Phrasing and speech rate effects on segmental and prosodic variability in Greek

Mary Baltazani1, Katerina Nicolaidis2

1University of Oxford, UK
2Aristotle University of Thessaloniki, Greece
mary.baltazani@phon.ox.ac.uk, knicol@enl.auth.gr

Abstract
This study examines how different phrasings in identical utterances in Greek are marked by pitch, pre-boundary lengthening, sandhi phenomena, and pausing, and the influence of speech rate on them. Eight speakers (4F, 4M) produced utterances in Greek containing structures of the type “noun (N) + clitic (C) + verb (V)” in two phrasings, [(N + C) (V)] (noun-final) and [(N) (C+V)] (clitic-final), in normal and fast speech rate. Results show that both phrasing and speech rate influence different measures. Specifically, in both phrasings, pre-boundary lengthening affected the phrase-final word, the phrase-final syllable, as well as the phrase pre-final syllable. Significantly higher scaling and earlier alignment of the H tone was found in the clitic-final than in the noun-final phrasing. Finally, the clitic-final phrasing induced greater degree of sandhi effects. In fast rate, sentence, word and segment durations were shorter than in normal rate. Overall, results on the effect of phrasing agree with previous literature on Greek and other languages, while novel findings on the patterns are revealed by the combined effect of phrasing and speech rate.

Index Terms: intonation, prosodic phrasing, prosodic variability, pre-boundary lengthening, s-voicing, speech rate

1. Introduction
Interfaces between prosodic organisation and other linguistic domains (morphology, syntax, pragmatics) have been established cross-linguistically [1, 2, 3]. Prosody provides information during the processing of the syntactic structure of utterances as shown in studies on adult sentence processing and infant language acquisition, e.g., [4, 5], which can be useful in the resolution of structural ambiguities in utterances with identical word strings. It has been shown that such processing involves different prosodic organization and language-specific, systematic variation of several acoustic parameters, including intonational changes, segmental lengthening, pauses and variation in sandhi phenomena.

Moreover, cross-linguistic research has shown that speech rate can also cause changes in these disambiguating cues. For example, at faster rates pitch accents have been reported to occur later in relation to the vowel they are associated with in several languages [6, 7, 8] and some boundaries may be deleted or reduced in strength [9].

Less work has been done on these issues for Greek [10, 11, 12], but it has documented the importance of intonational cues as well as pre-boundary lengthening and segmental sandhi processes in the parsing of relative clause attachment and other types of structural ambiguities. The presence of phrase-final lengthening has been established in [13, 14]. This occurs mainly in the pre-boundary syllable and in smaller degree in preceding syllables, especially the stressed ones. Boundary strength has been reported to influence the extent of pre-boundary lengthening [13, 14], pitch accent scaling and alignment [15].

Sandhi phenomena have also been examined as disambiguation cues of syntactically ambiguous sentences in Greek, among them the process of s-voicing before a voiced consonant. Recent analyses report /s/-voicing as a gradient phenomenon, whereby stronger boundaries are more likely to block it than weaker ones [16, 17, 18, 19].

The present study examines how different phrasings of constituents in utterances with identical word strings are marked by pitch, segmental processes (lengthening and sandhi) and pausing. It also examines the influence of speech rate on these parameters to study differences in their control under different temporal conditions.

Identification sentences with a clitic attachment either to the subject or the verb are examined. To our knowledge there have been no studies examining the acoustic correlates of different phrasings in these types of sentences in Greek, or the effect of speech rate differences on these correlates, despite their importance for (i) understanding the mechanisms underlying the implementation of different phrasings in disambiguating sentences with identical word strings, and (ii) understanding cue weighting in the processing of syntactic structure in adult speech and in infant language acquisition.

2. Methods
Eight native speakers of Standard Modern Greek (4F, 4M), age range 28-57, with no history of hearing or speech problems, produced sentences containing structures of the type “subject noun (N) + clitic (C) + verb (V) + object noun”. They were produced in two speech rates (comfortable/normal and fast), 3 repetitions, and two different phrasings where the first phrase ends in: (a) a clitic [(N + C) (V)], (henceforth ‘clitic-final’) or (b) a noun, [(N) (C+V)], (‘noun-final’). Table 1 presents the sentences classified in two types. In type 1, the subject noun ended in a vowel (i.e., [i]) and in type 2 in a consonant (i.e., /s/).

Type 2 sentences were included in the speech material to examine /s/-voicing before the initial voiced consonant of the clitic in the two phrasing conditions.

The written sentences were presented with brackets indicating the two different phrasings (Table 1) to the speakers who were instructed to produce them in a way which would convey the two different meanings indicated by each phrasing.

Recordings were carried out in a sound-treated studio with a Rode NT1-A cardiod condenser microphone. All annotations and segmentations were carried out in Praat [20]. Segmentation at sentence, phrase, word and segmental level was carried out. For segmental segmentation, the onset and end of each segment was marked on the basis of common segmentation criteria, e.g.
onset and end of formant structure for vowels, onset and offset of high frequency noise for fricatives, etc.

Table 1: Experimental sentences; cf= clitic-final, nf = noun-final.

<table>
<thead>
<tr>
<th>Phrasing</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>la (i ma rina mu) (do riz ‘mila)</td>
<td>My Marina cf offers apples</td>
</tr>
<tr>
<td>lb (i ma rina) (mu do ‘rizi ‘mila)</td>
<td>Marima offers nf me apples</td>
</tr>
<tr>
<td>2a (o dici yoros (ci ni to enodia ‘feron)</td>
<td>My lawyer cf is interesting</td>
</tr>
<tr>
<td>2b (o dici yoros (mu ci ni to enodia ‘feron)</td>
<td>The lawyer nf interests me</td>
</tr>
</tbody>
</table>

Most of the measurements presented below involve the subject noun and clitic, which together will be referred to as ‘test-words’ from now on. The duration of the sentence and the two test words were measured, as well as all segments from the consonant in the stressed syllable of the noun to the end of the clitic (in bold in Table 1), including the duration of /s/ in type 2 sentences. We hypothesize longer durations next to stronger (phrase) than weaker (word) boundaries, e.g., longer durations for the noun in lb compared to la. Presence or absence of pause was noted and pause duration was measured when present; we expect more and/or longer pauses next to strong boundaries. For type 2 sentences with a voiceless stop in verb-initial position ([ci ni]), the closure duration of [c] could not be distinguished from a preceding pause, so pause duration results are only reported for type 1 sentences. Final /s/-voicing duration in the subject word was measured in type 2 sentences, based on the voice report function in Praat [21]. More voicing is taken to indicate a weaker boundary, so more voicing is expected in the clitic-final than in the noun-final condition (where a word vs a phrase boundary follows /s/). Finally, the f0 peak in the vicinity of the subject noun was annotated and its f0 scaling and alignment were measured. Speakers are expected to differentiate between the two phrasings by using a continuation rise at the end of the first phrase, which would occur either at the end of the clitic (clitic-final condition) or at the end of the noun (noun-final condition) (Figure 1).

Regarding speech rate, we hypothesize that the magnitude of the durational, voicing, and intonational effects of phrasing will be reduced at the faster speech rate.

A mixed model ANOVA was conducted for each of the temporal and tone measures as dependent variables, phrasing (clitic-final, noun-final) and speech rate (normal, fast) as fixed factors, and speaker (EIR, TA, ANA, KN, TTSI, FOT, THA, TTSA) and sentence-type (1, 2) as random factors.

3. Results

3.1. Pre-boundary lengthening and speech rate

Temporal measures were influenced by both phrasing and speech rate. Sentence duration was longer in the noun-final condition (F(1, 180) = 6.91, p < 0.009), but post hoc tests for the phrasing by speech rate interaction (F(1, 180) = 4.20, p < 0.042) showed a significant difference only in the normal speech rate. Nouns were longer in noun-final than clitic-final position and clitics in clitic-final than noun-final position, as expected, due to pre-boundary lengthening (noun: F(1, 180) = 323.25, p < 0.001; clitic: F(1, 180) = 195.21, p < 0.001).

Sentences were significantly shorter in fast rate (F(1, 180) = 578.63, p < 0.001) and so were nouns (F(1, 180) = 425.52, p < 0.001) and clitics (F(1, 180) = 69.73, p < 0.001). Longer test words were found phrase-finally in both speech rates (phrasing by speech rate interaction: noun: F(1, 180) = 24. 50, p < 0.001; clitic: F(1, 180) = 17. 03, p < 0.001), (Figure 2 and Table 2).

Figure 1: Examples of a clitic-final (top) and noun-final (bottom) utterance.

Figure 2: Sentence duration by speech rate (top), noun duration (middle) and clitic duration (bottom) by speech rate and phrasing for all speakers
Table 2: Mean sentence, noun and clitic duration, in ms, by prosodic phrasing and speech rate.

<table>
<thead>
<tr>
<th></th>
<th>Clitic-final</th>
<th>Noun-final</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal rate</td>
<td>1859</td>
<td>1963</td>
</tr>
<tr>
<td></td>
<td>552</td>
<td>676</td>
</tr>
<tr>
<td></td>
<td>213</td>
<td>131</td>
</tr>
<tr>
<td>Fast rate</td>
<td>1558</td>
<td>1567</td>
</tr>
<tr>
<td></td>
<td>467</td>
<td>537</td>
</tr>
<tr>
<td></td>
<td>156</td>
<td>112</td>
</tr>
</tbody>
</table>

With reference to segmental duration (Table 3), the stressed and post-stressed vowels in the subject noun were significantly longer in noun-final than in clitic-final constructions (stressed V: F(1, 180) = 108.66, p < 0.001; post-stressed V: F(1, 180) = 217.29, p < 0.001) and so was /s/ in type 2 sentences, F(1, 85) = 103.12, p < 0.001), while for the consonants only the one in the final unstressed syllable was affected, being longer in noun-final position (F(1, 180) = 16.24, p < 0.001). The clitic consonant and vowel, on the other hand, were both significantly longer in the clitic-final than in the noun-final condition (C: F(1,180) = 24.19, p < 0.001; V: F(1,180) = 215.76, p < 0.001).

Speech rate influenced all above parameters (C in stressed syllable: F(1, 180) = 14.09, p < 0.001; C in unstressed syllable F(1, 180) = 16.69, p < 0.001; stressed V: F(1, 180) = 159.47, p < 0.001; unstressed V: F(1, 180) = 97.71, p < 0.001; /s/: F(1, 85) = 49.08, p < 0.001; clitic C: F(1, 180) = 9.79, p < 0.002; clitic V: F(1, 180) = 76.67, p < 0.001). All durations were shorter in fast speech (Table 3). The phrasing by speech rate interaction was significant for all the vowels of the noun and clitic words (stressed V: F(1, 180) = 11.42, p < 0.001; post-stressed V: F(1, 180) = 27.58, p < 0.001; clitic V: F(1, 180) = 22.57, p < 0.001) as well as for /s/ (F(1, 180) = 14.96, p < 0.001). Post-hoc tests showed significantly shorter duration in fast speech in both phrasings for all vowels. For /s/, significantly shorter duration in fast speech was found only when /s/ was in noun-final position.

Table 3: Duration of stressed C, V, post-stress C, V of the subject noun, /s/ duration and clitic vowel and consonant duration by phrasing and speech rate.

<table>
<thead>
<tr>
<th></th>
<th>Clitic-final</th>
<th>Noun-final</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal rate</td>
<td>45</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>129</td>
<td>161</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>76</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>122</td>
<td>57</td>
</tr>
<tr>
<td>Fast rate</td>
<td>35</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>108</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>78</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>77</td>
<td>44</td>
</tr>
</tbody>
</table>

For type 1 sentences where pause duration was measured, phrasing did not affect the pause duration, but speech rate did (F(1, 85) = 51.47, p < 0.001), with significantly longer pauses at normal than fast speech rates (Fig. 3 top). In fast rate, the pause was absent more frequently than in normal rate (Fig. 3 bottom).

3.2. Intonation effects

The results showed that the first phrase typically ended in a continuation rise, which has been reported as a L* H- tone in Greek [22], a L* nuclear pitch accent typically aligned with the stressed vowel, followed by an H- phrase accent. A significant main effect of phrasing, F(1, 180) = 36.27, p < 0.001, shows higher scaling for H in clitic-final than noun-final constructions (mean 251 vs 227 Hz). In addition, significantly higher scaling was present in normal than fast speech rate (F(1, 180) = 5.54, p = 0.02; mean 244 vs 235). Figure 4 presents H scaling in the two phrasing conditions and speech rates.

Figure 5 shows the mean distance of the H tone from the right boundary of the first phrase, in the phrasing and speech rate conditions for each speaker. The means for the main effect of phrasing (F(1, 180) = 7.91, p = 0.005) indicate that the H, which aligned close to the phrase boundary, was significantly later, i.e., closer to this boundary in the noun-final than in the clitic-final condition (mean -17 vs -28 ms; mean vowel duration 111 in noun-final and 100 ms in clitic-final condition). A main effect of speech rate (F(1, 180) = 5.13, p = 0.025) indicated that
the H aligned later at normal than fast rate (mean -18 vs -27). Figure 5 presents negative values for all but three speakers who show different strategies of alignment; it also presents further speaker variability across conditions.

![Graph 5](image)

**Figure 5.** H tone alignment from the end of the first phrase by speaker, phrasing and speech rate.

### 3.3. /s/-voicing

Significantly more voicing was present in the clitic-final constructions than in the noun-final ones (F(1, 85) = 143.25, p < 0.0001) as expected (mean unvoiced portion 0.125 vs 0.678). Speech rate had no effect on s-voicing (Figure 6).

![Graph 6](image)

**Figure 6.** Mean s-voicing by phrasing and speech rate.

### 4. Discussion

This paper examined the production of different phrasings of constituents in sentences with identical word strings in Greek, where a clitic can be grouped either with the previous noun or the following verb, resulting in a different interpretation. Our analysis showed that differences in phrasing are marked by multiple cues, including intonation, segmental processes (lengthening and sandhi) and pausing. In addition, our results revealed that speakers exercise fine control of these parameters under different temporal conditions of different speech rates.

In particular, in line with previous literature [1, 2, 3, 13], preboundary lengthening marked the phrase boundary, potentially serving as a disambiguating cue. Lengthening was found in the phrase-final word, the phrase-final syllable, but also the phrase pre-final stressed syllable. Vowels in both the stressed and unstressed syllables were lengthened and so was the consonant in the final unstressed syllable. Final /s/ in type 2 sentences was also longer in phrase final position. Fast speech rate resulted in shorter sentence, word, segment and pause durations. Pauses were also more frequently absent in the fast rate of production. Their absence suggests that are not a necessary boundary cue. The phrase-final lengthening cue was preserved, in fast speech rate.

Stronger boundaries were also marked by a reduction in the percentage of /s/-voicing, verifying reports in the literature that (a) /s/-voicing is a gradient, not a categorical phenomenon and (b) that stronger boundaries (in this experiment boundaries between phrases in the noun-final condition) are more likely to induce less voicing than weaker ones (boundaries between words in the clitic-final condition).

Turning to intonation, the results revealed that tone alignment and scaling varied systematically as a function of phrasing, meaning they could also potentially serve as disambiguating cues in Greek. Higher scaling was found for the H tone in the clitic-final than noun-final constructions. This is probably so due to tonal crowding pressure: there is more time for the f0 to reach its target in the clitic-final phrasing, as the end of the phrase is two syllables after the nuclear pitch accent (which is located on the stressed syllable) compared to the noun-final phrasing where the end of the phrase is one syllable away. Moreover, time pressure also played a role: significantly higher scaling was present in normal than fast speech rate, as f0 can reach its target when there is more time, i.e., at normal rather than fast speech rate. In other words, there is undershoot at fast speech rate, perhaps also a combined effect of phrasing, but there is variability.

With reference to alignment, H aligned within the phrase-final vowel: H alignment ranged from -28 to -17 ms from the boundary in the clitic-final vs noun-final condition respectively, that is 28% versus 15% into the vowel respectively. Thus, the H tone alignment was later in the noun-final condition, probably having less freedom because the L* nuclear pitch accent was nearer to the H in this condition than in the clitic-final condition.

In addition, our findings that the H aligns later at normal than fast rate do not agree with cross-linguistic reports that tones occur later in faster rates, demonstrating that such settings may be language specific and need to be examined on such a basis. They do agree, though, with reports of reduction of boundaries in faster rates, that is an association of higher scaling with stronger boundaries and lower scaling with weaker ones e.g., [9], since undershoot of H scaling was found in Greek at faster rates. In addition, the fact that pauses were absent more frequently at fast rates further corroborates the results in [9].

Perception tests are needed to determine whether these cues are actually used by listeners to disambiguate sentences and discover which cues weigh most in disambiguation. Absence of pausing in fast rate, for example, indicates that it is not a necessary cue in Greek. Further analyses are needed to estimate the contribution of different parameters in disambiguating these constructions. In addition, the production results in this study should also be investigated in natural speech. The experimental task was relatively artificial, so some of the differences that were produced between the two phrasings might be exaggerated to make the utterances distinguishable.

### 5. Acknowledgements

Special thanks are due to all the participants of the study and to George Vlahavas for assistance with the statistical analyses.
6. References