly participants were frequenting nearby coffee shops to meet with friends) to ensure a friendly environment.

In methods that diverge significantly in terms of ecological validity, and are highly directed, two considerations were met: first, participants were never asked to read sentences off a paper/screen, but instead their produced utterances were responses to aural and visual stimuli. Second, participants were always given trial sessions to ensure they understood what was being asked of them.

One final goal of the paper was to be able to offer a quantitative comparison across methods in terms of their success. We accomplished that in three ways: first, the same amount of speech was analysed from each speaker, second we measured whether dialectal speech had been elicited by counting the frequency of occurrence of phonological phenomena known to occur in the dialect of Epirus, and third we measured how often the prosodic and segmental variable under investigation appeared in each method. The second measurement allowed insights into whether dialectal speech was elicited, and the third answered the question as to whether the method was successful in terms of its linguistic merit, i.e., whether the linguistic goal of the experiment was met.

3.1 Method 1 – Conversational speech

Conversational speech provides the most naturally uttered and representative dialectal materials and can thus act as a baseline as to how often the phenomenon under investigation occurs in everyday speech.

Two female speakers from Northern Greece (aged 70-80 years old) were recorded at their homes during a single visit to Northern Greece. Interviews were run by a native speaker of the dialect and speakers were asked to describe stories from their life during wartime in Greece. Recordings were made using a Marrantz PMD660 recorder. Each conversation lasted approximately 30 minutes. The experimenter directed the conversation with questions regarding the participants' background, especially their experiences from World War II. The experimenter's participation was minimal, to allow participants to speak freely and for most of

the recording time. Very little overlapping speech ensued, which was not analysed. The quality of the recording was useable for phonetic analysis, in the sense that no background noise was recorded and the participants' amplitude was high enough.

3.2 Method 2 – Describe the differences

In method 2 we aimed again for free speech, but with some direction as to the possible words that participants would use. Participants were shown pictures on a computer screen using a powerpoint presentation and were asked to describe their differences (see Figure 2). The pictures always depicted items that would elicit words with high vowel deletion (e.g. in Figure 3 the word γουρούνι /yu'runi/ 'pig', pronounced /yu'run/ or even /'ɣrun/ in this dialect). Compared to Method 1, this is a more directed speech task, but with substantial amount of freedom during speech production.

The experiment was run by a native speaker of the dialect. Three female native speakers of Northern Greece (area of Ioannina) were recorded (aged 70-80 years old). Recordings took place at the participants' homes using a laptop for the powerpoint presentation and a Marrantz PMD660 for the recording. As for Method 1, no background noise was captured, and the amplitude of the speakers' productions was properly adjusted.

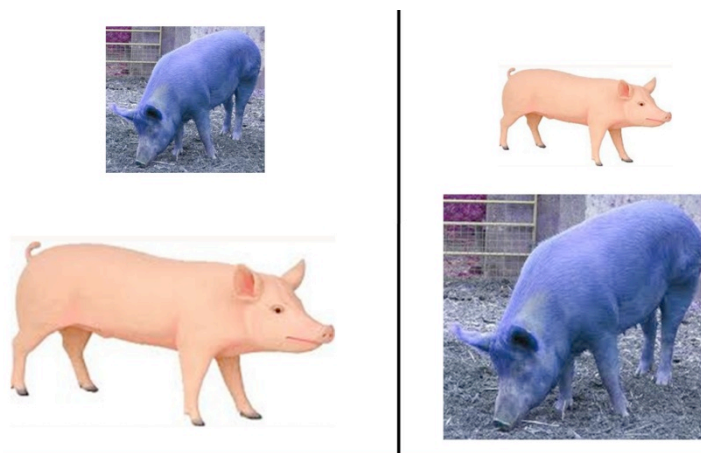


Figure 2: Example of pictures shown to participants to describe the differences. See text for details.

3.3 Method 3 – Two-action naming

Two-action naming is a speech task directly constructed to elicit as many appropriate² sentences as possible in a short amount of time, and therefore represents the method with the least amount of freedom thus far. Participants are placed in front of a computer and see powerpoint slides on the screen. Each slide contains two pictures – most commonly two people performing similar tasks, as for example in Figure 3. They were then prompted with a question of the type “In the picture we can see Eleni and Sula. What is Eleni and what is Sula picking?” (/sti fotoyra'fia

² *Appropriate* in the sense described at the end of Section 2.

'vlepume tin e'leni ce ti 'sula ti ma'zevi i e'leni ce ti i 'sula/ Στη φωτογραφία βλέπουμε την Ελένη και τη Σούλα. Τι μαζεύει η Ελένη και τι η Σούλα;). For each of the constructed sentences one of the names and the verb ended as often as possible in a high vowel, and contained appropriate metrical and segmental material³ to ensure frequent production of words with L*+H and high vowel deletion. The format of the questions was such to ensure that the resulting word order in the answers with place at least one of the names and, hopefully, the verb in pre-nuclear positions carrying the L*+H pitch accent. Moreover, the use of two names at the same time ensured that all names and verbs from the question would be used by the participant (given all these considerations, the answers' format could not differ significantly from “Eleni is picking lettuces and Sula lavender”, /i e'leni ma'zevi ma'ruʎa ce i 'sula le'vades/ Η Ελένη μαζεύει μαρούλια και η Σούλα λεβάντες).

This method was run in the same session as *Describe the differences*, therefore the same experimental procedure and participants were used. All questions directed to the participants had been pre-recorded by a native speaker and embedded in the powerpoint presentation for reasons of homogeneity of experimental process across participants.



Figure 3: Example of picture shown to participants during method *Two-actions naming*. See text for details.

3.4 Method 4 – One-action naming

This final method was run as a third experiment, using the experience acquired from all previous experiments. Participants were here allowed the least amount of freedom in their productions. They were shown a picture with one person performing one action, e.g. “What is Yannis using to cut the onion with?” (Με τι κόβει ο Γιάννης το κρεμμύδι;) (Figure 4) and were asked to respond using full sentences, e.g. “Yannis is cutting the onion with the knife”, (Ο Γιάννης κόβει το κρεμμύδι με το μαχαίρι). Participants listened to questions from the powerpoint; these had been pre-recorded by a native speaker of the dialect, as for Methods 2 and 3. The experiment was run using a native speaker of the dialect. All words within the sentence were potential sites of high vowel deletion. Moreover, we asked the participants to answer with a full declarative sentence in

³ Please see section 2 for a description of the requirements for appropriate metrical and segmental material.

the hopes of eliciting as many instances of the L*+H pitch accent as possible. Finally, the segmental and metrical structure of the words selected again accounted for metrical and voicing considerations.



Figure 4: Example of picture shown to participants for Method *One-action naming*. See text for details.

Three female native speakers of Northern Greek (area of Ioannina, 45-70 years old) were recorded producing sentences potentially involving deletion of high vowel, matched with sentences without deletion (e.g. “What are the two Yannis using to cut the onions?” Με τι κόβουν οι Γιάννηδες τα κρεμμύδια;). Recordings were made on a laptop using a Blue Yeti microphone.

3.5 Evaluation and comparison of methods

As mentioned at the beginning of this section, in order to ensure that comparisons across methods are experimentally sound, approximately the same amount of raw speech was analysed across methods (≈ 130 seconds of speech for each method). For conversational speech, while 130 seconds were analysed, these did not come directly from the first 130 seconds of the recording, but rather a selection was made from the whole dataset to ensure the speech excerpts contained as much as possible full sentences.

Unfortunately, it is not always easy to segment all parts of the produced utterances. Highly co-articulated passages, background noise, overlapping speech, coughs etc. necessitate that we discard several chunks of the conversation. For this reason, when reporting frequencies of how often a phenomenon applied within each method and the frequency of sites where it is expected to occur, these only apply to segments of speech that were analysable.

In order to measure the success of each method two questions were asked:

1. Have we managed to extract representative dialectal speech?
2. Have we managed to extract the needed pitch accents in deletion (L*+Hdel) and non-deletion (L*+H) environments?

To answer the first question we measured the frequency of appearance of three representative dialectal phenomena in each method. Northern Greek dialects are well-known to undergo (i) unstressed high-vowel deletion, e.g. /ðu'levi/ → [ðu'lev] 'he works', (ii) unstressed mid-vowel raising, e.g. /po'ta/ → [pu'ta] 'drinks', (iii) stressed mid-vowel diphthongization, e.g. \tsu'reci\ → [tsu'rec] 'brioche'.

These phenomena are quite common in this dialect but their application is not categorical. It is possible therefore that the right context for the application of these rules is present but nevertheless the rule does not apply. The exact details for the application of this rule have not so far been investigated to our knowledge. Our analysis accounts for that by performing two separate measurements; first, we counted the *potential* sites of application of these rules (that is, number of words where the phenomenon is expected to occur) and then we counted the actual number of tokens for which the rules did apply. These are called respectively *potential sites of application* and *phenomenon applied* in the Results section, and are compared across methods. For example, if a word like \pezi\ παίζει 'plays' were produced as ['pez] we would count one potential site of stressed-mid vowel diphthongization and one potential site for unstressed high-vowel deletion, with only the high-vowel deletion having actually applied.

To answer the second question we simply measured the frequency of appearance of L*+H and L*+Hdel instances in each method.

4 Results

4.1 Method 1 – *Conversational Speech*

This method proved the least successful of all. While unstressed /i/ deletion occurred 50% of the times, the potential sites of application are not that common. Similarly, potential sites for diphthongization are quite rare, while mid-vowel raising was more often. However, the rate of occurrence of each of the phenomena is quite high, with all phenomena appearing quite often, with the exception of /e/ diphthongization and raising.

The scarcity of occurrence of the vocalic phenomena makes it clear that we would need substantial amounts of conversational speech to be able to elicit enough material for an experimental analysis of L*+H pitch accent co-occurring with vocalic deletion. Moreover, even with a number of tokens big enough to allow a quantitative analysis, the context of tokens in spontaneous speech varies to a great extent and therefore it is difficult to extract reliable inferences, since parameters out of our control might be responsible for the experimental outcome. The success rate in terms of how often the phenomenon applied, on the other hand, clearly shows that dialectal speech has been elicited.

With respect to the second means of testing the success of the method, the pitch accents under investigation were extremely rare and sparse; only six instances of L*+H and two of L*+Hdel accents were identified in conditions allowing intonational analysis, clearly evincing the inadequacy of the current method for a quantitative analysis.

Phenomenon	Potential sites of application	Application of phenomenon	Percent
i unstressed (deletion)	20	10	50%
u unstressed (deletion)	5	2	40%
e stressed (diphthongization)	21	3	14.3%
o stressed (diphthongization)	19	8	42.1%
e unstressed (raising)	52	18	34.6%
o unstressed (raising)	32	15	46.9%

Table 1: Frequency of occurrence of each phonological process occurring in the dataset produced for Method 1- *Conversational Speech*. The second column indicates how many times the phenomenon could appear in the dataset, the third how many of those it actually happened, and the final column shows this relationship in percentage.

The reason conversational speech was not successful in extracting the specific pitch accents lies in the inherent freedom that comes with this method. Almost all experimental pre-requisites for the extraction of pre-nuclear declaratives are not met. While the potential for eliciting L*+H instances was common, in most cases the necessary criteria to see an analysable representation of the accent were not met. The most prevalent shortcoming of this method is that there were not always at least two unstressed syllables intervening between two stressed ones and therefore tonal crowding was not avoided. Furthermore, among other problems that prevented using some of the produced utterances, we note in particular the following: participants would often insert pauses giving rise to continuation rises, place early focus, or use intonation contours that are used in narration⁴ instead of conversational register (participants were asked to narrate their lives during war occupation).

4.2 Method 2 - *Describe the differences*

Table 2 shows the results for method *Describe the differences*. Overall, this method was more successful than conversational speech in eliciting both potential sites of phonological phenomena as well as actual raw instances mainly of high vowel deletion, but less successful in mid vowel raising. Percentage-wise, the two methods do not differ significantly, showing that the method *Describe the differences* resembles in terms of dialectal speech that of conversational speech.

This method was more successful than the previous one in eliciting the prosodic and segmental make-up under investigation since it exercised more control over the speaker utterances. All in all, 13 instances of L*+H were identified but only one L*+Hdel instance,

⁴ This is a register we encountered quite often with dialectal speakers, which deviates considerably in melodic patterns from conversational speech. To our knowledge no study has been published on the intonation of this register, which is an extremely interesting research topic.

clearly indicating that the method is not successful in eliciting the needed combination between the prosodic and segmental levels.

Phenomenon	Potential sites of application	Application of phenomenon	Percentage
i unstressed (deletion)	45	20	44.4%
u unstressed (deletion)	13	3	23%
e stressed (diphthongization)	28	10	35.7%
o stressed (diphthongization)	17	7	41.2%
e unstressed (raising)	23	3	13%
o unstressed (raising)	30	12	40%

Table 2: Frequency of occurrence of each phonological process occurring in the dataset produced for Method 2 –*Describe the differences*.

4.3 Method 3 – *Two-actions naming*

As can be seen in Table 3, *Two-actions naming* gave rise to more sites for potential application of vocalic phenomena than in conversational speech, and also more than *Describe the differences*. The three methods do not seem to differ significantly with respect to the actual percent of how often each vocalic phenomenon applied. This indicates that the last two methods have managed to elicit dialectal speech, especially when compared to conversational speech, which theoretically resembles the most free and dialectally successful method. Still, the method with the least amount of freedom, i.e., *Two-actions naming*, gave rise to most instances of high-vowel deletion in particular, and other vocalic phenomena in general (a needed prerequisite for quantitative analyses).

Phenomenon	Potential sites of application	Application of phenomenon	Percentage
i unstressed (deletion)	87	47	58%
u unstressed (deletion)	6	3	50%
e stressed (diphthongization)	23	5	21.7%
o stressed (diphthongization)	58	23	39.7%
e unstressed (raising)	30	2	6.6%
o unstressed (raising)	27	7	25.9%

Table 3: Frequency of occurrence of each phonological process occurring in the dataset produced for Method 3 – *Two-actions naming*.

Only five instances of L*+H pitch accents were identified in this method, but 14 instances of L*+Hdel. This renders this method marginally better than *Describe the differences*. If we compare *Describe the differences* and *Two action naming* we can see that exercising tighter control over the produced sentences produced the desired L*+Hdel pitch accents.

We should note furthermore, that sentences in Method 3 were designed to contain words that would give rise to vocalic deletion, but no matched sentences without deletion were designed. Given this fact, it was no surprise that L*+H on its own did not occur as often. Therefore, *Two-actions naming* was successful in extracting the material under investigation, but its design lacked the baseline conditions to compare against, something that we tried to rectify in the fourth experiment. In sum, the freedom of allowing speakers to describe pictures using their own words managed to extract the baseline conditions, but not the intended pitch accents.

There were two more problematic issues with the *Two-actions naming* process. Even though the questions posed to the participants' only involved the description of two people performing two actions, older generations often had problems remembering both names mentioned in each slide. Moreover, in cases where peculiar objects were depicted (e.g. lavender) participants often became stressed to produce the correct answer. All these were rectified in the following method.

4.4 Method 4 – *One-action naming*

In *One-action naming* we took advantage of the experience gathered from all previous experiments. Knowing that most instances of vocalic deletion and most L*+Hdel instances arose in the most tightly controlled method thus far, we decided to elaborate on this method, add baseline conditions to elicit L*+H pitch accents, and use sentences that do not require too much memory load on the part of the participants. Therefore, sentences only involved one – not two – actions and all words depicted everyday objects and actions, to avoid causing performance anxiety.

As seen in Table 4, this was a particularly successful method in eliciting potential sites of application of vocalic phenomena, especially high-vowel deletion. This is to be expected, since the segmental make-up of the words used was specifically designed to elicit those. Importantly,

in terms of frequency of occurrence of vocalic phenomena, no great differences can be seen from the previous methods, indicating that dialectal speech has again been elicited.

Phenomenon	Potential sites of application	Application of phenomenon	Percent
i unstressed (deletion)	143	81	56.5%
u unstressed (deletion)	31	8	25.8%
e stressed (diphthongization)	43	2	4.7%
o stressed (diphthongization)	31	7	22.6%
e unstressed (raising)	51	20	39.2%
o unstressed (raising)	69	35	50.7%

Table 4: Frequency of occurrence of each phonological process occurring in the dataset produced for Method 4 – *One-action naming*.

This method was also highly successful in eliciting the required intonational patterns; 14 instances of L*+H were extracted and 27 instances of L*+Hdel, making it the most successful method of all.

4.5 Overall comparison

Overall, all methods were successful in extracting vocalic phenomena, some more so than others. Figure 6 shows a comparison of the number of times (in raw numbers) each of the three vocalic phenomena appeared in each method. The most successful method in eliciting vocalic phenomena was *One-action naming*, with high success in eliciting unstressed high-vowel deletion (i.e., the phonological phenomenon necessary for the linguistic purposes of our experiment). This was expected, since the words in this method were especially designed to elicit the phenomenon. With respect to the remaining processes, *One-action naming* performed equally well (if not better, in the case of /o/ raising) as the other methods. The second more successful method was *Two-action naming*, which again elicited a high number of unstressed high-vowel deletions, for the same reason as *One-action naming*. Similarly, the method was quite successful at eliciting all other phenomena.

A clearer picture is seen with respect to across-methods comparisons when looking at percentages, that is how often a phenomenon applied with respect to how many potential sites of application arose in the corpus. Figure 7 shows that all methods were successful in often extracting vocalic phenomena, on the basis of which we can extrapolate that they were also successful in eliciting dialectal speech. Importantly, when compared to conversational speech, that is the most free version of speech out of the four methods, the highly controlled tasks *One-* and *Two-actions naming* performed equally well.

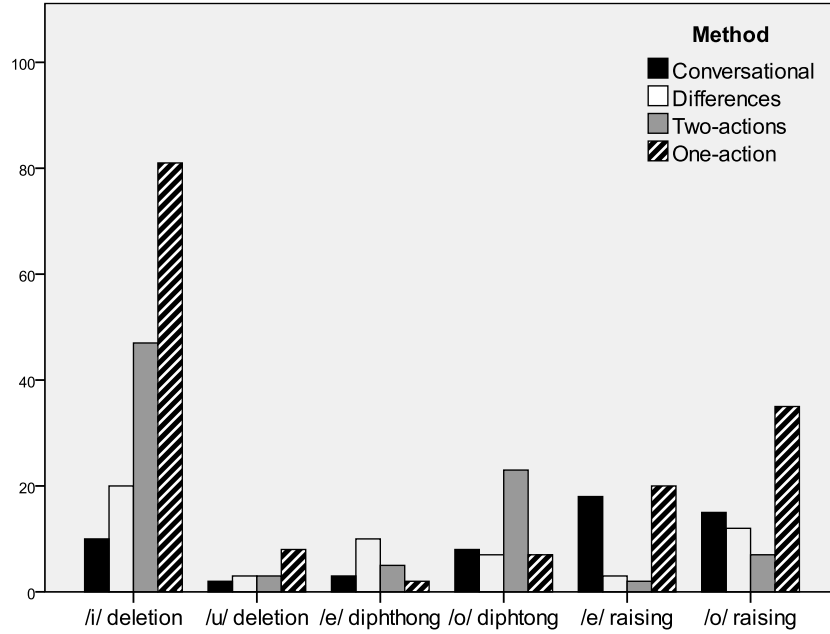


Figure 5: Comparison of frequency of occurrence of each vocalic phenomenon across methods (expressed in number of times of appearance).

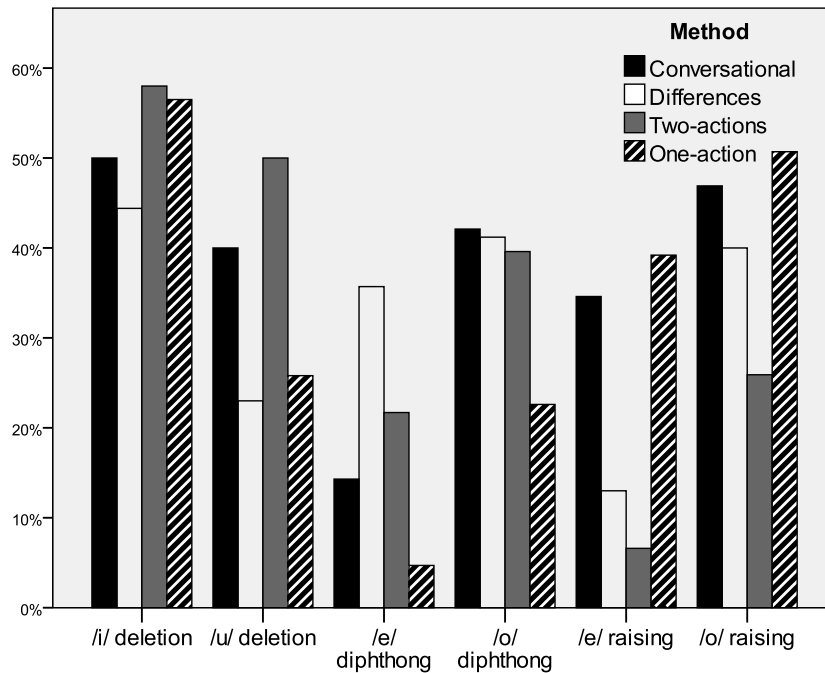


Figure 6: Comparison of frequency of occurrence of each vocalic phenomenon across methods (expressed in percent of times of appearance in the corpus).

An overall comparison in eliciting the pitch accent under investigation across methods is shown in Figure 8, where the clear success of *One-action naming* is seen in both L*+H and L*+Hdel.

This is the only method that was successful in eliciting both the baseline and target pitch accents, followed by *Two-actions naming*, where baseline productions were not specifically included to begin with. From this figure it is clear that conversational speech, while representative dialectally, is not appealing as a means of performing larger scale linguistic analyses.

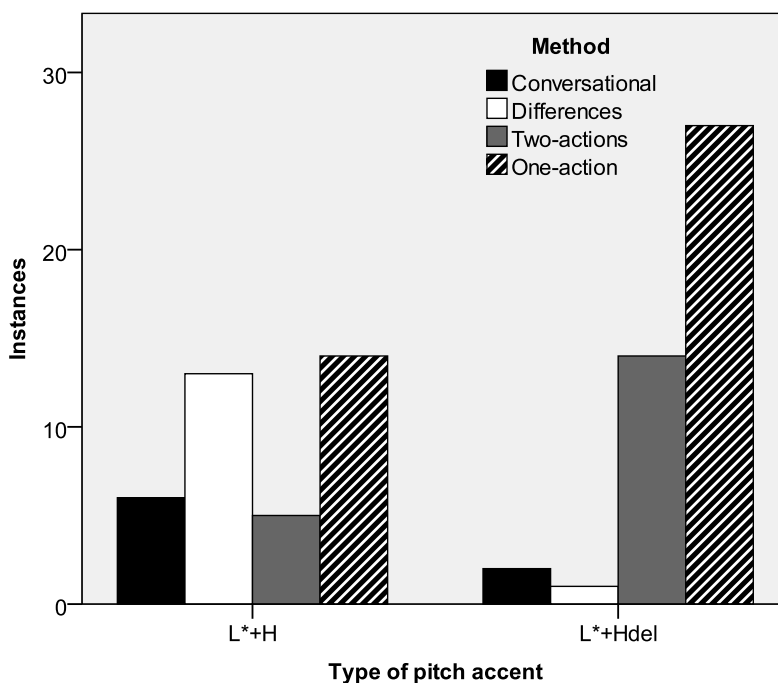


Figure 7: Comparison of frequency of occurrence of each pitch accent across methods (expressed in number of times of appearance).

5 Discussion

The most interesting result emerging from our investigation is that it is possible to elicit dialectally authentic material in a controlled elicitation experiment. In this way we can ensure elicitation of sufficient instances of the phenomenon under investigation to allow for a quantitative analysis of the results and to support strong claims about them. Moreover, without loss of authenticity of speech, we can control critical aspects of both linguistic and extralinguistic context, which cannot be controlled for in spontaneous speech corpora.

We propose that researchers performing linguistic fieldwork of a dialect, who are interested in unveiling specific linguistic phenomena, and therefore require very tightly manipulated experiments, should use a combination of conversational speech and of tightly-controlled speech tasks directed to elicit the specific phenomenon. The merit of using such a combination stems from the fact that conversational speech can act as a baseline as to whether/how often the phenomenon appears in everyday speech, as well as a baseline of each speaker's dialectal repertoire. Following an analysis of this type of data allows the researcher to understand what to expect from the specific dialect. The use of the tightly controlled experiment, on the other hand, allows for concrete control over the produced sentences, once the presence and nature of the

phenomenon has been established. Finally, avoiding the use of reading tasks makes sure that no accommodation to a more formal, “educated” or “standard” production surfaces.

However, there are a few considerations to be taken in mind, and which we suggest as an obligatory part of the protocol. Firstly, pilot testing with a subset of participants is important, not only to see if the experimental procedure is working, but also in order to gain insights into the community visited, such as which tasks they are prepared to perform, how they prefer the experimental procedure to take place, which materials are understandable and useable in everyday interactions, among many other decisions. Piloting the experiment might take more than one visit, since tweaking of the procedure might be considered necessary.

As it turned out in our fieldwork, elderly speakers were not able to carry out a controlled experiment for more than twenty minutes at a time. Younger generations too, however, were not comfortable in continuing with a tightly controlled experiment for more than approximately thirty minutes, especially if that included repetitions of picture tasks. Caution should be taken, therefore, as to how long a controlled experiment can last, and it is proposed that within the pilot protocol researchers include a provision for testing that.

Finally, an important consideration refers to the phonological phenomena we relied upon as markers of dialectal speech. The frequency of appearance of the vocalic phenomena in conversational speech was not as common as one would expect on the basis of traditional descriptions of the dialect. This clearly suggests that a synchronic phonological and acoustic analysis of vocalic phenomena of Northern Greece is in order. Importantly, however, in terms of methodology, it becomes clear that by piloting an experiment researchers are able to pre-test the synchronic frequency of phenomena they expect to use as baselines/potential markers of dialectal speech.

The findings of this paper can hopefully extend to other fields of linguistics, apart from phonetics and phonology. While it is obvious that each research project from each different field has its own research agenda, the common denominator across projects in linguistics remains that researchers are often interested in investigating *specific* phenomena either for syntactic, pragmatic, phonetic, phonological theoretical reasons. For researchers who are not just providing a description of the dialect, but who are also using the phenomenon of a dialect to investigate a deeper theoretical question and are in need of experimental data to allow statistical comparisons, we believe that this paper provides a liberating way forward, since it clearly showcases that conversational and controlled experiments can be similar in terms of extracting representative dialectal speech, and both can therefore be utilized in linguistic research.

References

- Anderson, Anne H., Bader, Miles, Bard, Ellen G., Boyle, Elizabeth, Doherty, Gwyneth, Garrod, Simon, Isard, Stephen, Kowtko, Jacqueline, McAllister, Jan, Miller, Jim, Sotillo, Catherine, Thompson, Henry S. and Regina Weinert. 1991. The HCRC Map Task Corpus. *Language and Speech* 34: 351–66.
- Archakis, Argiris, Lampropoulou, Sofia, and Dimitris Papazachariou. 2009. ‘As for the dialect, I will now tell you how we spoke at home’: On the performance of dialectal talk. *Journal of Greek Linguistics* 9(1): 5–33.

- Arvaniti, Amalia and Gina Garding. 2007. Dialectal variation in the rising accents of American English. In *Papers in Laboratory Phonology 9*, eds. J. Cole & J. H. Hualde, 547–576. Berlin, New York: Mouton de Gruyter.
- Arvaniti, Amalia and Mary Baltazani. 2005. Intonational analysis and prosodic annotation of Greek spoken corpora. In *Prosodic Typology: The Phonology of Intonation and Phrasing*, ed. Sun-Ah, 84–117. Jun Oxford University Press.
- Arvaniti, Amalia, Ladd Robert D. and Ineke Mennen. 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26: 3–25.
- Arvaniti, Amalia. 1998. Phrase accents revisited: comparative evidence from Standard and Cypriot Greek. *Proceedings of the 5th International Conference on Spoken Language Processing*, vol. 7: 2883–2886, Sydney.
- Arvaniti, Amalia. 2007. On the presence of final lowering in British and American English. In *Tone and Tunes, vol. 2: Experimental Studies in Word and Sentence Prosody*, eds. C. Gussenhoven and T. Riad, 317–347. Berlin and New York: Mouton de Gruyter.
- Atterer, Michaela and Robert D. Ladd. 2004. On the phonetics and phonology of “segmental anchoring” of F0: evidence from German. *Journal of Phonetics* 32: 177–197.
- Avanzi, Mathieu, Obin, Nicolas, Bardiaux, Alice and Guri Bortal. 2012. Speech Prosody of French Regional Varieties. *Proceedings of Speech Prosody 2012*.
- Barker, Geoffrey. 2005. *Intonation patterns in Tyrolean German: An autosegmental-metrical analysis*. Vol. 57. Peter Lang.
- Beebe, L. and Howard Giles. 1984. Speech accommodation theories: a discussion in terms of second language acquisition. *International Journal of the Sociology of Language* 46: 5–32.
- Bishop, Judith and Janet Fletcher. 2005. Intonation in six dialects of Bininj Gun-wok. In *Prosodic Typology: The Phonology of Intonation and Phrasing*, ed. Sun-Ah Jun, 91–142. Oxford University Press.
- Bruce, Gösta, Elert, Claes-Cristian, Engstrand, Olle, Eriksson, Anders and Pär Wretling. 1999. Database tools for a prosodic analysis of the Swedish dialects. In *Proceedings Fonetik* Vol. 99: 37–40.
- Bruce, Gösta, Granström, Björn and Susanne Schötz. 2007. Simulating intonational varieties of Swedish. *Proceedings of ICPHS XVI*.
- Dalton, Martha and Ailbhe Ni Chasaide. 2005. Tonal alignment in Irish dialects. *Language and Speech*, 48(4): 441–464.
- Dilley, Laura, Shattuck-Hufnagel, Stephanie and Mark Ostendorf. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, 24: 423–444.
- Edlund, Jens, Beskow, Jonas, Elenius, Kjell, Hellmer, Kahl, Strombergsson, Sofia, and David House. 2010. Spontal: A Swedish spontaneous dialogue corpus of audio, video and motion capture. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, eds. N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias, 2992–5. Valetta, Malta.
- Elordieta, Gorka and Nagore Calleja. 2005. Microvariation in accentual alignment in Basque Spanish. *Language and Speech* 48: 397–439.
- Fletcher, J., Grabe, E., and Warren, P. 2005. Intonational variation in four dialects of English: the high rising tune. In *Prosodic typology and transcription - a unified approach*, ed. Sun-Ah Jun, 390–409. Oxford, OUP.
- Foulkes, Paul, and Gerard J. Docherty. 1999. *Urban voices: Accent studies in the British Isles*. A Hodder Arnold Publication.

- Fournier, Rachel, Verhoeven, Jo, Swerts, Marc and Carlos Gussenhoven. 2004. Prosodic and segmental cues to the perception of grammatical number in two Limburgian dialects of Dutch. In *Speech Prosody 2004, International Conference*.
- Fournier, Rachel, Verhoeven, Jo, Swerts, Marc and Carlos Gussenhoven. 2006. Perceiving word prosodic contrasts as a function of sentence prosody in two Dutch Limburgian dialects, *Journal of Phonetics* 34: 29–48.
- Giles, Howard and Philip Smith. 1979. Accommodation Theory: Optimal Levels of Convergence. In *Language and Social Psychology*, eds. Giles, Howard, St. Clair, Robert N. Baltimore: Basil Blackwell.
- Giles, Howard, Bourhis, Richard and Donald M. Taylor. 1977. Toward a theory of language in ethnic group relations. In *Language Ethnicity and Intergroup Relations*, ed. Howard Giles, 307-349. Academic Press, London.
- Giles, Howard, Coupland, Joustine and Nicolas Coupland. 1991. Accommodation Theory: Communication, Context, and Consequence. In Giles, Howard; Coupland, Justine; Coupland, N. *Contexts of Accommodation*. New York, NY: Cambridge University Press.
- Goodwin, Marjorie Harness, Charles Goodwin, and Malcah Yaeger-Dror. 2002. Multi-modality in girls' game disputes. *Journal of pragmatics* 34(10): 1621–1649.
- Grabe, Esther. 2004. Intonational variation in urban dialects of English spoken in the British Isles. In *Regional Variation in Intonation*, eds. Gilles, P. and Peters, 9–31. J. Linguistische Arbeiten, Tuebingen, Niemeyer.
- Grabe, Esther and Brechtje Post. 2002. Intonational Variation in English. In *Proceedings of the Speech Prosody 2002 Conference*, eds. B. Bel and I. Marlin, 343–346. 11-13 April 2002, Aix-en-Provence: Laboratoire Parole et Langage.
- Grabe, Esther, Bretchje, Post, Nolan, Frances and Farrar, K. 2000. Pitch accent realisation in four varieties of British English. *Journal of Phonetics* 28.
- Grabe, Esther, Greg Kochanski, and John Coleman. 2007. Connecting intonation labels to mathematical descriptions of fundamental frequency. *Language and speech* 50(3): 281–310.
- Grice, Martin, D'Imperio, Mariapaola, Savino, Michelina and Cinzia Avesani. 2005. Strategies for intonation labeling across varieties of Italian. In *Prosodic Typology: The Phonology of Intonation and Phrasing*, ed. Sun-Ah Jun, 362–389. Oxford University Press.
- Gryllia, Stella, Arvaniti, Amalia, Baltazani, Mary, Adamou, Evi, Terkourafi, Marina, Vergis, Nikos and Stavroula Tsipakou. 2011. Identification and social evaluation of dialectal Greek intonation. Paper presented at the *10th International Conference on Greek Linguistics*, Democritus University of Thrace, September 1-4, 2011.
- Hanssen, Judith, Peters, Jörg and Carlos Gussenhoven. 2007. Phrase-final pitch accommodation effects in Dutch. *Proceedings of 16th ICPhS*.
- Hanssen, Judith, Peters, Jörg, Fournier, R. and Carlos Gussenhoven. 2008a. Regional variation in the realisation of IP-final intonation contours in the Netherlands. Poster presented at the *Third TIE Conference on Tone and Intonation in Lisbon, September 2008*.
- Hanssen, Judith, Peters, Jörg and Carlos Gussenhoven. 2008b. Prosodic effects of focus in Dutch declaratives. In *Proceedings of speech prosody 2008*.
- Himmelman, P. Nicolaus and Robert D. Ladd. 2008. Prosodic description: An introduction for fieldworkers. *Language Documentation and Conservation*, 2(2): 244–274.
- Jun, Sun-Ah and Foreman, Christina. 1996. Boundary tones and focus realization in African American English intonations. *The Journal of the Acoustical Society of America*, 100, 2826.
- Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania.

- Ladd, D. Robert. 2008. *Intonational phonology*. Cambridge Studies in Linguistics 79. Cambridge: Cambridge University Press, 2nd edition.
- Ladd, D. Robert, Mennen, Ineke and Astrid Schepman. 2000. Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America*, 107: 2685–2696.
- Ladd, D. Robert, Schepman, Astrid, White, Lawrence, Quarmby, L. May and Stackhouse, Rebekah. 2009. Structural and dialectal effects on pitch peak alignment in two varieties of British English. *Journal of Phonetics* 37: 145–161.
- Leemann, Adrian. 2009. Intonational Variation in Swiss German. *Language variation. European perspectives*, 135–143. Amsterdam/Philadelphia: Benjamins.
- Leemann, Adrian and Beat Siebenhaar. 2008. Perception of dialectal prosody. *Interspeech*, 524–527.
- Leemann, Adrian and Lucy Zuberbühler. 2010. Declarative sentence intonation patterns in 8 swiss German dialects. *Interspeech*, 1768–1771.
- Leemann, Adrian, Dellwo, Volker, Kolly, Marie-José and Stephan Schmid. 2012. Rhythmic variability in Swiss German dialects. *Speech Prosody 2012*.
- Mayo, Catherine, Aylett, Matthew and Robert D. Ladd. 1997. Prosodic transcription of Glasgow English: an evaluation study of GlaToBI. In *Intonation: Theory, Models and Applications*.
- Ortega, F. Vizcaino. 2002. A preliminary analysis of yes/no questions in Glasgow English. In *Speech Prosody 2002, International Conference*.
- Ostendorf, Mark, Price P. J. and Stephanie Shattuck-Hufnagel. 1995. The Boston University Radio News Corpus. Technical Report ECS-95-001, Electrical, Computer and Systems Engineering Department, Boston University, Boston, MA.
- Papazachariou, Dimitris. 2004. Intonation variables in Greek polar questions. *Graphematische Systemanalyse als Grundlage der Historischen Prosodieforschung*, 191–217.
- Papazachariou Dimitris and Argyris Archakis. 2001. The dialectal role and discourse function of six semantic-intonation variables in Northern Greece. In *First International Conference on Modern Greek Dialects and Linguistic Theory*, eds. A. Ralli, B. Joseph and M. Janse, 223–236. Patras: Patras University Press.
- Peters, Jörg. in press. A bitonal lexical pitch accent in the Limburgian dialect of Borgloon. In *Tone and intonation*, eds. C. Gussenhoven & T. Riad.
- Peters, Jörg, Gilles, Peter, Auer, Peter and Margret Selting. 2002. Identification of regional varieties by intonational cues. An experimental study of Hamburg and Berlin German. *Language and Speech*, 45:115–39.
- Post, Brechtje and Frances Nolan. 2012. Data analysis for prosodic analysis of continuous speech and dialectal variation. In *The Oxford Handbook of Laboratory Phonology*, eds. A.C. Cohn, C Fougeron, & M. Huffman. Oxford: OUP.
- Prieto, Pilar. 2012. Experimental methods and paradigms for prosodic analysis. In *Handbook in Laboratory Phonology*, eds. Cohn, A. C.; Fougeron, C.; Huffman, M. K, 527–547. Oxford University Press: Oxford.
- Prieto, Pilar, D’Imperio, Mariapaola, and Barbara Gili Fivela. 2005. Pitch accent alignment in Romance: Primary and secondary associations with metrical structure. *Language and Speech* (special issue on Variation in Intonation), ed. P. Warren. 48(4): 359–96.
- Schafer, Amy J., Shari R. Speer and Paul Warren. 2004. Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In *Approaches*

- to *Studying World-Situated Language Use*, eds. J. C. Trueswell and M.K. Tanenhaus, 209–225. Cambridge, MA: MIT Press.
- Schepman, Astrid, Lickley, Robert and Robert D. Ladd. 2006. Effects of vowel length and “right context” on the alignment of Dutch nuclear accents. *Journal of Phonetics*, 34: 1–28.
- Schoetz, Susanne, Bruce, Gosta and Bjorn Granstrom. 2012. Regional varieties of Swedish: Models and synthesis. In *Understanding Prosody: The Role of Context, Function and Communication*, 119–133.
- Swerts, Mark and René Collier. 1992. On the controlled elicitation of spontaneous speech. *Speech Communication* 11: 463–468.
- Themistocleous Charalambos. 2011. Prosodia kai plirophoriaki domi stin Atheniaki kai kypriaki Ellinici (Prosody and Information Structure in Athenian and Cypriot Greek). *PhD Thesis*. University of Athens Greece.
- Themistokleous, Charalambos. 2012. *Nuclear Accents in Athenian and Cypriot Greek*. In *Selected papers of the 10th ICGL*, eds. Z. Gavriilidou, A. Efthymiou, E. Thomadaki & P. Kambakis-Vougioukli, 796–805. Komotini/Greece: Democritus University of Thrace.
- Topintzi, Nina and Mary Baltazani. 2012. The acoustics of high-vowel loss in a Northern Greek Dialect and typological implications. In *Consonant Clusters and Structural Complexity*, eds. P. Hoole, L. Bombien, M. Pouplier, Ch. Mooshammer, and B. Kühnert, , 373–402. Interface Explorations series, de Gruyter.
- Tsiplakou, Stavroula, Arvaniti, Amalia, Baltazani, Mary, Adamou, Evi, Gryllia, Stella, Terkourafi, Marina, Themistocleous, Charalambos, Vergis, Nikos and Spyros Armosti. 2011. *Intonational variation in polar and wh-questions across seven varieties of Greek*. Paper presented at the *6th International Conference on Language Variation in Europe*, Freiburg, Germany.
- Ulbrich, Christiane. 2002. A Comparative Study of Intonation in Three Standard Varieties of German. In *Proceedings of the Speech Prosody 2002 conference*, eds. B. Bel, I. Marlien, 671–674. 11-13 April. Aix-en-Provence: Laboratoire Parole et Langage.
- Warren, Paul. 2005. Issues in the study of intonation in language varieties. *Language and Speech* 48(4): 345–358.
- Warren, Paul, Shari Speer and Amy Schafer. 2003. Wanna contraction and prosodic disambiguation in US and NZ English. *Wellington Working Papers in Linguistics*, 15: 31–49.