John Benjamins Publishing Company

This is a contribution from *Information Structure in Lesser-described Languages. Studies in prosody and syntax*.
Edited by Evangelia Adamou, Katharina Haude and Martine Vanhove.

CHAPTER 2

Prosodic separation of postverbal material in Georgian

A corpus study on syntax-phonology interface

Stavros Skopeteas, Caroline Féry and Rusudan Asatiani
Bielefeld University / Goethe-University Frankfurt / Tbilisi State University

A striking property of Georgian intonation is that focused postverbal material is prosodically separated from the core clause. The challenge of the present study is to assess the external validity of this experimental result by means of a corpus study. Corpus data is known to contain immense variability due to uncontrolled factors related to spontaneous speech production, such as segmental effects, intra-speaker variation, etc. The corpus study confirmed that the right edge of the verb is frequently associated with a prosodic boundary that separates the prosodic constituent encompassing the verb and the preverbal material from the postverbal domain. This boundary can be overwritten by information structure, in particular by postfocal dephrasing.

1. Assumptions about empirical data

The evidential basis of linguistic generalizations is currently debated in several fields of linguistic research. Proponents of supervised data gathering methods (so called lab speech) outline the merits of controlling data in order to answer relevant research questions (Xu 2010). Proponents of naturalistic data point to the risk of artefact in data collections involving manipulation of the spontaneous linguistic behaviour of native speakers; see Sampson (2007). This view implies a notion of authentic speech, attested in spontaneous communication, and a less authentic variety that speakers produce when they are observed. Based on this notion of 'authenticity', some scholars suggest that the target of enquiry should be the spontaneous variety in which minimum attention is paid to speech (Labov 1984: 29).

Beyond polarizing statements, which are not scarce in this debate, a consensus emerges that different data types are complementary. We need comparisons between data types in order to assess (a) the relevance of each of them for linguistic research and (b) the mapping of different data types to each other (Kepser & Reis 2005; Wagner et al. 2015). Different data types reflect different linguistic processes

https://doi.org/10.1075/slcs.199.02sko

and are systematically mapped to each other: this view is elaborated in a paradigm of studies comparing frequencies in corpus with grammaticality judgments (Adli 2011; Featherston 2005; Kempen & Harbusch 2005). Furthermore, different data types are evidence for different varieties of the same code, i.e., different styles (Labov 1972: 79–85; Wagner et al. 2015). Stylistic variation is part of a speaker's competence and the mapping between styles is a relevant object of study (Wertheim 2003).

The advantage of controlled data-gathering methods is that they are maximally informative for the research question at issue: a targeted empirical design contains the exact factors of interest crossed with the maximally relevant random factors. The question nevertheless arises whether it is worth replicating an investigation in a time-consuming non-laboratory setting. The reason for doing so is to assess the *external validity* of the observed reaction to the experimental cues, i.e., to examine whether the inference from the data is generalizable to real-world situations. External validity is maximized in experimental control by including representative random factors, such as a sample of speakers representing the population and a sample of items representing the possible lexicalizations of the linguistic phenomenon at issue (Clark 1973). However, any realistic experimental design is limited to the subset of maximally relevant random factors. Furthermore, laboratory data has an intrinsic restriction: it does not guarantee the validity of the inference across laboratory and non-laboratory settings (see Reis & Gosling 2010 for a similar view in social psychology). The generalizability of the linguistic behaviour in laboratory settings cannot be taken for granted: controlled data only contain a subset of the stylistic variation that appears in spontaneous settings, i.e., they are under-informative for those phenomena that depend on stylistic variation (see, e.g., Poplack 1993: 260 on code switching).

Studies on controlled and spontaneous data make use of different procedures in order to reduce the sources of variation. The crucial difference is that controlled data are 'idealizations' of real-world situations, while spontaneous data are interpreted through processes of 'abstraction' from real-world situations (see Stokhoff & van Lambalgen 2011). Controlled data are 'idealizations' for several reasons: speakers in the lab behave in a way in which the emotional involvement to the communication is eliminated; they avoid disfluencies that frequently occur in spontaneous speech; in most laboratory phonetic studies, target utterances are performed in isolation in order to eliminate irrelevant effects of a larger discourse; scripted speech is used in order to control the variation that results from different lexicalizations of the structure under investigation. All these sources of variation are available in spontaneous data to the effect that the crucial differences may not be easily established. An account of spontaneous data has to overcome these obstacles by undertaking the methodological risk of 'abstraction', i.e., by averaging data varying in all these respects. The critical assumption is that the sources of variation are outbalanced in the averaged results. Theoretically, the idealized patterns of the laboratory setting and the abstractions from real world-settings should lead to the same inference – unless stylistic variation between settings is at issue.

The comparison between different data types has been an important issue in prosodic research in the last decades. The maximally controlled type of data is scripted speech, i.e., the performance of written sentences designed by the experimenter. All types of non-scripted speech are considered to be instances of "spontaneous speech" (Llisterri 1992: 2; Beckman 1997: 7). This conception of spontaneous speech applies when sentence-planning is carried out by the speaker and covers a large array of data, from real-world conversations to narratives collected on a topic set by the instructor, as long as the speaker has the freedom to decide the exact content and form of the produced speech.

The relation between scripted speech and spontaneous speech has been an object of enquiry in several studies on intonation. Some studies point out that the phenomena found in scripted speech also appear in spontaneous data; e.g., post-focal downstep (see Bruce & Touati 1991: 13–2; Hansson 2003: 108ff. on Swedish; Face 2003: 125 on Spanish; Koch 2008 on Thompson Salish). Spontaneous data are frequently accentually underspecified in comparison to scripted speech: e.g., unaccented words in Spanish occur more frequently in spontaneous data than in scripted data (Face 2003: 122); the phonetic reflex of this difference is that the standard deviation and the range of F_0 values is greater in scripted data than in spontaneous data (see Blaauw 1991: 12–4 on Dutch). On the other side, reflexes of information structure, e.g., deaccenting given referents, are more consistently used in scripted speech (De Ruiter 2015 on German). A further difference relates to the fact that spontaneous speech is rich in triggers of prosodic phenomena such as emotions or speaker's attitudes towards the content of speech, signals of text structure, etc., that are normally excluded in scripted data. This results in a larger variation of accentual realizations in spontaneous speech in comparison to scripted speech: see Face (2003: 124f.) on accentual realization in Spanish narratives; Brehm et al. (2014) on the variation in accentual realization of Spanish imperatives; see also Beckman (1997: 12) on the need for complex models of discourse structure in order to account for spontaneous data.

The present study investigates the mapping of prosody to syntax in Georgian. Our starting point is a generalization obtained in an experiment with scripted speech: speakers consistently aligned the right edge of non-final verbs with a high target in the intonational contour (see Section 2). Our aim is to examine whether this experimental finding also applies to non-scripted speech (see Section 3). For this purpose, we compare the experimental findings with the intonational realization of utterances in a corpus of Georgian narratives (Section 4). Our study concentrates on the methodological challenge to test a prosodic hypothesis in data with not controlled lexicalizations: the target structures may be frequently identified in spontaneous speech, but they are realized with varying lexical and consequently segmental material (see Section 5). The final comparison between spontaneous and scripted speech shows that the phenomenon at issue can be replicated in both types of data (see Section 6). The findings are discussed in the light of previous research on the comparison between different types of data in Section 7.

2. Georgian prosody and syntax

Georgian is a verb-final language that allows for non-verb final orders. Verb-final and non-verb-final orders occur very frequently in discourse (see also corpus frequencies in Section 4.2) and both SOV and SVO orders are acceptable in all-new contexts. The observed word order variation is better explained on the basis of a V-final constituent structure because Georgian displays particular types of constructions that are expected for V-final syntactic projections, e.g., lexical verb-auxiliary order, right node raising (i.e., conjunct clauses with V-ellipsis in the first conjunct), and because V-final orders are generally preferred in a larger array of syntactic configurations than the non-V-final orders, e.g., the former are more frequent with intransitives and with non-compositional (idiomatic) VPs (see detailed discussions in Harris 1981; Skopeteas & Fanselow 2010).

An experimental study with scripted speech revealed a difference in the prosodic phrasing of SOV and SVO clauses. In a production experiment, eight native speakers realized scripted sentences with V-final and V-medial orders in several contexts (Skopeteas & Féry 2014). The speakers were instructed to perform scripted sentences as 'natural answers' to questions performed by a native-speaker instructor. They were told to avoid disfluencies and were allowed to repeat the task if they felt that their performance was not natural enough. The target sentences contained three constituents, a subject, an object and a verb, in different orders, as illustrated in (1). In this experiment, only arguments were included, no adjuncts.

(1) a. SOV
nino mama-s e-loliav-eb-a.
Nino[NOM] father-DAT IO.3.PASS-care-THM-PRS.PASS.SBJ.3SG
'Nino surrounds the father with care/attention.'

b. SVO
nino e-loliav-eb-a mama-s.
Nino[NOM] IO.3.PASS-care-THM-PRS.PASS.SBJ.3SG father-DAT
'Nino surrounds the father with care/attention.'

In Skopeteas and Féry (2014), we assume a model of Georgian prosodic phrasing framed in the Match theory, a syntax-based account (see Selkirk 2011). We assume three layers of prosodic constituency: phonological words are mapped to Prosodic Words (ω), lexical projections to Prosodic Phrases (φ), and root clauses to Intonational Phrases (ι) (Gussenhoven 1984; Nespor & Vogel 1986; Selkirk 1984, 2011; Féry 2017: 36). These layers define hierarchically ordered types of prosodic constituents, which does not mean that every syntactic entity of this kind will be marked through unambiguous prosodic signals. Every argument may be phrased in a prosodic phrase (φ-phrase) individually by virtue of a constraint matching a maximal projection in the syntax to a φ-phrase, MATCH(XP,φ), as indicated in (2) by a subscripted φ in the phrasing. A verb-adjacent constituent is additionally phrased in a larger φ-phrase encompassing both the constituent and the verb. The verb itself forms a prosodic word, but not a φ-phrase on its own. A subscripted ι stands for intonation phrase (ι-phrase), a constituent matched to a syntactic clause: this is achieved by the constraint MATCH(CP,ι). The phrasing indicated in (2) is the result of the algorithm mapping syntax to prosody, and more specifically by the constraints just mentioned, MATCH(XP,φ) and MATCH(CP,ι). Elimination of boundaries may appear through the interaction with further constraints. But it should be mentioned that the pitch excursions illustrated below only reflect a subset of the possible phrasal boundaries.

(2) a. ${}_{\iota}({}_{\varphi}(\ S\ )_{\varphi}\ {}_{\varphi}(\ {}_{\varphi}(\ O\ )_{\varphi}\ V\ )_{\varphi})_{\iota}$
b. ${}_{\iota}({}_{\varphi}(\ {}_{\varphi}(\ S\ )_{\varphi}\ V\ )_{\varphi}\ {}_{\varphi}(\ O\ )_{\varphi})_{\iota}$

The difference between the phrasing of SOV and SVO clauses relates to the fact that the basic constituent order in Georgian is V-final. Postverbal material is reported to be prosodically separated from the verb in several languages with basic V-final order; see Hale and Selkirk (1987: 161) on Tohono O'odham (Uto-Aztecan), Gordon (2005: 306f.) on Chikasaw (Muskogean), Mahjani (2003: 53) on Modern Farsi (Indo-Iranian), Boeder (1991) on Old Georgian. The assumption in (2) that only preverbal material is integrated in the φ-phrase of the V has been proposed for Tohono O'odham (see Hale and Selkirk 1987: 156; Truckenbrodt 1999: 231).

The experimental study on scripted speech confirmed a difference in prosody depending on the position of the verb, as illustrated in Figure 1. The pitch tracks illustrate the prosodic realization of SOV and SVO as answers to the question 'What happened?', thus in an all-new information-structural context. The default pattern of Georgian intonation in declaratives is a sequence of gradually downstepped rising contours that encompass the prosodic constituents (Skopeteas, Féry, & Asatiani 2009: 112; Vicenik & Jun 2014); see Figure 1a. In non-verb final orders, the right edge of the verb is associated with a high target in the intonational contour that is reset at the pitch level of the first high target of the utterance (associated with the subject); see Figure 1b.[1] The tonal targets are aligned with the edges of prosodic

1. In the figures showing pitch tracks, we use smoothed contours. This allows us to concentrate on the general prosodic properties of the contours and to ignore phonetic details when they are irrelevant (the pitch excursions were smoothed at smooth level 4 in PRAAT; Boersma & Weenink 1992–2016).

constituents. Word stress has weak phonetic correlates in Georgian (and its placement is debated in Georgian phonology) and is not relevant for the realization of the prenuclear rising contours in Georgian intonation (see Skopeteas & Féry 2010).

a. [SOV]$_F$

b. [SVO]$_F$

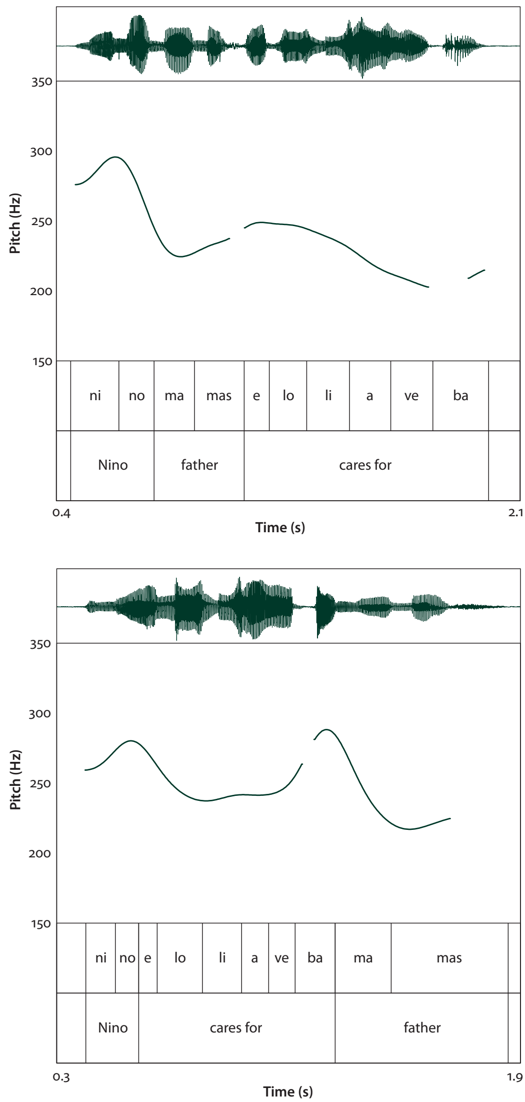


Figure 1. Phrasing in all-new contexts (speaker LEL, item 1)

The prosodic realizations in Figure 1 illustrate the preferred options in all-new contexts. In other contexts, prosody is influenced by information structure. The generalization obtained in the experimental data is that the high boundary at the right edge of the V appears if the postverbal material is part of the focus domain, i.e., in [SVO]$_F$ and S[VO]$_F$, and is even more frequent in cases of postverbal narrow focus, i.e., SV[O]$_F$. The SVO order with final narrow focus is illustrated in Figure 2a, which essentially has the same prosodic structure as the one in the SVO order in an all-new context (compare with Figure 1b). However, if the postverbal material is given, i.e., in [S]$_F$VO and S[V]$_F$O, no boundary separates the right-edge of the verb from the rest of the clause. This option is illustrated in Figure 2b: the material after the initial focus is de-accented and de-phrased and the right edge of the verb is not marked with a high boundary.

a. SV[O]$_F$

b. [S]$_F$VO

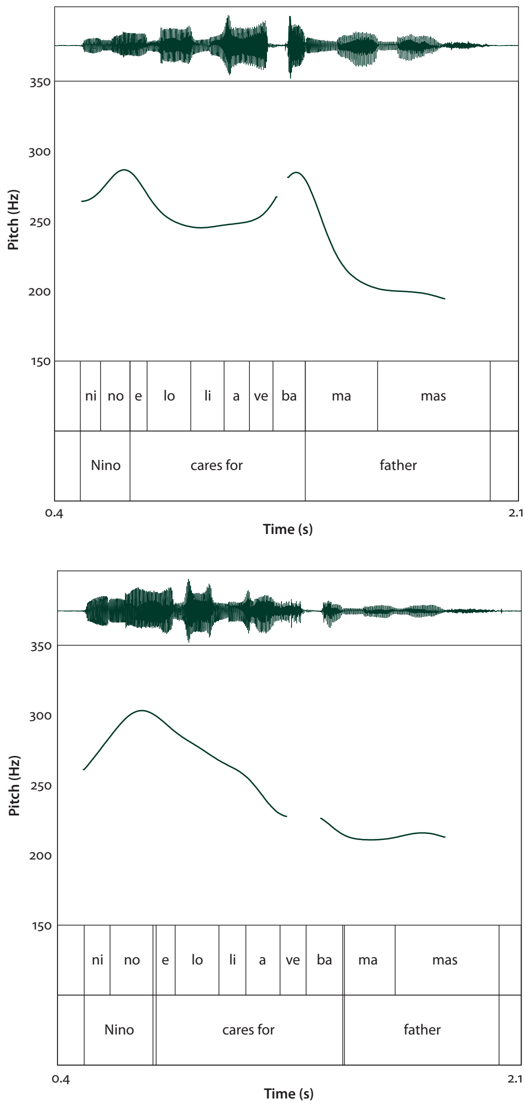


Figure 2. Prosodic realization of SVO clauses with narrow focus domains (speaker LEL, item 1)

The presented data leads to the following generalization for the Georgian prosody:

(3) In utterances with non-final verbs, the verb's right edge is associated with a H-boundary in the tonal contour, unless the postverbal material is background information.

In a nutshell, we interpret the high tone on the final syllable of the verb in Figure 1b as a high phrase tone at the level of the φ-phrase. This tonal event demarcates the right edge of a prosodic domain that contains the verb and the material on its left and separates this domain from the postverbal material, as shown in (2b). To explain the difference in tonal realization between all-new and narrow focus, we assume that the result of the syntax-prosody mapping illustrated in (2) can be overwritten by specific constraints. The Match constraints can be overwritten both by information structural needs and by well-formedness constraints that can change or correct the initial syntax-based phrasing. More specifically, a focused constituent is preferably aligned with the right or the left boundary of a φ-phrase. In Figure 2b, the default phrasing shown in (2b) is replaced by the phrasing in (4) due to the focus on the subject.

(4) $_{\iota}(_{\varphi}(\ S_{F}\ )_{\varphi}\,_{\varphi}(\ V\ O\ )_{\varphi})_{\iota}$

3. External validity of experimental findings

The findings summarized in Section 2 were obtained through an experiment with scripted speech. They present an idealized version of linguistic behaviour. The basic assumption is that speakers simulate real-world situations. They were instructed to produce the selected structures in a way that is close to their natural behaviour in discourse. A realistic experimental design must always exclude several sources of variation under the assumption that these sources are irrelevant for the research question at issue. This decision is based on plausibility or previous knowledge and is indispensable in order to get interpretable results. However, the final result is only generalizable for the random factors that vary in the experimental design. In the experiment presented in Section 2, the examined lexicalizations excluded several sources of variation based on assumptions established through plausibility. The lexicalizations only contained open syllables with non-complex voiced onsets. The findings may be not generalizable for all types of onsets, since we know that different types of onset can cause F_0 perturbations at the micro-prosodic level and that these phenomena do not reflect tonal targets at the layer of the sentential prosody. Second, the presented orders were SOV and SVO (furthermore, the experimental study also contained OSV and OVS sentences). Hence, the result is not generalizable to all types of preverbal and postverbal constituents, e.g., adverbs or adpositional phrases. Based on the current assumptions concerning the prosody-syntax interface, the alternative positions of the V may be crucial if they relate to different constituent structures, but the exact category of the postverbal constituent is not expected to play a role. Finally, the verbs used in the experiment were in a particular inflectional form, namely in the perfective past, in order to avoid the voiceless third person suffix in present tense, which limits the generalizability of the results to a particular tense. This choice in the experimental design is also plausible, since we do not know of any intonational language in which prosody depends on tense. Hence, the limitations in the experimental design are plausible. Nevertheless, the inference from the experimental result to the intended generalization partially relies on the prior knowledge that no relevant effect has been yet discovered for certain potential sources of variation.

A further limitation is the validity across styles: since the effect of the verb position is obtained through the performance of scripted sentences, it is not generalizable for the entire repertoire of styles of Standard Georgian speakers. Scripted speech involves maximal attention to the speech and is known to induce a bias against casual styles (Labov 1972: 80; Milroy & Gordon 2003: 202). The validity of the inference relies on the assumption that the behaviour of speakers who realise scripted sentences reflects their behaviour in natural discourse.

In order to verify the external validity of the generalization in (3), the present study examines whether this phenomenon holds true for spontaneous data, in which the sources of variation have not been controlled. The challenge for the interpretation of the spontaneous data is the enormous variation that is caused by different speakers, different lexicalizations, disfluencies of spontaneous speech production, etc. In order to test the generalization in (3), we need methods that abstract away from this variation in the data.

4. Corpus study: Method

The aim of this study is to examine whether the difference in phrasing between SOV and SVO utterances that we found in the scripted data is confirmed in non-scripted data. For this purpose, we investigated the intonational realization of clauses in a corpus of Georgian narratives, introduced in Section 4.1. The relevant data for our research question are defined in Section 4.2. The annotation procedure is outlined in Section 4.3.

4.1 Corpus

The aim of the present data collection was to create and analyze a corpus of less-controlled speech recordings of narratives containing comparable data from different speakers. In terms of Beckman's classification of spontaneous data, this data belongs to the category of "unstructured narratives", elicited by an open-ended question and without a pre-defined structure. Since speakers are aware that they are recorded, effects of attention to the speech are possible. However, this data certainly represents the lower end of self-conscious control of speech production in laboratory situations. Studies on sociolinguistic interviews report that, depending on their personal involvement in the topic under discussion, native speakers may forget the context of the interview and behave in a way that does not substantially differ from spontaneous communicative situations (Beckman 1997: 16; Schilling-Estes 1998; Labov 2004). The narratives in our corpus contain all properties of spontaneous data, e.g., laughter, self-repair, occasional discussions with the interviewer. Sentence-planning is spontaneous and does not differ to real-world situations. The context of an interview may have an impact on the choice of style (leading to less casual styles), which means that our findings are not generalizable across styles.

Twenty speakers produced five monologues each with the same instructions; see English translation of the instructions in (5). The corpus contains (20 × 5 =) 100 short narratives with a total duration of 91.1 min (average duration per narrative: 54.7 sec). The interviews took place in a private home in Tbilisi, in September 2012, and were conducted by Rusudan Asatiani (Tbilisi State University) and Veronika Ries (Bielefeld University); the native speakers were instructed in Georgian and were paid for their participation. Recordings were made with an Olympus LS-11 linear PCM recorder at a sampling rate of 44.1 kHz/16 bit (stereo), using the integrated microphones. The collected narratives were transcribed, glossed and translated by Rusudan Asatiani (October 2012-March 2013). Recordings and transcriptions are available as Open Resource in the TLA archive.[2]

(5) a. Ancestor story (AN)
'Please tell me how you imagine that Ancient Georgians lived. It is not a problem if you are not sure about the details. Just tell me the story of your ancestors – as far as you know it.'

2. TLA archive > Donated Corpora > XTYP lab > Georgian; <https://hdl.handle.net/1839/00-0000-0000-0021-4DA3-5@view>

b. Event description (EV)
'Please tell me how you celebrated the last New Year's feast: what did you prepare for the feast, who was there, what did you do, what did you think.'

c. Path description (PA)
'Please tell me how to go from *Vake* to *Marjanishvili* (= locations in Tbilisi). Please give exact descriptions, so that I can recognize the way that I have to follow.'

d. Activity description (AC)
'Please tell how you make a *khachapuri* (= Georgian cheese bread). Do not worry if there are some details that you do not know, just give a clear description, such that another person can do the same.'

e. Comparative description (CO)
'Please tell me how you perceive the major differences between Georgian and Russian.'

All speakers were inhabitants of Tbilisi and had Georgian as first native language. They were highly educated (either university students or already in possession of a university degree) and used the standard variety of Georgian in speaking and writing in their everyday communication (18 women, 2 men; age range 20–58 at the time of the recordings; mean 28.3; standard deviation of the sample: 10.9). As a result, this corpus represents the repertoire of styles used by native speakers of Georgian with higher academic education.

4.2 Data selection

In order to examine the prosodic contour of non-final verbs, we compare utterances with a verb (V) and two phrasal constituents (XP_1, XP_2) in two different linearizations: (a) XP_1 XP_2 V (baseline), (b) XP_1 V XP_2 (case of interest); see illustration in (6). The purpose of the annotation is to identify syntactic constituents (not prosodic constituents), in order to test a hypothesis on syntax-prosody mapping.

(6) a.
| XP_1	| XP_2		| V	|
sač'mel-eb-i	*dzalian*	*mart'iv-ad*	*k'et-d-eb-a*	
food-PL-NOM	very	simple-ADV	do-PASS-THM-SBJ.3SG	

'food is made very easily' (AC, speaker 17)

b.
| XP_1		| V	| XP_2	|
čem-i	*c'inap'r-eb-i*	*cχovr-ob-d-nen*	*sopel-ši*	
my-NOM	ancestor-PL-NOM	live-THM-IMP-SBJ.3PL	village-in	

'my ancestors were living in a village' (AN, speaker 3)

The following categories were labelled as 'V': verbs and non-verbal predicates accompanied by a copula, e.g., non-finite verb-forms, *šemo-s-ul-i=a* (PR-go.INF-PRC-NOM=be.PRS.SBJ.3SG) 'is brought' (AN, speaker 20), adjectives, e.g., *k'arg-i=a* (good-NOM=be.SBJ.3SG) 'is good' (AC, speaker 10), and PPs, *gramat'ik'a-ši=a* (grammar[DAT]-in=be.PRS.SBJ.3SG) 'is in the grammar' (CD, speaker 11). 'XP_1' and 'XP_2' stand for maximal projections: DPs, PPs and adverbial phrases. Adjacent XPs (either preverbal or postverbal) forming a single constituent were coded as one phrase, since these units are likely to be mapped on a single φ-phrase due to the recursive Match syntax-prosody algorithm that we use. This applies to syntactic phrases embedded within DPs, e.g., $|_{XP}$ *meχute sauk'un-is kartul-i* | (fifth century-GEN Georgian-NOM) 'Georgian of the fifth century' (AN, speaker 18), as well as to coordinated phrases, e.g., $|_{XP}$ *me=da čem-i kmar-i* | (1SG.NOM=and my-NOM husband-NOM) 'me and my husband' (EV, speaker 3). The XPs at issue may be arguments or adjuncts; see, e.g. the subject XP_1 in (6a–b), the adverb in (6a) and the locative in (6b).

The target sequence (XP_1 XP_2 V or XP_1 V XP_2) may be part of a larger syntactic configuration containing further phrasal constituents before or after it. In other words, an XP_1 is not necessarily the first phrasal constituent in the clause in XP_1 XP_2 V or XP_1 V XP_2 (see Figure 4a below) and the right edge of an XP_2 in XP_1 V XP_2 does not necessarily coincide with the right edge of a clause (see Figure 5b below). These further constituents are not relevant for our purposes, since the research question relates to the tonal events that appear at the right edge of the verb.

Some XP_1 XP_2 V/ XP_1 V XP_2 tokens were nevertheless left out of consideration. Our expectations for the prosodic realization of the verb only hold for declarative main clauses. Subordinate clauses or non-declaratives (e.g., questions or exclamatives) were left out, as they may have different prosodic properties. Furthermore, the prosodic measurements were used to calculate pitch movements between syllables (see Section 4.3). Since monosyllabic prosodic constituents may show different patterns due to the truncation of tonal targets, we only considered utterances in which the target constituents (V, XP_1, XP_2) contain at least two syllables. Monosyllabic XPs, e.g., *čven* (1PL.NOM/ERG/DAT) 'we', *ik* 'there', and monosyllabic verbs, e.g., *a-kv-s* (SINV.3.CV-HAVE-OINV.3) 'he has it', were not retained for analysis. Utterances with embedded clauses (argument or adjunct clauses or relative clauses embedded in nominal projections) are left out of consideration, since clausal constituents are often mapped on separate ι-phrases. Finally, utterances containing laughter, self-repairs or large breaks (adopting an arbitrary threshold of 300 msec) were excluded from analysis.

The relevant tokens for our research question are summarized per speaker in Table 1. We found tokens of both conditions (XP_1 XP_2 V and XP_1 V XP_2) in the subcorpus of each speaker, i.e., we were able to test a model including by-speakers random slopes with respect to the effect of order. The frequencies of orders indicate that clauses with postverbal constituents are more frequent (66.2%) than clauses without postverbal constituents, thus with a final verb (33.8%); no speaker produced more V-final than V-non-final clauses. The proportion of clauses with non-final verb is higher than in written registers (Apridonidze 1986 reports 46.1% of clauses with a non-final verb in a corpus of 23.253 clauses of written Georgian). Georgian certainly differs from rigid V-final languages (e.g., Japanese or Korean) in which V-medial orders occur rarely.

Table 1. Valid tokens per speaker

speaker	**XP_1 V XP_2**		**XP_1 XP_2 V**		**total**
	n	%	*n*	%	*n*
1	*7*	54	*6*	46	*13*
2	*5*	63	*3*	38	*8*
3	*19*	79	*5*	21	*24*
4	*12*	60	*8*	40	*20*
5	*7*	64	*4*	36	*11*
6	*9*	69	*4*	31	*13*
7	*10*	59	*7*	41	*17*
8	*14*	70	*6*	30	*20*
9	*8*	62	*5*	38	*13*
10	*9*	82	*2*	18	*11*
11	*4*	57	*3*	43	*7*
12	*7*	64	*4*	36	*11*
13	*4*	50	*4*	50	*8*
14	*10*	50	*10*	50	*20*
15	*11*	58	*8*	42	*19*
16	*10*	83	*2*	17	*12*
17	*10*	67	*5*	33	*15*
18	*11*	79	*3*	21	*14*
19	*13*	87	*2*	13	*15*
20	*7*	54	*6*	46	*13*
total	***187***	**66**	***97***	**34**	***284***

4.3 Annotation and analysis

The data were annotated in PRAAT (Boersma & Weenink 1992–2016). In an interval tier of a Textgrid object, we annotated the voiced part of the initial and the final syllable of each target constituent (V, XP_1, XP_2); see illustration in Figure 3, first tier (σ = critical syllables (initial and final) of the target constituents).

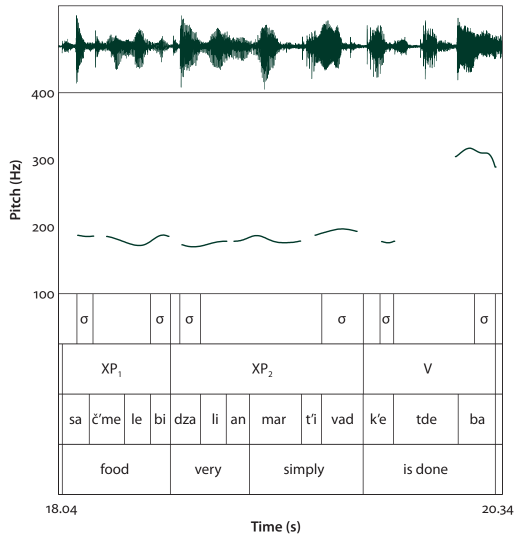


Figure 3. Critical frames for F_0 measurements; speaker 17, text AC, see (6a)

In order to estimate the influence of information structure, we annotated the post-verbal argument noun phrases for givenness: referential noun phrases were annotated as [given], if their referent was already introduced in discourse. Further distinctions relating to the discourse status of the referents may have an impact on prosody as well (Baumann 2006; Baumann & Riester 2013 for German). However, we restricted the annotation to previous mention of the referents in discourse, since this is the maximally uncontroversial criterion for the annotation of textual data.

For each speaker separately, we identified an optimal pitch range that minimized measurement errors – in particular octave jumps. View pitch range was set at this optimal pitch range for each speaker. A PRAAT script written by the first author extracted the time of start and end of each annotated interval (voiced part of the target syllables) and divided the interval in three frames of equal duration. The following measurements were extracted for each frame: F_0 maximum, F_0 minimum, F_0 mean, temporal alignment of the F_0 maximum and F_0 minimum in the syllable. Pitch measurements were extracted with the autocorrelation method.

All calculations on the extracted measurements were made in *R* (R Core Team 2016). Pitch values in Hz were converted into semitones with a reference value of 50 Hz. The Hz-to-semitone conversion was done with the formula $f_{\text{semitones}} = 12(\log_2 \cdot f_{\text{HZ}}/50)$ (see Nolan 2003 for the goodness of fit of the semitone scale; see Grice *et al.* 2007 for a previous conversion with a reference value of 50). The central tendency of F_0 measurements is reported with median values that are less affected by outliers than the means. The variability of F_0 measurements is estimated with the median absolute deviation, which is the median of the absolute differences of the data points to the median of the sample: MAD = $\mathit{median}_i\ (|x_i - \mathrm{median}_j(x_j)|)$. This measure is more robust than the standard deviation of the mean, since it is not influenced by outliers and by the sample size. Statistical evaluation of the findings was done with a linear mixed-effects model (see Section 5.4) that was calculated with *R*-package *lme4*, version 1.1–12 (Bates *et al.* 2016).

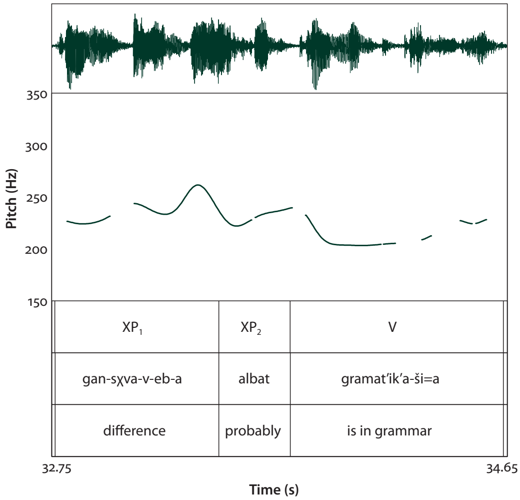
5. Corpus results

The extracted measurements and the effects of the relevant factors are summarized in the following sections. Section 5.1 outlines the effects of word order, in particular the difference between the prosodic realization of V-final and V-medial orders. Section 5.2 addresses the issue whether information structure may account for a part of the observed phenomena. Section 5.3 considers the impact of random factors, in particular effects of segmental variation on the pitch excursion. Finally, statistical models are fitted to the data in Section 5.4.

5.1 Effects of word order

Figure 4 presents two illustrative examples of V-final utterances; see the morphological transcriptions in (7)–(8). In Figure 4a, the preverbal constituents, XP_1 and XP_2, are realized with rising contours reaching a high tonal target at the right edge of the corresponding constituent, as compared to the following verb. Figure 4b illustrates another type of prosodic realization of V-final utterance. The left edge of the initial constituent is associated with a high target in the intonational contour and the overall contour of this constituent tends downwards. The second preverbal constituent is not separated from the verb by means of a high phrase tone; this indicated that this constituent and the verb together form a joint φ-phrase. The intonational contour ends with a rising pattern that targets a H-tone associated with the right-edge of the clause (continuation rise) and separates this clause from the subsequent one.

a. speaker 11, text CD; see (7)



b. speaker 9, text PA; see (8)

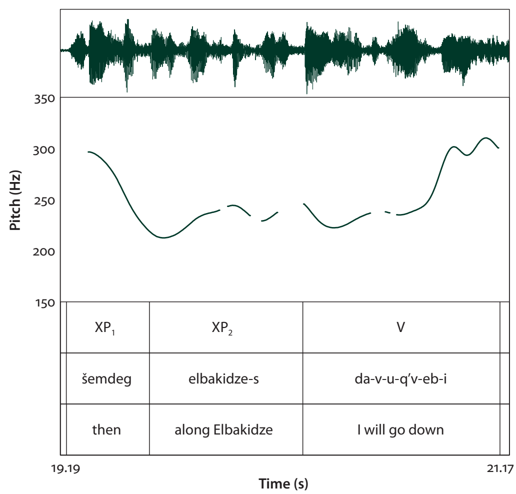


Figure 4. Prosodic realization of XP_1 XP_2 V

We assume the syntax-prosody mappings briefly introduced in Section 2. The contour illustrated in Figure 4a is a single Intonation Phrase (ι). This realization corresponds to the phrasing in (7), as predicted by a strict application of the Match constraints. The rising contours within this ι-phrase signal separate prosodic subconstituents. The first XP, *gansχvaveba* 'difference', is the subject of the clause. It is realized in a φ-phrase enclosed with a H-target on its right edge. The second XP, *albat* 'probably', is a sentential adverb. This XP is also realized with a rising contour, which is evidence that it is phrased as a separate φ-phrase. Crucially, the H-target at the right edge of this prosodic constituent is downstepped. Downstep is motivated by the prosodic structure; see (7), compare with (2a). The H-target of the immediately preverbal constituent is associated with the right edge of a prosodic subconstituent, which is embedded into a φ-phrase that encompasses the adverb and the verb, and is the prosodic sister of φ-phrase of the subject. Downstep is evidence for prosodic sisterhood (see Ladd 1990; Féry & Truckenbrodt 2005). In the case at issue, it applies to the relative pitch height of the phrase tone of the φ-phrase of the subject and the subsequent φ-phrase of the adverb that is embedded in the sister of the subject (the extended V-projection).

(7) $_\iota$($_\varphi$(*gansχvaveba*)$_\varphi$ $_\varphi$($_\varphi$(*albat*)$_\varphi$ *gramat'ik'a-ši=a*)$_\varphi$)$_\varphi$)$_\iota$
difference[NOM] probably grammar-in=be.SBJ.3SG
'Probably, the difference lies in grammar.' (speaker 11; text CD)

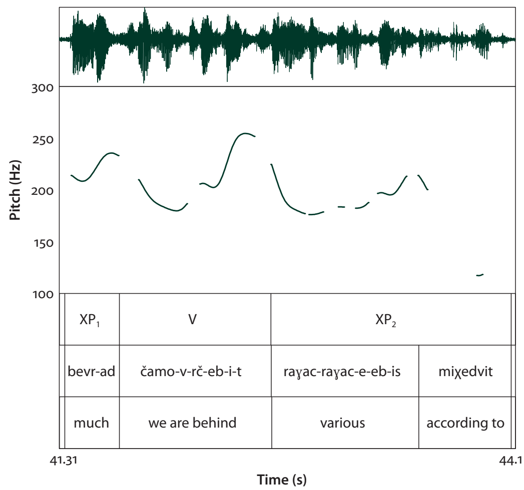
In Figure 4b, the preverbal constituents, XP_1, XP_2 and verb, are integrated into a single φ-phrase; see (8). The falling contour of XP_1 starts with a H-target that is associated with the left edge of the ι-phrase in this contour, while the rise at the end of the verb delimits this ι-phrase from the material that follows. The tonal events in Figure 4b indicate the boundaries only weakly – if at all. They are compatible with boundaries of prosodic words, rather than φ phrases. Again it is expected that the pre-verbal constituent and the verb are joined in a single φ-phrase, but in the present case, since the temporal adjunct is phrased together with the adjunct, it is also included in the larger φ-phrase. In this case, we assume that the prosodic boundaries predicted by the Match constraints are deleted. This can happen because of information structure, especially givenness, as shown in Section 2 but also because of speech tempo, or any individual preference. It may also be the case that the prosodic and tonal phrasing shown in (8) is likely to be realized when XP1 is not the subject, but part of the VP.

(8) $_\iota$($_\varphi$(*šemdeg elbakidze-s* *da-v-u-q'v-eb-i*)$_\varphi$)$_\iota$
then Elbakidze-DAT PR.FUT-SBJ.1-IO.3.CV-go_down-THM-PRS/FUT
'Then (I)'ll go down along Elbakidze.' (speaker 9; text PA)

Utterances with non-final verbs are illustrated in Figure 5. The preverbal material is either integrated into the same φ-phrase with the verb, as in Figure 5b, or is mapped by a separate φ-phrase, as in Figure 5a. But crucially, the verb is always falling together with the right edge of a φ-phrase.

In Figure 5a, XP_1 is realized with a rising contour that separates it from the verb, as predicted by a strict application of the Match constraints. However, the H-target at the right edge of the verb is non-downstepped; compare with Figure 4a. The absence of downstep reflects the assumption in (2b) that there is a φ-phrase

a. speaker 23, text AN; see (9)



b. speaker 10, text EV; see (10)

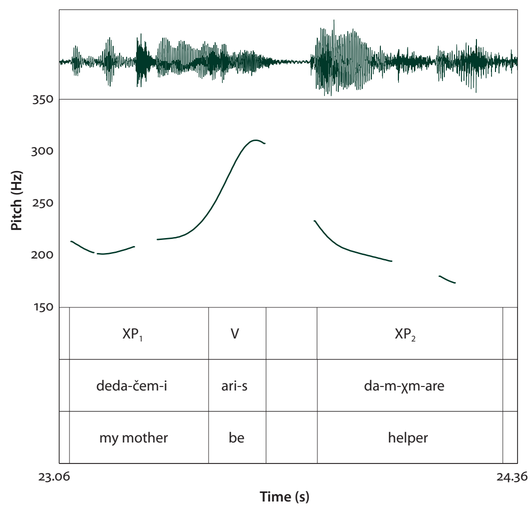


Figure 5. Prosodic realization of XP_1 V XP_2

separating the verb and the material on its left from the postverbal domain (XP_2); see assumed prosodic structure in (9).

(9) $_\iota(_\varphi(_\varphi($*bevr-ad*$)_\varphi$ *čamo-v-rč-eb-i-t*$)_\varphi$
much-ADV PR-SBJ.1-be_backward-THM-PRS-SBJ.1.PL
$_\varphi($*raɣac-raɣac-e-eb-is* *miχedvit*$)_\varphi)_\iota$
some-some-0-PL-GEN according_to
'We are lying very much behind with respect to various things.'

The preverbal material in Figure 5b (XP_1) is integrated within the φ-phrase that contains the verb. There is no boundary tone at the right edge of XP_1, which would indicate the edge of a prosodic constituent. The H-target at the right edge of the verb delimits this φ-phrase from the φ-phrase that contains the postverbal material (XP_2). Thus, the pre-verbal constituent and the verb jointly form a φ-phrase.

(10) $_{\iota}(_{\varphi}($*deda-čem-i* *ari-s*$)_{\varphi}$ $_{\varphi}($*damχmare*$)_{\varphi})_{\iota}$
mother-my-NOM be.PRS-SBJ.3SG helper[NOM]
'My mother is (my) assistant.'

The relevant difference between the XP_1 XP_2 V utterances in Figure 4 and the XP_1 V XP_2 utterances in Figure 5 lies in the prosodic separation of the postverbal material that is marked by the H-target at the right edge of the verb in the SVO pattern. The question is whether this phenomenon holds true in the totality of V-final and V-medial utterances in the sample. Figure 6 presents the time-normalized average F_0 measurements of all tokens of Table 1 (see Xu 1999 on averaging intonational contours). The F_0 contours are the averages of the F_0 means of the initial and the final syllable of each target constituent (V, XP_1, XP_2), measured in three frames of equal duration and converted into the semitone scale (see Section 4.3). The rising contour of the final constituent is the result of the frequent continuation rises in narratives. A rising contour of the final word appears in 160 out of 284 clauses in our corpus (56%). The high frequency of continuation rises in narrative texts is known from previous corpus studies and reflects the fact that low boundaries mainly appear in declaratives that close a discourse unit (see Oliveira 2000 on Brazilian Portuguese & van Donzel 1999 on Dutch). The crucial result presented in Figure 6 is the difference in the realization of the medial constituent of V-final and V-medial clauses. The contour of medial verbs displays an average rise that is larger than the baseline established by the medial constituent of V-final clauses. This finding descriptively confirms the first part of the generalization in (3), namely that the right edge of a non-clause final verb is associated with a H-boundary in the contour.

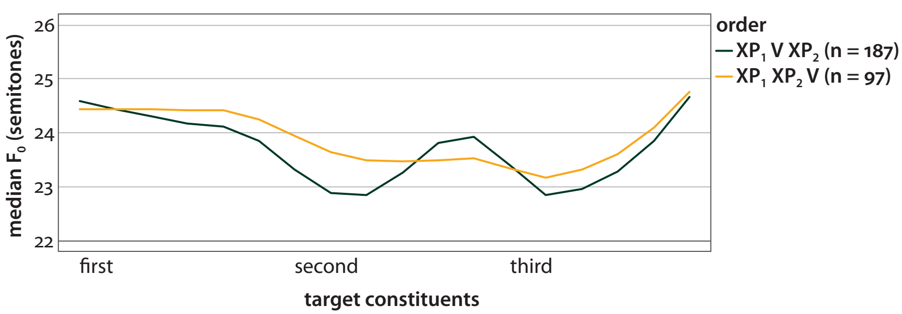


Figure 6. F_0 measurements in V-final and V-medial orders (three frames of equal duration of the initial and the final syllable of each constituent)

5.2 Effects of information structure

The generalization in (3) contains a condition on the information structure of the postverbal material: the right edge of a non-clause final verb is associated with a H-boundary unless the postverbal material is background information. In the experimental manipulation, focus was elicited through context questions. In order to estimate the effects of focus in the corpus data, we annotated the discourse status of the referents: referential noun phrases were annotated as [given], if their referent was already introduced in discourse; see Section 4.3. In utterances with given postverbal noun phrases, the prosodic realization was reminiscent of the experimental findings: there was no high boundary tone on a verb preceding given constituents; see Figure 7. The postverbal arguments in this utterance had already been introduced in the immediately preceding sentence; see (11). The focused part of the critical utterance is the verb. The postverbal object is prosodically integrated with the verb and both constituents are mapped by a single φ phrase.

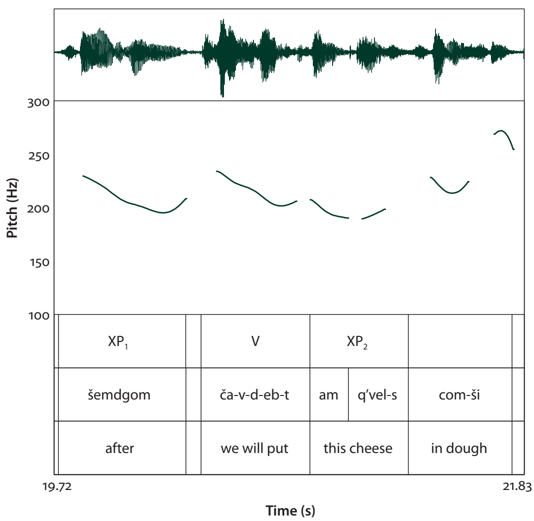


Figure 7. Prosodic realization of XP_1 V $XP_{2[given]}$
speaker 11, text AC; see (11)

(11) Pretext: 'And, in order to make good khachapuri, we will choose the cheese, and correspondingly the dough.'

$_{\iota}(_{\varphi}$(*šemdgom*)$_{\varphi}$ $_{\varphi}$(*ča-v-d-eb-t* *am* *q'vel-s* *com-ši*)$_{\varphi}$)$_{\iota}$

after PR.FUT-SBJ.1-put-THM-SBJ.1PL this cheese-DAT dough-in

'Then we will put this cheese in the dough.' (speaker 11; text AC)

Figure 8 plots the 187 ‘XP_1 V XP_2’ utterances of the corpus depending on the givenness of the postverbal constituent. The utterances with a given XP_2 contain referential noun phrase arguments whose referent has previously been introduced into the discourse (n = 36; 19.3%). The complementary group contains non-referential constituents (i.e., adverbs, adpositional phrases, non-referential noun phrases) as well as referential noun phrases whose referent is new. Figure 8 shows that the prosodic properties identified for verbs in Section 5.1 do not hold true for utterances with a final given XP_2. In line with the experimental data, there is no high boundary tone on a verb preceding given constituents. The prosodic pattern of the V-medial clauses in Figure 6 reflected the prosodic pattern of the utterances with a new postverbal argument, a feature of the majority of the cases in our corpus (151 out of 187; 80.7%) (see statistics in Section 5.4). In Figure 8, however, it is shown how the verb looks like when the postverbal XP_2 is new in comparison with when it is given.

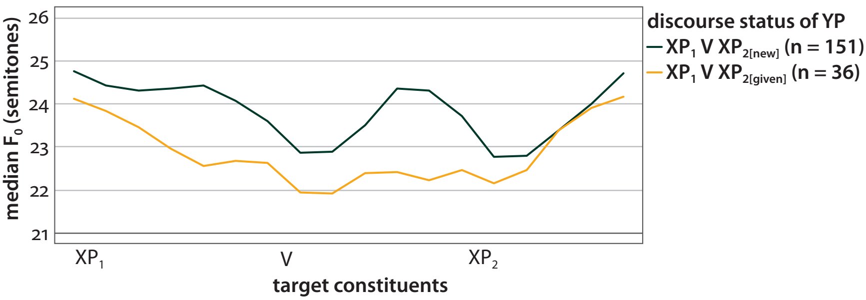


Figure 8. F_0 measurements of V-medial orders depending on the discourse status of the postverbal constituent

5.3 Effects of random factors

In contrast to the scripted data, the corpus data contain variations caused by several non-controlled factors; see Section 1. These random factors have been included in the statistical model so that their effect was subtracted from the effects of interest. In the following, we consider two phonetic factors that had an effect on F_0, namely the different types of syllabic onsets and the different vowels in the syllabic nucleus.

Some syllabic onsets cause micro-prosodic perturbations in the intonational contour. In particular, voiceless onsets lead to a higher F_0 of the following vowel than voiced onsets (Gussenhoven 2004: 7). In order to capture the part of the variation that is due to the onset, we classified the possible Georgian onsets to four types (see inventory in Shosted & Chikovani 2006: 255; see previous studies on segmental effects: Jilka & Möbius 2006; van Santen & Hirschberg 1994).

(12) Types of onset
  a. no onset
  b. voiceless obstruents
    [pʰ], [p̣], [kʰ], [ḳ], [tʰ], [ṭ], [q̣], [ts], [tṣ], [tʃ], [tʃ̣], [s], [ʃ], [χ], [h]
  c. voiced obstruents
    [b], [d], [g], [v], [z], [ʒ], [ɣ], [dz], [dʒ]
  d. sonorants
    [m], [n], [r], [l]

The by-onset aggregated measurements for the initial and final syllable of the medial constituent (i.e., the XP$_2$ in 'XP$_1$ XP$_2$ V' and the V in 'XP$_1$ V XP$_2$') are presented in Figure 9. The descriptive data reveal a difference between voiceless obstruents and the other types of onset. While the average measurements generally display a falling contour in the initial syllable and a rising contour in the final syllable of the medial constituent (compare the averages in Figure 6), syllables with a voiceless onset start with a fall in both cases. This finding is in line with previous observations (Gussenhoven 2004: 7).

a. initial syllable of the medial constituent **b.** final syllable of the medial constituent

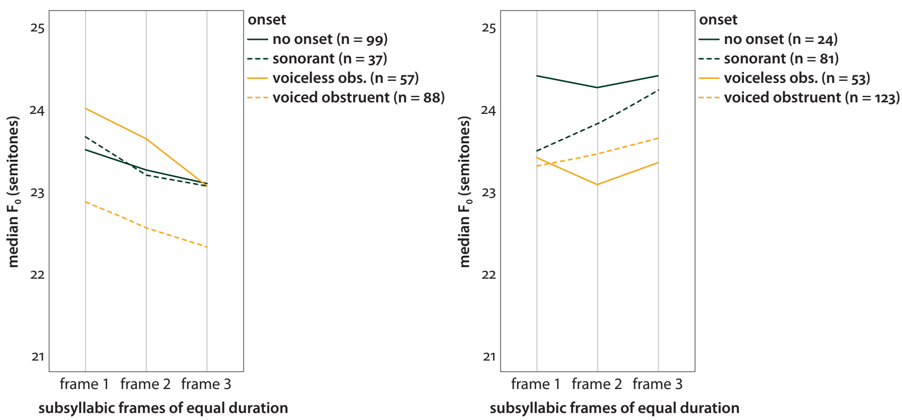


Figure 9. Pitch excursion of the target syllables depending on onset

Furthermore, vowels' intrinsic pitch is known to influence F_0 height and peak alignment, such that higher vowels correlate with (a) higher peak values and (b) later peak alignment (Jilka & Möbius 2007). The available phonetic facts on the acoustic properties of Georgian vowels reveal the following scale of vowel height (as reflected in the central values of sample F_1 measurements): i > ɛ > u > ɔ > a (Shosted & Chikovani 2006: 255). The by-nucleus aggregated measurements are plotted in Figure 10. The F_0 measurements do not directly reflect the scale of vowel height, which predicts that [i] and [ɛ] should be realized with the highest pitch. This finding indicates that the variance in spontaneous data is too large for confirming hypotheses about the intrinsic pitch of vowels. The expectation of a later peak for high vowels is descriptively confirmed by [i] in both initial and medial syllables (compare with German in Jilka & Möbius 2007).

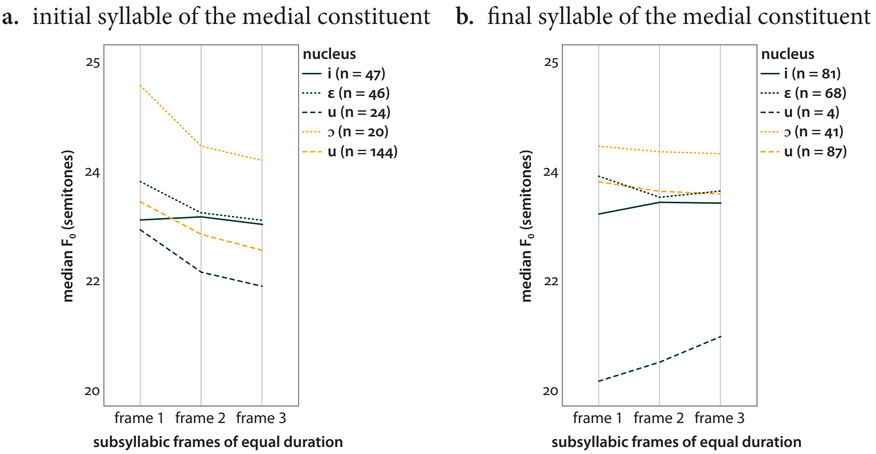


Figure 10. Pitch excursion of the target syllables depending on nucleus

The descriptive facts in this section show that different onsets and different nuclei have a potential effect on F_0. The fact that the overall tendencies presented in Figure 9 and Figure 10 only very roughly correspond to the prediction about intrinsic pitch of onsets and nuclei suggests that these sources of variation are not crucial in spontaneous data. Our purpose is not to confirm the segmental effects but to test whether the differences of interest (due to word order and information structure) are independent from the variation that is due to phonetic micro-variation. In the experimental study outlined in Section 2, these factors were controlled, i.e., the experimental result is not generalizable for the totality of syllable types. The result of the corpus study is generalizable for these sources of variation if the difference of interest be significant under consideration of the random effects of phonetic micro-variation.

5.4 Statistical modelling

In order to assess the statistical significance of the reported findings, we fitted linear mixed-effects models on the rise of the medial constituent that was shown to be influenced by word order and givenness. The rise was calculated as the difference between the F_0 maximum of the final syllable minus the F_0 minimum of the initial syllable of the medial constituent.[3] The random factors of the model capture the relevant aspects of variation in the data across which our generalization is assumed to apply (SPEAKERS, ONSETS, NUCLEI). The factor ONSET relates to the onset of the final syllable and has the four levels defined in (10): (a) zero; (b) –voiced; (c) +voiced, –sonorant; (d) +sonorant. The factor NUCLEUS relates to the vowel quality of the nucleus of the final syllable and has five levels: (a) i; (b) ɛ; (c) u; (d) ɔ; (e) a (see Section 5.3).

The fixed factors of the model should capture the factors of interest (order and givenness). Since the effect of final given referents on the realization of a prosodic boundary at the right edge of the verb is not observable with V-final orders, givenness is nested within the level 'XP_1 V XP_2' of the factor ORDER. This opens a set of alternative models that were tested in the data; see Table 2. The FULL MODEL contains a fixed factor with three levels, i.e., the permutations of orders and givenness in the data. The ORDER MODEL distinguishes between the V-medial and V-final orders and ignores the role of givenness. The ORDER/GIVENNESS MODEL examines the contrast between utterances with postverbal material that is new with the remaining utterances (V-final utterances and V-medial utterances with given postverbal material). According to the findings of the experiment summarized in Section 2, the maximal goodness of fit was expected to be reached by the latter model, since the verbs are associated with a high boundary at their right edge unless the postverbal material is postfocal.

Table 2. Model comparison

	XP_1 XP_2 V	XP_1 V XP_2		model comparison		
		XP_1 = –given	XP_2 = given	*BIC*	χ^2 (df)	*p*
Full Model	α	β	γ	1572.9	1.3 (1)	.3
Order Model	α	β		1568.5	.6 (0)	–
Order /Givenness Model	α		β	1568.0		

The results in Table 2 confirm this expectation. The maximal goodness of fit was reached by the ORDER/GIVENNESS MODEL: the least value according to the *Bayesian Information Criterion* (BIC = 1577.2) indicates the maximally informative model, i.e., the model that reaches the best fit with the minimal number of assumed

3. As F_0 maximum we consider the highest value among the F_0 means of the three averaged frames (see Section 4.3). These measurements are less affected by extreme values.

differences. Log-likelihood tests comparing the deviance of the FULL MODEL with the deviance of the reduced models result in chi-square values that are associated with non-significant *p*-levels. This means that the loss of information that results from adopting a model with fewer parameters than the FULL MODEL is not significant. The fixed factor of the winner model (ORDER/GIVENNESS MODEL) cannot be further reduced: the comparison with a model containing only random effects reveals a chi-square of 7.5, $df = 1$, which corresponds to a significant *p*-value ($p < .01$).

The estimates of the model with the maximal goodness of fit (ORDER/GIVENNESS MODEL) are given in Table 3.[4] The effect of the contrast between V-medial sentences with non-given constituents and the remaining utterances is significant, as shown by the chi-square value obtained by the difference between the deviance of this model minus the deviance of a model containing only the random factors. This result takes the variation induced by several speakers and several segmental factors into account, i.e., it is generalizable for the population from which the speakers come from as well as for the various types of syllable of Georgian. The significant *t*-value is confirmed by a log-likelihood test: the difference between the deviance of the ORDER/GIVENNESS MODEL and the deviance of a model without fixed effects is 7.53, which corresponds to a significant *p*-value ($p < .01$) in the χ^2-distribution.

Table 3. Linear mixed-effects model on the rise of the medial constituent

	factors	**estimate**	**st. error**	***t*-value**	***p***	**st. dev.**
medial rise =	intercept +	1.194	.353	3.377	< .001	
	ORDER/GIVENNESS +	1.240	.443	2.801	< .01	
	NUCLEUS +					1.281e-07
	ONSET +					7.254e-08
	SPEAKER					5.889e-01

6. Variability of scripted and spontaneous data

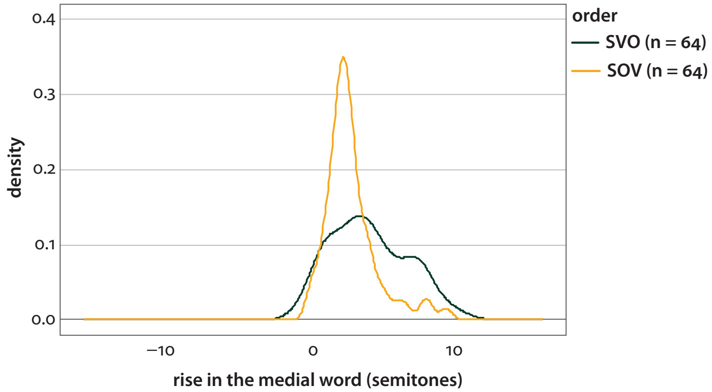
The results reported in Section 5 confirm that the generalization in (3) is also valid for spontaneous data. Non-final verbs in Georgian are characterized by a rising contour reaching a high target at the final syllable of the verb – unless the postverbal material is given.

4. A model including slopes and intercepts for all random effects cannot be evaluated because it does not converge with the FULL MODEL. The findings reported in Table 3 only include random intercepts and are obtained in *R* with the formula: lmer(rise~ order.givenness + (1|speaker)+(1|onset)+(1|nucleus), data).

In order to estimate the differences in the distribution of the data, we plotted the probability density of the discussed datasets. Notice, however, that the two datasets are only partly comparable due to the difficulty of identifying information structure in observational research. Hence, we compared the realization of all-new sentences in the scripted data with the realization of utterances independently of information structure in spontaneous data. The background assumption was that the average of all utterances in the corpus corresponded to a 'default' prosodic realization, which was likely to appear in neutral contexts.

Figure 11 presents the probability density of the two datasets discussed in this study (see descriptive statistics in Table 4). The *x*-axis displays the rise in the medial

a. scripted speech (all new)



b. spontaneous speech (various information structures)

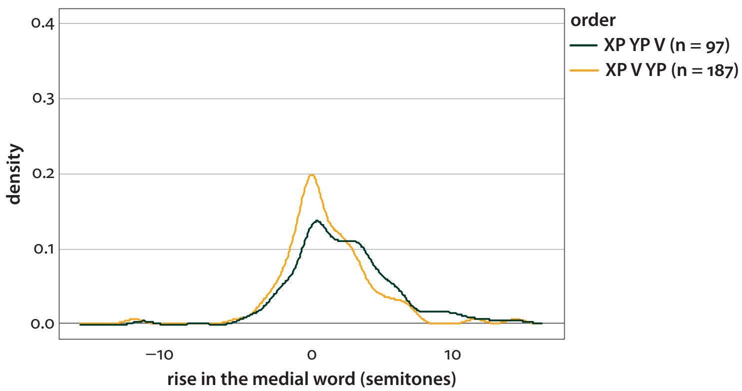


Figure 11. Word order effects: probability density

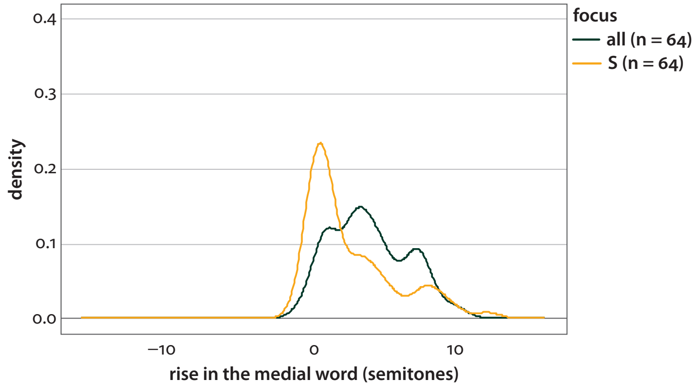
word (in semitones) calculated as the difference between the F_0-max of the final syllable minus the F_0- min of the initial syllable. Both in scripted and spontaneous speech, the central value is higher for V-medial orders (which indicates a larger average rise in the medial word). The difference between medians is 1.5 semitones in scripted speech (see Table 4) and 1.3 semitones in spontaneous speech. All distributions are positively skewed, i.e., the median is smaller than the mean. Variability is generally larger in the spontaneous data (average of median absolute deviations: 2.65) than in the scripted data (average of median absolute deviations: 2.25). The rise is dispersed around a larger range of values in the V-medial than in V-final orders in both datasets.

Table 4. Rise of the medial word (in semitones; ref. 50) depending on word order

	scripted speech		**spontaneous speech**	
	SOV	**SVO**	**XP_1 XP_2 V**	**XP_1 V XP_2**
mean	2.7	3.9	1.1	2.3
median	2.2	3.7	.5	1.8
median absolute deviation	1.1	3.4	2.4	2.9
n	64	64	97	187

A similar contrast can be observed with respect to the effect of information structure. Again, the two datasets do not display identical conditions with respect to information structure. In scripted speech, the focus domain is manipulated with context questions; in spontaneous speech, we annotated given referents (see Section 4.3). The generalization in both types of data is that the average rise is lower if the postverbal constituent is given. In scripted speech (Figure 12a), this generalization is achieved by the contrast between all focus (median rise: 3.7 semitones) and non-final focus (median rise of S-focus: 1.1 semitones) (see Table 5). The same pattern also appears in spontaneous speech (Figure 12b): V-medial sentences with a given postverbal XP_2 display a median rise of 2.2 semitones, while the corresponding sentences with a non-given postverbal XP_2 display a median rise of 1.5. Median absolute deviations are higher for the spontaneous data (average: 2.9) than for the scripted data (average 2.33). Hence, both types of data reveal the same phenomenon, but the spontaneous data display smaller contrasts and larger dispersion.

a. scripted speech
$[SVO]_F$ vs. $[S]_F VO$



b. spontaneous speech
XP_1 V $XP_{2\ [new]}$ vs. XP_1 V $XP_{2\ [given]}$

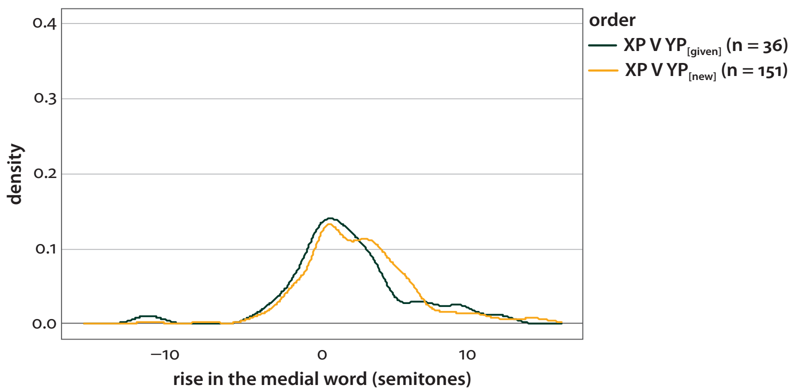


Figure 12. Effects of information structure: probability density

Table 5. Rise of the medial word (semitones, rf. 50) depending on information structure

	scripted speech		spontaneous speech	
	all-focus	S-focus	$XP_{2\ [new]}$	$XP_{2\ [given]}$
mean	3.9	2.6	2.4	1.7
median	3.7	1.1	2.2	1.5
median absolute deviation	3.4	1.7	3.1	2.7
n	64	64	151	36

7. Discussion

The main contrasts observed in scripted data are replicated in the spontaneous data with a greater variability. This result confirms that the properties of scripted speech are not artefacts of the lab situation. This is in line with the view expressed in previous studies that speakers in the lab select patterns that reflect the phonology of the language and are learnt in real world communication (see similar observations in Bruce & Touati 1991: 13–2; Koch 2008 in Section 1).

A problem for testing hypotheses in spontaneous data is the immense variability, which is due to the large array of uncontrolled factors (Xu 2010: 334; Face 2003: 124f.). An empirical solution to this problem is the use of mixed models that take account of the relevant sources of variability in their random-effect structure (see Section 5.4). In the statistical model of the spontaneous data we examined the impact of several onsets on pitch. Voiceless obstruents have an influence on pitch, as the present corpus study has shown (see Figure 9). The frequency of voiceless onsets is not equal across conditions. In the onset of the final syllable of the medial constituent, voiceless consonants occur in 31 out of 187 XP_1 V XP_2 tokens (17%), and in 22 out of 97 XP_1 XP_2 V tokens (23%) in our corpus. Descriptively speaking, the difference is small and the segmental effects on fundamental frequency are not large enough to explain the difference between V-final and V-medial sentences (compare segmental effects in Figure 9 with the word order effects in Figure 6). A statistical model with the factor ONSET in its random structure confirms that the differences of interest are not reducible to the impact of ONSET, i.e., the finding is generalizable across onsets; see Table 3. This procedure appropriately treats the obstacle of variability in spontaneous data.

Spontaneous data are known to be frequently accentually underspecified (see observations of Face 2003 & Blaauw 1991 in Section 1). This difference predicts that the contrasts between the conditions of interest will be clearer in scripted speech, which is confirmed by our data; Figures 11–12. In particular, deaccenting given referents is consistent in scripted data, but less so in spontaneous data (De Ruiter 2015). The comparison between scripted and spontaneous data in Figure 12 exactly reflects this difference; see the high density of sentences around 0 rise in Figure 12a.

Finally, a source of variability in spontaneous data is that the contextual predictors in corpus are less precise than the manipulated environments in an experimental study. This problem is immanent in the study of information structure by means of observational research in corpora. Contextual properties such as the discourse status of the referents offer useful predictors, but they are only approximations of the factors that are assumed to be crucial in phonological theory, i.e., the domain of focus. It is not excluded that a referent already introduced in discourse is part of the focus domain of the utterance, although it is expected that this will be less likely the case than with new referents. In order to refine the contextual predictions in corpus data, we need richer models of the processes that take place in discourse (see Beckman 1997: 12). The operational decision in the present corpus study was to use givenness as a maximally straightforward predictor, since givenness can be unambiguously annotated on the basis of previous mentions of the referent in the text.

8. Conclusions

We investigated a hypothesis about Georgian intonation in spontaneous speech and compared our findings with previous findings from scripted speech. The mapping of both datasets reveals that the same contrasts are available, with larger differences and less variability in the scripted data. The presented corpus study shows that getting prosodic generalizations from spontaneous data is a realistic task and that the immense variability in this data can be effectively managed with an appropriate corpus design and a statistical procedure that takes the sources of variation into account.

Stylistic differences may come from the fact that scripted data involve maximal attention to the speech that results in a bias for formal styles, while speakers producing narratives may forget the laboratory situation and move to styles that occur in casual communication (see Section 4.1). The fact that the same pattern is observed in both datasets implies that it is generalizable for the range of styles that are represented in these datasets.

However, the primary difference in the compared datasets is not the potential stylistic variation, but the fact that the speakers are forced to read aloud a written text in the scripted data while they are free to use their own words in the spontaneous data. The methodological consequence of this difference is the variability of the spontaneous data. Relevant sources of variation are controlled in scripted data in order to avoid irrelevant influences on the effect of the factors of interest. These sources of variation vary freely in spontaneous data, i.e., it cannot be excluded that their effects are confounded with the effects under investigation. The solution to this problem is to include the relevant sources of variation to the random-effects structure of the statistical model.

A problem for generalizations on corpus data comes from the lack of control for intraspeaker variation (Xu 2010: 332). This difficulty can be managed with better corpus designs. In this study, the corpus contained narrative texts produced by 20 speakers under identical conditions and with identical instructions. The advantage of this design is that variation between registers is not confounded with the intra-speaker variation. A mixed-linear model that considers the effect of the random factor SPEAKER is the appropriate empirical procedure to ensure that the result is independent from the variation caused by different individuals.

It is not surprising that the findings of scripted speech correspond to phenomena in spontaneous speech. The prosodic patterns that speakers produce in scripted speech reflect the phonological properties of the language and are expected to apply in real world communication too. There is no reason to assume that native speakers invent a new grammar of the syntax-prosody interface when they are asked to perform scripted sentences. This implies that the idealization assumed for laboratory data is not an artefact that only exists in the lab. Finally, our study shows that it is feasible to use uncontrolled data for confirmatory research and that it is possible to address specific relevant questions in a dataset with enormous variation by entertaining the necessary steps of abstraction.

Abbreviations

1	first person	OINV	inverted object
3	third person	PASS	passive
ADV	adverbial case	PL	plural
CV	characteristic vowel	PR	preverb
DAT	dative	PRC	participle
ERG	ergative	PRS	present
FUT	future	SBJ	subject
GEN	genitive	SG	singular
IMP	imperfective	SINV	inverted subject
IO	indirect object	THM	thematic marker
NOM	nominative		

References

Adli, Aria. 2011. On the relation between acceptability and frequency. In *The Development of Grammar: Language Acquisition and Diachronic Change* [Hamburg Studies on Multilingualism 11], Esther Rinke & Tanja Kupisch (eds), 383–404. Amsterdam: John Benjamins. https://doi.org/10.1075/hsm.11.20adl

Apridonidze, Šukia. 1986. *Sit'q'vatganlageba axal kartulši* (Word order in modern Georgian). Tbilisi: Mecniereba.

Bates, Douglas, Maechler, Martin, Bolker, Ben & Walker, Steven. 2016. lme4: Linear mixed-effects models using 'Eigen' and S4 classes. R package, version 1.1–12. <http://cran.rproject.org/web/packages/lme4/index.html>

Baumann, Stefan. 2006. *The Intonation of Givenness: Evidence from German*. Tübingen: Niemeyer. https://doi.org/10.1515/9783110921205

Baumann, Stefan & Riester, Arndt. 2013. Coreference, lexical givenness, and prosody in German. *Lingua* 136: 16–37. https://doi.org/10.1016/j.lingua.2013.07.012

Beckman, Mary E. 1997. A typology of spontaneous speech. In *Computing Prosody*, Yoshinori Sagisaka, Nick Campbell & Norio Higuchi (eds), 7–26. New York NY: Springer. https://doi.org/10.1007/978-1-4612-2258-3_2

Blaauw, Eleonora. 1991. Phonetic characteristics of spontaneous and read-aloud speech. In *ESCA Workshop on Phonetics and Phonology of Speaking Styles*, *Barcelona, Spain*, 30 September – 2 October. ISCA Archive. <www.isca-speech.org/archive>

Boeder, Winfried. 1991. The phrasing of Old Georgian according to scribal punctuation. In *Seventh Conference on the Non-Slavic Languages of the Soviet Union*, Chicago, May 1991, 22–24.

Boersma, Paul & Weenink, David. 1992–2016. Praat: Doing Phonetics by Computer (Version 6.0.17) [Computer program]. <http://www.praat.org/>

Brehm, Annika, Lausecker, Alina & Feldhausen, Ingo. 2014. The intonation of imperatives in Mexican Spanish. In *Proceedings of the 10th International Seminar on Speech Production (ISSP), Cologne, Germany*, Susanne Fuchs, Martine Grice, Anne Hermes, Leonardo Lancia & Doris Mücke (eds), 53–56. Cologne: University of Cologne. <http://www.issp2014.uni-koeln.de/wp-content/uploads/2014/Proceedings_ISSP_revised.pdf>

Bruce, Gösta & Touati, Paul. 1991. On the analysis of prosody in spontaneous speech with exemplification from Swedish and French. Presented at the ESCA Workshop on Phonetics and Phonology of Speaking Styles, Barcelona, Spain, 30 September – 2 October.

Clark, Herbert H. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior* 12(4): 335–359. https://doi.org/10.1016/S0022-5371(73)80014-3

De Ruiter, Laura. 2015. Information status marking in spontaneous vs. read speech in storytelling tasks – Evidence from intonation analysis using GToBI. *Journal of Phonetics* 48: 29–44. https://doi.org/10.1016/j.wocn.2014.10.008

Face, Timothy L. 2003. Intonation in Spanish declaratives: Differences between lab speech and spontaneous speech. *Catalan Journal of Linguistics* 2: 115–131.

Featherston, Sam. 2005. The decathlon model: Design features for an empirical syntax. In *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*, Stephan Kepser & Marga Reis (eds), 187–208. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110197549.187

Féry, Caroline. 2017. *Intonation and Prosodic Structure*. Cambridge: CUP.

Féry, Caroline & Truckenbrodt, Hubert. 2005. Sisterhood and tonal scaling. *Studia Linguistica* 59(2–3): 223–243. https://doi.org/10.1111/j.1467-9582.2005.00127.x

Gordon, Matthew K. 2005. Intonational phonology of Chikasaw. In *Prosodic Typology: The Phonology of Intonation and Phrasing*, Sun-Ah Jun (ed.), 301–330. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199249633.003.0011

Grice, Martine, Baumann, Stefan & Jagdfeld, Nils. 2007. Evidence for tonal identity from peak scaling under pitch span variations. In *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS), Saarbrücken, Germany*, Jürgen Trouvain (ed.), 977–980. Saarbrücken: Saarland University.

Gussenhoven, Carlos. 1984. *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.

Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511616983

Hale, Kenneth & Selkirk, Elisabeth. 1987. Government and tonal phrasing in Papago. *Phonology Yearbook* 4: 151–183. https://doi.org/10.1017/S0952675700000804

Hansson, Petra. 2003. *Prosodic Phrasing in Spontaneous Swedish*. Lund: Department of Linguistics and Phonetics.

Harris, Alice. 1981. *Georgian Syntax: A Study in Relational Grammar*. Cambridge: CUP.
Jilka, Matthias, & Möbius, Bernd. 2006. Towards a comprehensive investigation of factors relevant to peak alignment using a unit selection corpus. In *Interspeech 2006 and 9th International Conference on Spoken Language Processing*, 2054–2057. Pennsylvania PA: International Speech Communication Association (ISCA).
Jilka, Matthias & Möbius, Bernd. 2007. The influence of vowel quality features on peak alignment. In *8th Annual Conference of the International Speech Communication Association (Interspeech 2007)*, 2621–2624. Antwerp: International Speech Communication Association (ISCA).
Kempen, Gerard & Harbusch, Karin. 2005. The relationship between grammaticality ratings and corpus frequencies: A corpus study into argument linearization in the midfield of German subordinate clauses. In *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*, Stephan Kepser & Marga Reis (eds), 329–350. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110197549.329
Kepser, Stephan & Reis, Marga. 2005. Evidence in Linguistics. In *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*, Stephan Kepser & Marga Reis (eds), 1–6. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110197549.1
Koch, Karsten A. 2008. Spontaneous speech, lab speech, and effects on intonation: some useful findings for fieldworkers (and laboratory phonologists). In *Proceedings of the 2008 annual conference of the Canadian Linguistic Association*, Susie Jones (ed.), 1–14. Vancouver BC: University of British Columbia.
Labov, William. 1972. *Sociolinguistic Patterns*. Oxford: Blackwell.
Labov, William. 1984. Field methods of the project on linguistic change and variation. In *Language in Use: Readings in Sociolinguistics*, John Baugh, & Joel Sherzer (eds), 28–53. Englewood Cliffs NJ: Prentice Hall.
Labov, William. 2004. Ordinary events. In *Sociolinguistic Variation: Critical Reflections*, Carmen Fought (ed.), 31–43. Oxford: OUP.
Ladd, D. Robert. 1990. Metrical representation of pitch register. In *Papers in Laboratory Phonology, I: Between the Grammar and Physics of Speech*, John Kingston & Mary Beckman (eds), 35–57. Cambridge: CUP. https://doi.org/10.1017/CBO9780511627736.003
Llisterri, Joaquim. 1992. Speaking styles in speech research. In *Proceedings of ELSNET/ESCA/SALT Workshop on Integrating Speech and Natural Language*, Dublin, Ireland, 15–17 July.
Mahjani, Behzad. 2003. An Instrumental Study of Prosodic Features and Intonation in Modern Farsi (Persian). MA thesis, University of Edinburgh.
Milroy, Lesley & Gordon, Matthew. 2003. *Sociolinguistics: Method and Interpretation*. Oxford: Blackwell. https://doi.org/10.1002/9780470758359
Nespor, Marina, & Vogel, Irene. 1986. *Prosodic Phonology*. Dordrecht: Foris.
Nolan, Francis. 2003. Intonational equivalence: An experimental evaluation of pitch scales. In *Proceedings of the 15th International Congress of Phonetic Sciences*, *Barcelona, Spain*, Maria-Josep Solé (ed.), 771–774. Barcelona: ICPhS Organizing Committee.
Oliveira, Miguel. 2000. Prosodic Features in Spontaneous Narratives. PhD dissertation, Simon Fraser University, Ottawa.
Poplack, Shana. 1993. Variation theory and language contact. In *American Dialect Research: An Anthology Celebrating the 100th Anniversary of the American Dialect Society*, Dennis R. Preston (ed.), 251–286. Amsterdam: John Benjamins. https://doi.org/10.1075/z.68.13pop
R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Reis, Harry T. & Gosling, Samuel D. 2010. Social psychological methods outside the laboratory. In *Handbook of Social Psychology*, Susan T. Fiske, Daniel T. Gilbert & Gardner Lindzey (eds), 82–114. Hoboken NJ: Wiley & Sons. https://doi.org/10.1002/9780470561119.socpsy001003
Sampson, Geoffrey R. 2007. Grammar without grammaticality. *Corpus Linguistics and Linguistic Theory* 3(1): 1–32. https://doi.org/10.1515/CLLT.2007.001
Schilling-Estes, Natalie. 1998. Investigating 'self-conscious' speech: The performance register in Ocracoke English. *Language in Society* 27(1): 53–83. https://doi.org/10.1017/S0047404500019722
Selkirk, Elisabeth O. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge MA: The MIT Press.
Selkirk, Elisabeth O. 2011. The syntax-phonology interface. In *The Handbook of Phonological Theory*, 2nd edn, John Goldsmith, Jason Riggle & Alan Yu (eds), 435–484. Oxford: Blackwell. https://doi.org/10.1002/9781444343069.ch14
Shosted, Ryan K. & Chikovani, Vakhtang. 2006. Illustrations of the IPA: Standard Georgian. *Journal of the International Phonetic Association* 36(2): 255–264. https://doi.org/10.1017/S0025100306002659
Skopeteas, Stavros & Fanselow, Gisbert. 2010. Focus in Georgian and the expression of contrast. *Lingua* 120: 1370–1391. https://doi.org/10.1016/j.lingua.2008.10.012
Skopeteas, Stavros & Féry, Caroline. 2010. Effect of narrow focus on tonal realization in Georgian. In *Speech Prosody 2010*, Mark Hasegawa-Johnson (ed.).
Skopeteas, Stavros, & Féry, Caroline. 2014. Focus and intonation in Georgian: Constituent structure and prosodic realization. Ms, University of Göttingen and Goethe University Frankfurt.
Skopeteas, Stavros, Féry, Caroline & Asatiani, Rusudan. 2009. Word order and intonation in Georgian. *Lingua* 119: 102–127. https://doi.org/10.1016/j.lingua.2008.09.001
Stokhoff, Martin & van Lambalgen, Michel. 2011. Abstractions and idealizations: The construction of modern linguistics. *Theoretical Linguistics* 37(1–2): 1–26. https://doi.org/10.1515/thli.2011.001
Truckenbrodt, Hubert. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30(2): 219–255. https://doi.org/10.1162/002438999554048
van Donzel, Monique. 1999. Prosodic Aspects of Information Structure in Discourse. PhD dissertation, University van Amsterdam.
van Santen, Jan P. H. & Hirschberg, Julia. 1994. Segmental effects on timing and height of pitch contours. In Proceedings of *3rd International Conference on Spoken Language Processing* (ICSLP 94), Yokohama, Japan, 719–722.
Vicenik, Chad & Jun, Sun-Ah. 2014. An autosegmental-metrical analysis of Georgian intonation. In *Prosodic Typology, II: The Phonology of Intonation and Phrasing*, Jun Sun-Ah (ed.), 154–186. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199567300.003.0006
Wagner, Petra, Trouvain, Jürgen, & Zimmerer, Frank. 2015. In defense of stylistic diversity in speech research. *Journal of Phonetics* 48: 1–12. https://doi.org/10.1016/j.wocn.2014.11.001
Wertheim, Suzanne. 2003. Rethinking the observer's paradox and data "purity". In *Proceedings of the 28th Meeting of the Berkeley Linguistics Society*, Julie Larson & Mary Paster (eds), 511–521. Berkeley CA: BLS.
Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of *F*0 contours. *Journal of Phonetics* 27: 27–55. https://doi.org/10.1006/jpho.1999.0086
Xu, Yi. 2010. In defense of lab speech. *Journal of Phonetics* 38: 329–336. https://doi.org/10.1016/j.wocn.2010.04.003