Slovenian clitics prefer to cliticize to the right

Guy Tabachnick, Franc Marušič, Rok Žaucer
University of Nova Gorica

May 8, 2026

Abstract

Slovenian second-position clausal clitics are typically described as being either enclitic by default (Golden & Milojević Sheppard 2000) or prosodically neutral (Bošković 2001). However, Orešnik (1984) argues that they are usually proclitic. In this paper, we present an experiment testing this description. Subjects were presented a series of sentences with an added beep in various positions in the vicinity of these clitics and asked to identify the location where they perceived the beep. The results confirm previous descriptions that clitics can attach in either direction if needed, but suggest that by default, Slovenian clitics are perceived as proclitic, attaching to the word following them, supporting the claims of Orešnik (1984).
keywords: Slovenian · clitics · prosody · perception experiment

1 Introduction

Slovenian clausal clitics (including pronominal and auxiliary clitics) are second-position (Wackernagel) clitics (Golden & Milojević Sheppard 2000, Franks & King 2000), meaning that they usually occur after the first constituent of the sentence, regardless of what the first constituent is, as shown in (1).[1]

(1) a. Micka **mu** **je** včeraj podarila knjigo.
Micka him.DAT AUX.2SG yesterday gave book
'Micka gave him a book yesterday.'

b. Včeraj **mu** **je** Micka podarila knjigo.
yesterday him.DAT AUX.2SG Micka gave book
'Micka gave him a book yesterday.'

c. Knjigo **mu** **je** včeraj podarila Micka.
book him.DAT AUX.2SG yesterday gave Micka
'Micka gave him a book yesterday.'

d. Podarila **mu** **je** včeraj Micka knjigo.
gave him.DAT AUX.2SG yesterday gave Micka
'Micka gave him a book yesterday.'

Unlike their Bosnian/Croatian/Montenegrin/Serbian (BCMS) counterparts, which are typically seen as being strictly enclitic (see e.g. Browne 1974, 1975, Radanović-Kocić 1988, Schütze 1994, Franks & King 2000, Bošković 2001), Slovenian clitics are considered to be freer in the specification of their attachment direction, or even as lacking specification altogether.

The ability of Slovenian clitics to prosodically attach in both directions is well-established. When they appear in the first position of a clause, as in (2-a), or follow a pause/intonational break, as in (2-c), Slovenian clausal clitics can only look for a prosodic host on their right, in which case they are clearly proclitic. However, when they occur at the end of the sentence, as in (2-b), or when they precede an intonational break/clause boundary, as in (2-d), they can only prosodically attach to the preceding prosodic word (Franks & King 2000, Golden & Milojević Sheppard 2000).

[1]The clausal clitics under discussion are bolded in the examples throughout the paper.

(2) a. **Si=** **ga=** videl?
AUX.2SG him.ACC saw
'Did you see him?'

b. Videl **=sem** **=ga**.
saw AUX.1SG him.ACC
'I saw him.'

c. Prešeren, || največji slovenski pesnik, || **se=** **je=** rodil v Vrbi.
Prešeren greatest Slovenian poet REFL.ACC AUX.3SG born in Vrba
'Prešeren, the greatest Slovenian poet, was born in Vrba.'

d. Videl **=sem** **=ga**, || ko je skočil.
saw AUX.1SG him.ACC when AUX.3SG jumped
'I saw him when he jumped.'

Given that Slovenian clitics can be either proclitic or enclitic, one question to ask is how they behave in environments where sentential prosody does not force either left or right attachment, such as in a sentence with no large prosodic boundaries or pauses either preceding or following the clitic cluster, as in (3).

(3) Metka **si** **ga** **je** dobro ogledala.
Metka REFL.DAT him.ACC AUX.3SG well looked
'Metka took a good look at him.'

The usual assumption is that clitics in examples such as (3) are enclitic (Toporišič 2000, Golden & Milojević Sheppard 2000), but Orešnik (1984) argues that Slovenian clitics should be seen as proclitic by default. In this paper, we explore this disagreement in the literature with an acoustic perception experiment: we asked participants to locate a beep artificially inserted into a recording of a sentence, under the hypothesis that listeners' perception of beeps aligns with prosodic breaks.

In what follows, §2 presents the background on pro-/encliticization and the rationale of the study in more detail, §3 presents the experimental design, §4 presents the results, §5 contains a discussion, and §6 concludes.

2 Proclitics vs. enclitics

Slovenian clausal clitics are Wackernagel clitics forming a cluster which is (typically) located after the first syntactic constituent (Toporišič 1992, Golden & Milojević Sheppard 2000). Unlike BCMS clitics (e.g. Browne 1974, Radanović-Kocić 1988), Slovenian clitics only rarely split syntactic constituents. Accordingly, Golden & Milojević Sheppard (2000) propose that they are adjoined to the first syntactic head of the clause (assuming that the first syntactic constituent occupies the first specifier). Describing the prosodic attachment of clitics, Golden & Milojević Sheppard (2000: 192) write, "they are hosted by the first [. . .] maximal projection and are enclitic to the last word in it", with the addition that "[u]nlike Serbo-Croatian clitics, Slovene clitics may also be proclitic" (Golden & Milojević Sheppard 2000: 192).

The (default) location of BCMS and Slovenian clitics has been proposed to be the head of C (see e.g. Progovac 1996 for BCMS and Golden & Milojević Sheppard 2000 for Slovenian) and it has been shown that, if prosodic requirements are satisfied, clitics can also appear in any lower head (Bošković 2001, Marušič 2008). The theory of phasal spell-out (Chomsky 2001) predicts that (at least in the default case, when clitics are located in the C phase-head) clitics should be spelled out with the specifier preceding them. This is because spell-out always takes place at the phrasal level; that is, the complement of the phase head (a phrase) gets spelled out as a unit, to the exclusion of that head and its specifier. As a result, the C head, hosting clitics, should get spelled out together with its preceding specifier as a prosodic/phonological unit with the sentence-initial constituent preceding them (assuming that heads are always linearized to the right of their specifiers, following the Linear Correspondence Axiom of Kayne 1994). Thus, clitics should normally encliticize to the preceding constituent.

It seems, therefore, that phase theory predicts that Slovenian clitics should only be enclitics. However, this is not what we see. We have already mentioned two instances where clausal clitics behave as proclitics in §1: whenever they follow a prosodic boundary, as in (2-c), and in yes–no questions, as in (2-a). Orešnik (1984) argues further that Slovenian clausal clitics are usually proclitics.

One of his arguments involves stressed/emphasized clitics. When an utterance is composed of the clitic cluster alone, as in (4), one of the clitics must get stressed (see Dvořák 2007 for abundant data of this type).[2] In such cases it is always the last clitic that gets stressed, never the first one; that is, the initial clitic (as well as any other clitic from the cluster) is procliticized onto the last one.

(4) A: **Si=** **si=** **ga=** pogledal?
AUX.2SG REFL.DAT him.ACC watched
'Did you watch it?'
B: **Sem=** **si=** **gá**.
AUX.1SG REFL.DAT him.ACC
'I did.' (based on Orešnik 1984)

This sentence shows a genuine stressed clitic, with seemingly the same syntax as other clitic clusters. We see no evidence of overt movement, nor is it possible that stressed *ga* is actually a full (non-clitic) pronoun (cf. Toporišič 2000). The full form of the third-person accusative masculine pronoun is *njéga* 'him', with stress on the first syllable, but it is ungrammatical to replace the stressed *ga* in (4) with *njéga* (Franks 2016: 97). Further, in examples such as (4), if the order of the clitics is different—that is, if the last clitic in the cluster is the 3rd person auxiliary clitic *je*—then it is this one that ends up carrying the stress and acting as the host of the preceding clitics, as in (5).

(5) A: **Si=** **ga=** **je=** pogledala?
REFL.DAT him.ACC AUX.3SG watched
'Did she watch it?'
B: **Si=** **ga=** **jé**.
REFL.DAT him.ACC AUX.3SG
'She did.'

According to Orešnik (1984), sentences like (4) and (5) provide arguments for a general preference for Slovenian clitics to attach to the right (in these cases, the two unstressed clitics attach to the following stressed clitic). However, his arguments do not necessarily generalize: sentences with regular clitic placement like (3) have overt material on either side of the clitics, while the sentences in (4) and (5) do not and must involve some *repair mechanisms*. If all three clitics in (4) and (5) are in the same head, as standardly assumed, then they are all spelled out in the same phase. In the answers (the B sentences), a requirement for each prosodic word to have stress forces one of the clitics to be stressed; the direction is determined by a general phonological preference for rightward attachment (cf. Franks 2016: 99).[3] In the questions (the A sentences), the clitics are likely in C, a phase head, with an empty (or covert) specifier. Thus, they comprise their own phase syntactically and likewise need to be a part of a prosodic word with stress. Here, though, the repair is to attach them to material spelled out in the previous phase—that is, the post-clitic word—rather than to stress one of the clitics.[4]

By contrast, in (3), the prosodic grouping inherited from the syntax is well-formed, in that the clitics

[2]Toporišič (2000: 113) notes briefly that clitics in such examples are not stressed (they lack lexical stress) but rather just emphasized (receive sentence-level stress). Orešnik (1984) does not make this distinction, nor do we, as we are not aware of any phonetic or phonological evidence suggesting a difference between the stressed clitics in (4) and (5) and syllables receiving regular lexical stress.

[3]This effect would be akin to the emergence of the unmarked in phonology (see Becker & Potts 2011): a default preference is outweighed in normal cases but emerges under particular circumstances. However, while such effects are usually argued to reflect *universal* considerations of markedness, the case of stressed clitics seems to reveal a preference specific to Slovenian.

[4]It is possible in principle that the stressed and unstressed clitics in (4) and (5) are in different syntactic positions. This is certainly the case in (i), where an adverb separates the stressed clitic from the unstressed ones (these examples, especially the first, are only likely to be encountered in the spoken language).

are spelled out together with preceding material. If the default preference for rightward attachment only emerges when phonology needs to repair the grouping from syntax, the clitics in (3) will attach leftward. On the other hand, in line with the assumption of Orešnik (1984), the phonological preference for rightward attachment may be strong enough to force rebracketing in sentences like (3). Another possibility is that the reasoning or assumptions of the phase-based syntactic account are wrong, and a different syntactic account would predict a default rightward attachment for clitics in concert with the phonology. Regardless of what turns out to be correct, ascertaining the direction of cliticization in Slovenian is not only descriptively but also theoretically relevant.

While Toporišič (2000), a reference grammar of Slovenian, states that clitics in examples like (3) above or (6-a) below are enclitics (Toporišič 2000: 65), speakers' judgments do not seem to be clear. Some speakers intuit that example (6-a) can be pronounced either as (6-b) or as (6-c), without interpretational differences between the two.[5]

(6) a. Metka **ga** **je** videla.
Metka him.ACC AUX.2SG saw
'Metka saw him.'
b. Metka **=ga** **=je** videla.
Metka him.ACC AUX.2SG saw
'Metka saw him.'
c. Metka **ga=** **je=** videla.
Metka him.ACC AUX.2SG saw
'Metka saw him.'

In this study, we address this question experimentally.

3 Experiment

The goal of this experiment is to determine whether Slovenian speakers treat clitics as enclitics (attaching them to the preceding prosodic word) or proclitics (attaching them to the following prosodic word).

The direction of clitic attachment is difficult to detect. It is a purely phonetic (or phonological) property of the clitics as it does not correspond to any difference in meaning. We test attachment direction through the location of prosodic boundaries: if clitics attach to the left (as enclitics), this means there is a prosodic word boundary separating the cluster from the following word and none between the cluster and the preceding word. On the other hand, if clitics attach to the right (as proclitics), this means there is a prosodic boundary at the start of the cluster and none at its end. The prosodic word boundary at either edge of the clitic cluster may also be the boundary of a larger (intermediate) prosodic domain, but we

(i) a. A **pa** **si** res **gá** udaril?
Q PART AUX.2SG really him.ACC hit
'But did you indeed hit him?' (verum focus)
b. A **pa** **ga** res **jé** udaril?
Q PART him.ACC really AUX.3SG hit
'But did he indeed hit him?' (verum focus)

However, these examples probably do not offer any insight into the position of clitics in (4)–(5). First of all, (i) has a verum focus interpretation, which does not seem to be the case with (4) and (5). Secondly, the fact that either a pronominal or auxiliary clitic can receive the stress and split off from the rest of the cluster under verum focus suggests that the position of the clitics in (i) is not particularly syntactic either: the last clitic, whatever it is, gets unusual stress and position. This may be the outcome of a PF operation, in which case we have further evidence for a phonological repair mechanism that stresses the last clitic in a clause when necessary. We leave (i) and similar examples for future work.

[5]Lack of clarity in speakers' intuitions about the direction of clitic attachment in examples like (6-a) is not a recent development. Škrabec (1895), arguing against the view (apparently common in the late nineteenth-century) that pronominal clitics attach to the left, says that he feels no strong intuition either way and no difference between these pronominal clitics and the verbal negation marker *ne*, which must always directly precede the finite verb and presumably attaches rightward to form a prosodic word with it. Thus, his description can also be seen as an early predecessor to Orešnik (1984)'s claim that Slovenian clitics attach by default to the right.

cannot say this for sure.

Previous methods used to detect prosodic boundaries are not applicable to our research question. For example, prosodic boundaries do show some phonetic effects, like stronger articulation of consonants at edges of prosodic units (see e.g. Fougeron & Keating 1997, Cho & Keating 2001, Cho et al. 2007) and lengthening of final syllables before prosodic boundaries (cf. Wightman et al. 1992, Tabain 2003); some researchers (e.g. Cambier-Langeveld 1997) have looked for prosodic boundaries by measuring these criteria in speakers' production. However, these effects can be quite small at low-level prosodic boundaries, especially given interaction with word stress, and may require articulatory data to detect. Other experiments have participants judge the strength or acceptability of a prosodic boundary on a scale or a slider (e.g. Gussenhoven & Rietveld 1992, Cambier-Langeveld et al. 1997, Krivokapić 2007, Krivokapić & Byrd 2012) or press a button when they perceive the end of a group of words (e.g. Simon & Christodoulides 2016). These studies generally involved manipulating factors of the audio like the length of a pause or the preceding syllable. This sort of methodology is most appropriate for detecting larger (phrase-level) prosodic boundaries and thus may not reliably detect the prosodic word boundary at the edge of the clitic cluster. Finally, studies (e.g. Scott 1982, Gollrad 2013, Petrone et al. 2017) that rely on identifying differences in meaning implied by different prosodic parses are likewise unable to distinguish clitics' direction of attachment. With these challenges in mind, we selected an older methodology, described in the next section.

3.1 Methods

We conducted a perception study designed to detect the prosodic parse that Slovenian speakers assume for clitics: given an ambiguous input, do speakers attach clitics to the previous or following prosodic word? To answer this question, we used a methodology intended to detect perceived prosodic boundaries: participants heard sentences with beeps inserted and marked the location where they perceived the beep. Although this technique has lain dormant for about half a century, many early attempts to find experimental evidence for syntactic structure (e.g. Fodor & Bever 1965, Garrett et al. 1966, Bever et al. 1969) asked participants to identify the location of a non-speech sound (a click) in running speech. This line of work claims to show that clicks and/or beeps tend to be perceived at boundaries between syntactic clauses—specifically, between units corresponding to sentences at the level of deep syntactic structure (according to the theory of transformational generative grammar prevalent at the time). However, Reber & Anderson (1970) and Reber (1973) argue that these experiments' results are due primarily to non-linguistic factors like attention, memory, and response bias rather than syntactic structure.

Much of the interpretation of the earlier studies is grounded in now-outdated assumptions about syntax and the syntax–phonology interface. In particular, while attempts are made to tease apart syntactic structure and prosodic effects like intonation (Garrett et al. 1966, Geers 1978), there is no mention of prosodic structure as such. From a modern perspective, an utterance has higher-level prosodic structure that is closely related to its syntactic structure but deviates from it; we assume that listeners impose prosodic structure, not just syntactic structure, upon heard utterances, whether or not this prosodic structure is phonetically detectable in the input (cf. Fodor 2002). Thus, we hypothesize that speakers are more likely to perceive beeps at prosodic boundaries in their constructed prosodic parse; like other prosodic effects, this should be stronger at larger prosodic boundaries.

3.1.1 Participants

A total of 50 participants were recruited through Prolific. We selected for participants who indicated that their first language was Slovenian; all were born and educated in Slovenia except for one participant born in Croatia and one who attended high school in North Macedonia. Their ages ranged from 20–54, with a median of 25 and a mode of 23. They were each paid 4 euros for completing the study. One participant was removed due to technical issues, so results include data from 49 participants.

3.1.2 Experimental conditions

The stimuli were divided into five experimental conditions that varied according to the material near the clitic cluster, as shown in Table 1 (the element labelled *noun* after the clitics is an adverb in a few sentences). The division between each element is marked with its expected prosodic boundary. In most cases, each element contains one prosodic word (or possibly more, in the case of the *rest* at the end of each sentence), delineated by a | boundary. The relative and adjunct clauses should be bounded on both sides—or, at the very least, on the side next to the clitics—with a larger phrase boundary, which is marked with || (the exact level of this boundary is left undetermined). In the three test conditions (DCD, MCD, and DCM—named mnemonically, with C, D, and M standing for Clitics, one Disyllabic word, and two Monosyllabic words, respectively), the direction of cliticization is unknown, so the clitics are surrounded with ?, indicating that one of their edges should contain a prosodic word boundary and the other, no such boundary. In conditions |C and C|, the direction of cliticization is known: since the subordinate clause forces a larger prosodic boundary across which clitics cannot attach to a prosodic word, they must attach in the other direction, so there should be no prosodic word boundary after the clitics in condition |C or before the clitics in condition C| (their names are mnemonic devices for their known boundaries, and thus their direction of cliticization).

condition	sentence			
DCD	noun	? clitics ?	noun	| rest
MCD	modifier | noun	? clitics ?	noun	| rest
DCM	noun	? clitics ?	modifier | noun	| rest
|Ca	noun || relative clause	|| clitics	noun	| rest
|Cb	noun || relative clause	|| clitics	modifier | noun	| rest
C|a	noun	clitics ||	adjunct clause	|| rest
C|b	modifier | noun	clitics ||	adjunct clause	|| rest

Table 1: Experimental conditions, with expected prosodic boundaries (| prosodic word, || larger prosodic phrase, ? unknown)

Each condition contained six sentences; conditions |C and C| were divided into two subconditions with three sentences each (all experimental sentences are given in the supplementary materials - see Data section at the end). Each clitic cluster contained two clitics, a one-syllable pronominal clitic (e.g. the third person singular accusative masculine clitic *ga*) and one auxiliary clitic. Of these, the third-person singular past and future auxiliaries (*je* and *bo*, respectively) follow the pronominal clitic, while others (e.g. first-person singular past *sem* and third-person plural past *so*) precede it. The nouns had two syllables, with stress adjacent to the clitic cluster—that is, nouns with final stress like *bombáš* 'bomber' appear before the clitic cluster in conditions DCD, DCM, and C|a, while nouns with initial stress like *sáblje* 'sabres' appear after the clitic cluster in conditions DCD, MCD, and |Ca. The modifier–noun pairs in conditions MCD, DCM, |Cb, and C|b comprised two monosyllabic words, like *bel prt* 'white tablecloth' or *dva dni* 'two days' (in the names of the test conditions, D and M stand for disyllabic and monosyllabic, so e.g. condition DCM has one disyllabic word before the clitic cluster and two monosyllabic words after it). This configuration ensures that the noun and modifier–noun conditions do not differ in their stress pattern: in the three test conditions, the clitic cluster always has adjacent stressed syllables (which are substantially longer than unstressed syllables) on both sides. The content contained in the *rest* at the end of each sentence varied somewhat by condition, but always included the verb and at least one other phrase (in that order, in the test conditions; see supplementary materials).

Each sentence appeared with beeps in nine different locations: six target locations and three fillers. For the target locations, beeps were placed at the midpoints of the syllables preceding and following the clitic cluster and four evenly spaced points in between these points (see §3.2.2 for a concrete example). The filler beep locations varied by condition, and are shown schematically in Table 2.

condition	sentence			
test	NP	clitics	NP	$_{F1}$ V $_{F2}$ XP $_{F3}$
|C	NP $_{F1}$ [C clitic(s) $_{F2}$ X(P) $_{F3}$ X(P)]	clitics	NP	X(P) X(P)
C|	NP	clitics [C clitic(s) X(P) X(P)]		$_{F1}$ X(P) $_{F2}$ X(P) $_{F3}$ XP

Table 2: Experimental conditions with locations of filler beeps (F1–3), with subordinate clauses in brackets (NP is a one- or two-word nominal phrase, XP is a phrase, V is the main verb, X(P) is a phrase or inflected verb, C is a complementizer or relativizer)

3.1.3 Materials

The stimuli comprised recordings of thirty sentences, listed in an appendix in the supplementary materials, spoken by a 32-year-old female native speaker of Slovenian (from Western Slovenia) with experience as a radio presenter. She was instructed to read the sentences with neutral intonation and without pauses where possible. None of the sentences have identifiable periods of silence at either side of the clitics except in the case of voiceless stops (in which case it is usually difficult to tell whether the length of the silence exceeds the expected length of a stop closure). There are other possible cues that occasionally yield the perception (in the authors' subjective view) of a prosodic boundary in a particular location, in particular intonation and phrase-final lengthening. Such cases are not very common, though perceptually, the sentences in condition C| had more of a break after the clitics than the others, in line with our results; indeed, presaging our results, it was particularly difficult to get our speaker to pronounce them otherwise. In general, visual inspection of results for all sentences (see supplementary materials) does not indicate that sentences for which the natively Slovenian-speaking coauthors judged a more perceptible prosodic break had largely different results from others in the same condition. We did not modify the stimuli, as we wanted them to be naturalistic; we discuss the implications of these cues on the interpretation of our results in §5.

The resulting recordings were processed in version 6.3.17 of Praat (Boersma & Weenink 2023). First, ambient noise was removed. All sentences were normalized to have an average intensity of 70 dB. Next, the first author (not a native Slovenian speaker) segmented the sentences into syllables. Ambiguity in segmenting the crucial junctures (within and at the edge of the clitic cluster) was reduced by ensuring that these junctures had at least one non-glide consonant: thus, there were no clusters like *mi je* 'me.DAT AUX.3SG', where the cluster-internal juncture consists of a vowel directly followed by a glide, nor sequences like *mama jo bo* 'mother her.ACC AUX.FUT.3SG', where a glide at the start of the cluster directly follows a vowel, etc. Finally, beeps were inserted (by a Praat script) into the sentences centered at the locations described in §3.1.2. The beeps were pure tones with a frequency of 500 Hz, a length of 100 ms with fade-in and fade-out times of 10 ms each, and an amplitude of .9 Pa, equivalent to an intensity of 90.05 dB (used for normalization; participants were free to adjust the volume to their preferred level).

All recordings, Praat TextGrids, and scripts are available in the supplementary materials.

3.1.4 Experimental task

The experiment was run on PCIbex (Zehr & Schwarz 2022). In each trial, participants heard an audio file of a speaker reading one of the 30 sentences with an inserted beep. Half a second after the completion of the recording, a series of radio buttons appeared. Participants had to indicate where they heard the beep by selecting a button, which was associated with a syllable or a boundary between two syllables (see Figure 1, which shows a sentence from condition DCD). Spaces between words and punctuation (other than the period at the end of the sentence) were removed; syllable boundaries within words were placed according to the principles in Toporišič (1992: 377): for word-internal consonant clusters, the syllable boundary is placed such that the second syllable begins with the (first) consonant of the lowest sonority.[6]

[6] Toporišič (1992: 377) gives the following examples: *ka*[w]-*ra* 'rutabaga', *pra*[w]-*da* 'lawsuit', *moj-ra* 'Fate', *o-tka* 'tool for cleaning plowshares', *sla-dka* 'sweet'.

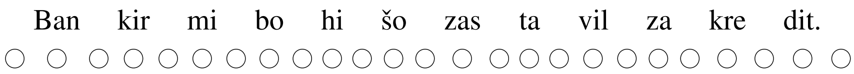


Figure 1: Sample experimental prompt, in which participants had to select where they heard the beep in the sentence *Bankir mi bo hišo zastavil za kredit.* 'The banker will mortgage my house for a loan.', with clitic cluster *mi bo* 'me.DAT FUT.3SG'

After a short training period, each participant saw 90 trials, divided internally into three blocks of thirty (the blocks were not separated by a break). Each block contained each of the thirty sentences once—thus, each participant heard each sentence three times in total. Within each of the five experimental conditions, the six sentences were divided randomly into three groups of two as shown in Table 3.[7] In each third (block) of their experiment, a participant would hear each of the sentences in each condition once with six different beep locations, two each from groups A, B, and C.

group	target beeps	filler beeps
A	1, 4	1
B	2, 5	2
C	3, 6	3

Table 3: Division of sentences within each condition by beep location

Altogether, this block-and-grouping structure ensured that each participant heard two sentences from each of the five conditions with each of the nine beep locations—that is, the distribution of stimuli was balanced across condition and beep location. However, sentences were assigned into the three groups in Table 3 for each participant separately and randomly; we did not balance the assignment of sentences into groups *across* participants. This means that the number of times each individual sentence was assigned to each group (that is, was used with a given beep location) could vary (minimum 9, maximum 23, mean 16.33, standard deviation 3.15). We do not believe that this sentence-level imbalance had any substantive effect on the results.

3.2 Analysis

3.2.1 Measuring perceptual attraction

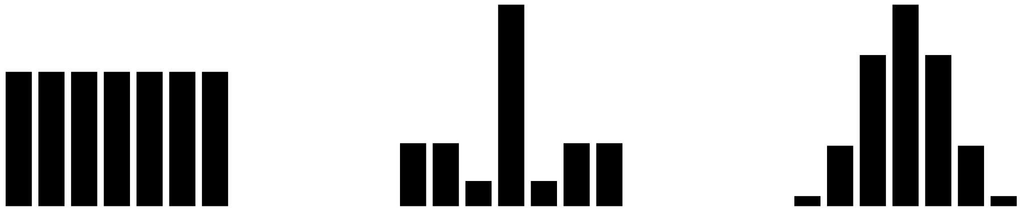
The goal of our analysis is to determine the extent to which beeps are perceived at our target locations—the start of the cluster (where we expect a large prosodic boundary in condition |C) and the end of the cluster (where we expect a large prosodic boundary in condition C|).

The null hypothesis is that prosody has no effect on beep perception—that is, speakers hear beeps where they actually are, or have some other bias unrelated to linguistic structure. This is the situation depicted in Figure 2a: the prosodic boundary has no effect on beep perception.

If we do find an effect that beeps are more likely to be perceived at prosodic boundaries—in other words, prosodic boundaries are *perceptual attractors* of beeps—there are at least two possible explanations for why. First, beeps in the middle of a word or a prosodic constituent are intrusive, so hearers might snap them to boundaries where they cause less of an interruption. If so, we would expect a steady distribution of beeps around the boundary with a large spike at the boundary itself. There may even be a dip in locations near the boundary, as hearers should be more likely to snap a beep to a boundary when it is close to it. This distributional pattern, which we call *absolute attraction*, is shown in Figure 2b.

Second, hearers may be more likely to perceive beeps at boundaries because boundaries and the syllables near them are perceptually stronger: these boundaries are often marked by pauses and lengthening of adjacent syllables, which speakers expect to be longer than others (Gussenhoven & Rietveld 1992).

[7] In conditions |C and C|, which each had two subconditions, the three groups each contained one sentence from subcondition a and one from subcondition b.



(a) No attraction (b) Absolute attraction (c) Relative attraction

Figure 2: Expected distributions of perceived beeps around a prosodic boundary for no, absolute, and relative attraction

Accordingly, the salience of prosodic boundaries and their surroundings may manifest as (real or perceived) length, so there will be more "perceptual time" for the beeps to fall into. In this case, we would instead expect beeps to be perceived more often near the boundary than far away from it, with nearby syllables also getting a boost. We call this distributional pattern *relative attraction*, and it is shown in Figure 2c.

We quantified perceptual attraction by measuring the distances of a response (that is, the perceived location of a beep) from a particular landmark. This measurement is most sensitive to relative attraction effects, in which responses get more frequent as we get closer to a boundary. It is robust to slight deviations in the center of the distribution: if the distribution of responses peaks adjacent to a boundary rather than on it, we should still find a distance effect for the boundary, even if it is less strong than expected. Measuring distance does also capture absolute attraction effects, in which a large number of responses are clustered a distance of 0 from the boundary.

Alternately, we can measure the likelihood of a response at a particular landmark. This most clearly captures absolute attraction effects, in which responses at a boundary are much more likely than responses elsewhere. While it should also capture relative attraction effects (responses are still somewhat more likely at a boundary than elsewhere), it is not robust to deviations in the center of the distribution. The supplementary materials include models with a binary dependent variable (response at the landmark or not); their results are broadly similar to those presented below.

3.2.2 Statistical models

We fitted two mixed models measuring the absolute value of the distance of responses (that is, perceived beeps) from different landmarks: the start and end of the clitic cluster, respectively. Distance is measured in syllables—for example, a response on the first clitic is a distance of .5 syllables from the start of the clitic cluster, as is a response on the last syllable before the cluster. Similarly, the boundary between the two clitics is 1 syllable from both the start and end of the cluster, while the start of the syllable before the cluster is 1 syllable from the start of the cluster and 3 syllables from its end. All of these models include trials with the six target beep locations and exclude trials with the three filler beep locations. They were constructed in version 4.3.1 of R (R Core Team 2023): using the `lmer` function from 3.1-3 of R's *lmerTest* package (Kuznetsova et al. 2017), which augments the function of the same name from *lme4* by using Satterthwaite's method to estimate p values for fixed effects in mixed linear regressions.

Both regressions have a fixed effect of condition and fixed linear and quadratic effects of actual beep location. The maximal models for our experimental design would include by-sentence and by-participant random slopes for the quadratic effect of beep location, a by-participant random slope for condition, and correlations between random effects. As Barr et al. (2013) show, omitting experimentally justified random effects can lead to false positives among fixed effects. However, the maximal models have convergence and singularity issues; thus, following the suggestions in Bates et al. (2018), we reduced the random effects to yield more parsimonious models with no overfitting issues. Both models have random intercepts for sentence and participant, by-sentence and by-participant random slopes for the linear effect of beep location, and correlations for participant; in addition, the second model (measuring distance from the end of the cluster) has a by-sentence random slope for the quadratic effect of beep location. None of the changes to the random effects substantially change the fixed ones—thus, our results are not false positives due to an insufficiently specified model.

Condition is dummy-coded with condition Cl as the reference level, as this is predicted to differ from the test conditions by the experimental hypothesis (see §3.2.3). We additionally calculated estimated marginal means for individual conditions and pairwise comparisons between them using the `emmeans` and `pairs` functions from version 1.9.0 of R's *emmeans* package (Lenth 2023).

Unlike the dependent variable, beep location is coded as *signed* distance from the landmark. As described in §3.1.2, target beeps 1 and 6 are centered at the midpoints of the syllables preceding and following the clitic cluster, respectively; these are straightforwardly .5 syllables before and 2.5 syllables after the start of the clitic cluster ($-.5$ and 2.5 away from the landmark in signed distance). Target beeps 2–5 are evenly spaced between these two landmarks, but the syllables themselves are of different lengths—for example, the stressed syllables usually preceding and following the clitic cluster are often twice as long, or longer, than each clitic within the cluster. This length also varies somewhat from sentence to sentence. Accordingly, the locations of these beeps were calculated according to their position within a syllable, as shown in Table 4.

syllable	time of boundary (s)
⋮	
	0.305
bor	
	0.574
ti	
	0.701
bo	
	0.836
da	
	1.041
⋮	

beep	time (s)	distance from start of cluster (syllables)
1	0.439	–0.500
2	0.539	–0.129
3	0.639	0.514
4	0.739	1.283
5	0.839	2.014
6	0.939	2.500

Table 4: Syllable boundary locations and target beep locations and codes for the stimulus sentence *Suh bor ti bo danes polomil vrtno ograjo.* 'A dry pine will break your garden fence today.'

The beeps in the sentence in Table 4 are spaced .1 seconds apart between the middle of *bor* and the middle of *da*. However, the codes are not evenly spaced: the difference between the codes for beeps 1 and 2 is $-.129-(-.5)=.371$ syllables, while for beeps 2 and 3, the difference is $.514-(-.129)=.643$ syllables. This is because of the difference in syllable lengths: .1 seconds is 37% the length of *bor*, but 79% the length of *ti*. This distortion reflects the experimental task (see Figure 1), in which each syllable was given a single button despite being of different lengths.

This mismatch in syllable length leads to some variation in the location of the beeps relative to the clitic cluster. Beeps 1 and 6 are fixed and beeps 3 and 4 are always in the middle of the first and second clitic, respectively. Beep 2 usually precedes the start of the clitic cluster, but in five sentences, it is placed within the first clitic. Likewise, beep 5 is usually within the syllable following the clitic cluster, but in ten sentences, it is instead within the second clitic.

The linear and quadratic term of beep location together define a parabola whose inflection point represents where the dependent variable is at a minimum (see supplementary materials for details). This captures the fact that responses are expected to be close to the actual beeps, so beeps placed closer to a given landmark should be perceived as closer to that landmark. Since the inflection point is calculated from the parameters, it is robust to an overall shift in perception—for example, if beeps are overall perceived as occurring earlier than their actual placement (as we find in our study), beeps placed slightly *after* a landmark should be perceived as closer to that landmark than beeps placed at the landmark itself, meaning that the inflection point will be after the landmark rather than at it.[8]

[8] If beep location had been encoded as the absolute value of the distance from the landmark rather than signed distance, we would not have needed the quadratic term but it would not have been robust to an overall shift in perception, since absolute value is symmetric around the landmark itself.

We removed 29 outlier responses that were more than 3 syllables away from the clitic cluster. There are two reasons to do this. First, such responses are very far away from the actual beeps and are thus less likely to reflect legitimate perception of beeps (for example, they may represent random choice due to a lapse in attention); thus, it would be inappropriate to include these tokens, which in fact have an outsized effect on the model given their greater distance from the landmarks. Second, including outliers 2 syllables or more from the start of the cluster has an outsized effect for condition |C, since the other conditions only have 2 syllables before the cluster. Removing the outliers allows for better comparison across conditions. In this light, the question arises why we put the cutoff at 3 syllables from the cluster rather than 2. As shown in Figure 3 below, condition |C does have a few responses 2 to 3 syllables before the cluster, then tails off fully. Thus, 3 syllables before the cluster seems to be the earliest point of the distribution for "legitimate" responses, and it would perhaps be inappropriate to cut off part of the distribution.

All fixed effects in the final models significantly improve the model's fit according to an F-test, calculated with the `anova` function from R's *lmerTest* package; all random effects except for the random intercept of sentence in the first (start-of-cluster) model and the by-sentence random slope for the linear effect of beep location in the second (end-of-cluster) model significantly improve the model's fit according to a chi-squared test, calculated with the `ranova` function from *lmerTest*. Each model's marginal (fixed effects only) and conditional (including random effects) R^2 was calculated using the `r2` function from version 0.10.4 of R's *performance* package (Lüdecke et al. 2021).

All models and calculations are available in the supplementary materials.

3.2.3 Hypotheses

Our experiment is designed to test two types of hypotheses. The main question, of course, is whether clitics by default attach prosodically to the preceding or following word (we call this the experimental hypothesis). However, since our experimental task of beep perception is somewhat novel, we also check whether it is in fact doing what we expect (that is, that task effects reflect the presence of prosodic boundaries) through a number of task hypotheses.

Conditions |C and C| have an unambiguous prosodic parse: condition |C has a large prosodic boundary at the start of the cluster and no boundary at the end of the cluster, while condition C| has the opposite configuration. The most basic task hypothesis—for which we find support in our experiment—is that we can detect this difference between a large boundary and no boundary (here and throughout, $X < Y$ means that beeps are perceived a *smaller distance* from X than Y—that is, that X is a greater perceptual attractor of beeps than Y):

(7) *Task hypothesis 1 (large boundary < no boundary)*
Beeps are perceived as closer to the start of the clitic cluster in condition |C than in condition C|, and closer to the end of the clitic cluster in condition C| than in condition |C.

If the task hypothesis in (7) were incorrect, we would expect no effect of condition at all; the supplementary materials contain a regression limited to trials in conditions |C and C|, the clearest evidence for an effect of condition confirming the task hypothesis and rejecting the null hypothesis.

Our experimental hypothesis is that clitics in the test conditions (DCD, MCD, and DCM) attach to the following word. This is only testable by our experiment under the assumption that the task is sensitive to prosodic boundaries, i.e. that we find evidence supporting the task hypothesis in (7). The experimental hypothesis is operationalized as the test conditions patterning with condition |C:

(8) *Experimental hypothesis*
Beeps are perceived as closer to the start of the clitic cluster in the test conditions than in condition C|, and closer to the end of the clitic cluster in condition C| than in the test conditions.

If we also find support for the experimental hypothesis in (8), this is evidence that the test conditions have a prosodic boundary at the start of the clitic cluster (this is what we find in our experiment). However, this boundary is expected to be smaller than the boundary at the start of the cluster in condition |C, since the latter is at a clause edge. Our second task hypothesis is that beep perception is sensitive not just to the *presence* of prosodic boundaries, but also to their *size*. If this is the case, and the previous two hypotheses are also confirmed, condition |C should trigger a greater task effect than the test conditions:

(9) *Task hypothesis 2 (large boundary < smaller boundary)*
Beeps are perceived as closer to the start of the clitic cluster in condition |C than in the test conditions.

This hypothesis is partially supported by our results.

Finally, our three test conditions were designed to test an additional task hypothesis. There is a small (prosodic word) boundary between the two monosyllabic words preceding the cluster in condition MCD and, likewise, between the two monosyllabic words following the cluster in condition DCM. By contrast, there is no prosodic boundary between the syllables of the disyllabic words on either side of the cluster in condition DCD, before the cluster in condition DCM, and after the cluster in condition MCD. Thus, we test whether the task is sensitive enough to detect the attractive effect of a much smaller prosodic boundary:

(10) *Task hypothesis 3 (small boundary < no boundary)*
Beeps are perceived as closer to the start of the syllable before the clitic cluster in condition MCD than in conditions DCD and DCM; likewise, beeps are perceived as closer to the end of the syllable after the clitic cluster in condition DCM than in conditions DCD and MCD.

Unlike the previous hypotheses, (10) was not supported by the experiment. It also differs from the others in that it requires looking at different landmarks (one syllable before and after the cluster). As task hypothesis 3 is secondary to the main question about the attachment of clitics, and also because of space restrictions, we discuss the results of this hypothesis in the supplementary materials; for the other hypotheses, the three test conditions are predicted to show the same behavior.

The results predicted by our hypotheses for the start and end of the clitic cluster—that is, excluding (10)—are shown schematically in Table 5.

landmark	prediction	hypothesis
start of clitic cluster	|C < C|	(7) large boundary < no boundary
start of clitic cluster	|C < DCD, MCD, DCM	(9) large boundary < smaller boundary
start of clitic cluster	DCD, MCD, DCM < C|	(8) clitics lean to the right
end of clitic cluster	C| < |C	(7) large boundary < no boundary
end of clitic cluster	C| < DCD, MCD, DCM	(8) clitics lean to the right

Table 5: Experimental predictions by landmark and condition ($X < Y$ means that beeps are expected to be perceived a *smaller distance* from a given syllable boundary in condition X than condition Y—that is, that the boundary is a greater perceptual attractor of beeps in condition X than condition Y)

The comparisons between condition C| and the other four conditions–that is, the experimental hypothesis and task hypothesis 1 (large boundary < no boundary) are tested by the fixed effect of condition in the two models. Task hypothesis 2 (large boundary < no boundary) is tested by comparison of the estimated marginal means of condition |C and the test conditions in the model measuring response distance from the start of the cluster.

4 Results

4.1 Descriptive summary

Figure 3 shows responses for target stimuli (beeps 1–6) by condition. Vertical lines mark the junctures between syllables; responses of syllables fall between these lines. The black lines indicate the prosodic boundaries expected in each condition, as shown in Table 1. Condition |C has a thick black line at the start of the clitic cluster representing the larger prosodic boundary at the end of the relative clause; condition C| has a similar thick line marking the border between the clitic cluster and the following adjunct clause. The thinner black lines mark lower-level prosodic boundaries: following the post-clitic noun in the test conditions and |C, and in between the adjective and noun in condition MCD (preceding the cluster) and DCM (following the cluster). In conditions |C and C|, the subconditions are collapsed, so no prosodic word boundary is shown between the adjective and noun in conditions |Cb and C|b. Finally, the clitic clusters themselves are delineated by dashed lines in the test conditions, indicating that the location of the prosodic word boundary is unknown. In the test conditions and |C, only two syllables (a two-syllable noun or one-syllable adjective and noun) precede the clitic cluster, so the gray rectangle delineates the beginning of the sentence. Figure 3 includes the area within 3 syllables of the clitic cluster; 29 outliers appearing earlier (possible in condition |C only) or later (possible in all conditions) are not pictured.

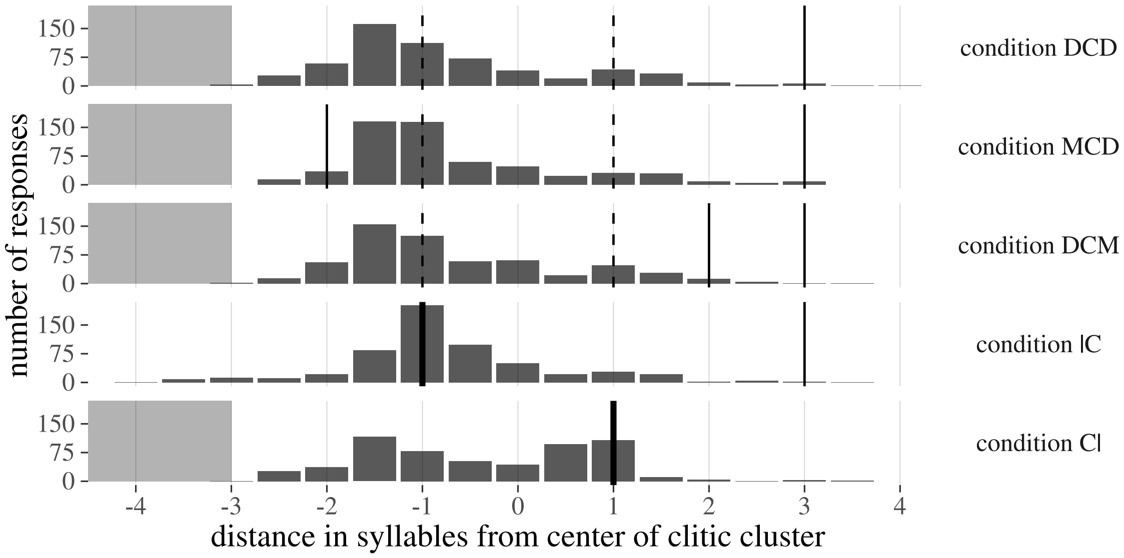


Figure 3: Experimental responses by condition for target beeps, with expected prosodic boundaries according to Table 1 (with 29 outliers removed)

Figure 4 presents the same results further separated out by beep. To make this figure more readable, the test conditions have been grouped together and some of the vertical lines have been removed.

The clearest pattern is in condition |C, where we see a strong pattern of relative attraction as defined in §3.1.4: a distribution centered around the start of the prosodic boundary at the start of the clitic cluster. In Figure 4, we see that this attraction pattern is strongest in beeps 2–4 (and especially beep 3), where beeps were perceived most often at the start of the cluster. Beeps 1 and 5 show the peak on the syllables preceding and following the start of the cluster, respectively, while beep 6 shows a more distributed pattern with smaller peaks around the end of the cluster and the syllable following the start of the cluster.

In Figure 3, the test conditions pattern similarly to condition |C, though their main peak is on the syllable preceding the start of the cluster rather than the boundary itself. As we see in Figure 4, this peak is concentrated on beeps 1 and 2; as we move along to beeps 3–5, the highest peak is at the start of the cluster itself. Meanwhile, beeps 5–6 pattern more like beep 6 of condition |C, with a second peak at or just following the end of the cluster.

Condition C|, in turn, has a starkly different pattern from the other four: Figure 3 shows a pattern with a much smaller peak at the end of the cluster and another, similarly sized one at the syllable before the start of the cluster. In Figure 4, we see that in beeps 4–6, condition C| roughly mirrors condition |C: a sizable peak on the end of the cluster in beep 6, increased responses on the syllable preceding the end of the peak in beep 5, and a second peak appearing near the start of the cluster in beep 4. On the other hand, for beeps 1 and 2, the pattern in condition C| looks more like that of the test conditions, with strong

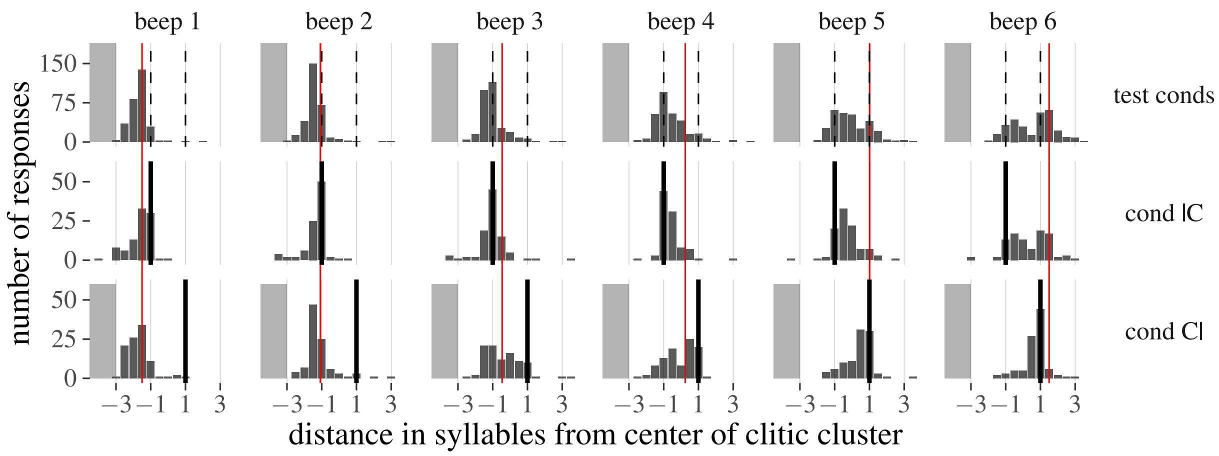


Figure 4: Experimental responses by condition and beep for target beeps, with expected prosodic boundaries according to Table 1 in black and average beep locations in red (with 29 outliers removed)

peaks on the syllable preceding the start of the cluster.

Given that conditions |C and C| have large boundaries on opposite sides of the cluster, it is striking that their results are not mirror images: in Figure 3 condition |C has one peak, while condition C| has two. Figure 4 clarifies the situation: as mentioned above, conditions |C and C| seem to be reflections of each other, but the center is around beep 5, not in the middle of the cluster (between beeps 3 and 4) as expected. This suggests that the distribution of perceived beeps is shifted from their actual location, so that a beep near the end of the clitic cluster (that is, beep 5) instead behaves as if it were in the middle of the cluster.

Indeed, there is a persistent pattern of beeps being perceived earlier than they actually occurred. This is shown in Table 6, which shows the mean difference between a beep's actual and perceived location for each beep in the test conditions (with no outliers removed). This leftward shift is smallest at beeps 1 and 2 (where perception is pulled to the right by the prosodic boundary in the cluster and stopped from going too far left by the start of the sentence) and filler beep 3, which is at the very end of the sentence (a particularly salient location); it is largest towards the end of the cluster (where perception appears to be pulled to the left towards the start of the cluster).

	beep	location	mean leftward shift
target	1	middle of pre-clitic syllable	0.15
	2	near beginning of first clitic	0.27
	3	near middle of first clitic	0.56
	4	near middle of second clitic	0.86
	5	near end of second clitic	1.12
	6	middle of post-clitic syllable	0.87
filler	1	before main verb	0.91
	2	after main verb	1.04
	3	end of sentence	0.29

Table 6: Mean leftward shift (number of syllables by which a beep's perceived location precedes its actual location) by beep for test conditions

4.2 Landmark: start of the clitic cluster

The model presented in Table 7 predicts the distance (in absolute value) of a perceived beep from the start of the clitic cluster. This model has a marginal R^2 of .216; once the random effects are considered, the marginal R^2 goes up to .346. In both models, a *positive* coefficient represents *greater* distance from the start of the cluster, so *negative* coefficients represent *stronger* attraction to that landmark.

	β coef	SE	t value	estimated p
Intercept	**0.78**	**.04**	**19.01**	**<.0001**
Beep location				
linear	**–0.21**	**.05**	**–4.67**	**<.0001**
quadratic	**0.25**	**.01**	**16.92**	**<.0001**
Condition (default: C|)				
DCD	**–0.18**	**.05**	**–3.63**	**.0013**
MCD	**–0.31**	**.05**	**–6.13**	**<.0001**
DCM	**–0.25**	**.05**	**–4.93**	**<.0001**
|C	**–0.36**	**.05**	**–7.05**	**<.0001**

Table 7: Fixed effects of mixed linear regression predicting the distance of experimental responses from the start of the clitic cluster

Beeps were perceived as further from the start of the clitic cluster in condition C| than in any other condition; this difference is significant for all other conditions. Moreover, the model in Table 7 is significantly better than one without condition as a factor ($\chi^2 = 36.00$, $p < .001$). Comparison of estimated marginal means finds that condition |C and condition DCD are significantly different (estimated $p = .016$), but condition |C and the other two test conditions are not (estimated $p = .846$ for condition MCD and $p = .233$ for condition DCM). None of the differences between the test conditions are significant (estimated $p \geq .136$).

The quadratic effect of actual beep location is positive, indicating that the distance of responses from the start of the cluster is smallest for beeps at a certain location and greater as the beep gets further away from that location. This location is .42 syllables after the start of the cluster (see supplementary materials for calculation), reflecting the slight leftward shift (beeps are perceived earlier than they actually occur) discussed in §4.1.

4.3 Landmark: end of the clitic cluster

The model in Table 8 predicts the distance of responses from the end of the cluster—that is, the strength of the end of the cluster as a perceptual attractor. This model is a substantially better fit than the equivalent model for the start in Table 7, with a marginal R^2 of .480 and a conditional R^2 of .599.

	β coef	SE	t value	estimated p
Intercept	**0.66**	**.08**	**8.42**	**<.0001**
Beep location				
linear	**–0.49**	**.03**	**–14.57**	**<.0001**
quadratic	**0.05**	**.02**	**2.90**	**.0044**
Condition (default: C|)				
DCD	**0.66**	**.09**	**7.41**	**<.0001**
MCD	**0.65**	**.09**	**7.24**	**<.0001**
DCM	**0.54**	**.09**	**6.04**	**<.0001**
|C	**0.65**	**.09**	**7.30**	**<.0001**

Table 8: Fixed effects of mixed linear regression predicting the distance of experimental responses from the end of the clitic cluster

As expected, participants perceived beeps significantly closer to the end of the cluster in condition C| than in any other condition. Comparison of the estimated marginal means confirms that this condition stands out from the others: the test conditions and condition |C were each significantly different from condition C| (estimated $p < .001$) and not significantly different from one another (estimated $p \geq .652$). Again, this model is a substantially better fit than one without condition as a factor ($\chi^2 = 42.47$, $p < .001$).

The linear and quadratic factors for beep location indicate that beeps are perceived as closest to the end of the cluster when they are 5.30 syllables after it. As shown in Table 6, even the latest target beep is perceived, on average, before the end of the cluster. Thus, the "ideal" location for a beep to be perceived at the end of the cluster is later than any in the experiment, so the frequency of responses at the end of the cluster is expected to monotonically increase. This can be captured with a linear factor. A peak of 5.30 syllables after the end of the cluster is thus merely notional, and the quadratic effect must reflect some characteristic of the data other than symmetry around a point.

5 Discussion

Our primary finding is that condition C| behaves differently from the other four conditions at both ends of the cluster: beeps are perceived as being further from the start of the cluster and closer to its end in condition C| than from the others; furthermore, model comparison shows that the experimental task is highly sensitive to condition. Our experiment thus provides statistical support for task hypothesis 1 (large boundary < no boundary): the clause boundary at the end of the cluster in condition C| attracted more beeps than the lack of boundary at that location in condition |C, and likewise, the clause boundary at the start of the cluster in condition |C attracted more beeps than the equivalent location in condition C|, where there is no prosodic boundary.

We also find support for the experimental hypothesis: the test conditions likewise had stronger perceptual attraction to the start of the cluster and weaker attraction to the end of the cluster than condition C|, suggesting that the test conditions have a prosodic boundary at the start of the clitic cluster—that is, that clitics prosodically attach by default to the following word. Our findings thus support the arguments of Orešnik (1984): the phonological preference for rightward clitic attachment holds generally, not just under special conditions that require phonological repair like sentences comprised entirely of clitics.[9] This, in turn, raises a theoretical issue: as discussed in §2, phasal spell-out predicts that Slovenian clitics—assumed to usually be in C, a phase head—should be spelled out together with the preceding phrase in the specifier of CP, meaning that clitics should attach by default to the left. We proposed two explanations for the discrepancy between this predicted syntactic grouping and our observed prosodic grouping: either phonology receives leftward-leaving clitics from syntax and consistently overrides this to reattach them to the right prosodically, or something in our syntactic assumptions is incorrect. The present study cannot distinguish between these two outcomes, and we leave further investigation for future work.

Task hypothesis 2 (large boundary < smaller boundary) predicted that the start of the cluster should be a stronger perceptual attractor of beeps in condition |C (which has a clause boundary there) than in the test conditions (which have some sort of lower-level phrase boundary). The graph by condition in Figure 3 shows a substantial difference between condition |C and the test conditions: the former has a substantial peak of responses at the start of the cluster, while the latter have gentler peaks split between the start of the cluster and the preceding syllable. When we look at the results for each target beep in Figure 4, we see that this difference seems to be located in a more extreme tendency for beeps close to

[9] A reviewer points out that all of our stimuli have the verb following the clitics, and that this may affect the direction of cliticization. Indeed, in a production study, Połomská & Ziková (2025) find that second-position clitics in Czech are more prosodically integrated with preceding sentence-initial verbs than nominal phrases. They attribute this difference to a syntactic distinction: sentence-initial verbs are in the same complex head as the clitics, while nominal phrases are—as phrases—in a preceding specifier (Matushansky 2006). Thus, we can say that our results hold when the pre-clitic position is filled by a phrase, but further work is required to test verb-initial sentences as well.

the start of the cluster (especially beeps 2–4) to be perceived at the start of the cluster itself. However, the model presented in §4.2 only found a significant difference between condition DCD and condition |C; conditions MCD and DCM were not significantly different from condition |C. Given the apparent observable difference between condition |C and the test conditions, it may be that our experiment was not powerful enough for this difference to achieve statistical significance, and further study would more convincingly show the beep test's ability to distinguish between different levels of prosodic boundaries.

In §3.2.1, we discussed two explanations for increased perception of beeps by prosodic boundaries. One (absolute attraction) is that hearers will perceptually align beeps at prosodic boundaries because they are less intrusive there. This predicts a large spike of responses at prosodic boundaries and no spike (or even a slight drop) in adjacent syllables. The other explanation (relative attraction) is that prosodic boundaries and their surrounding syllables have increased (actual and perceptual) salience—pauses, phrase-final lengthening, phrase-initial strengthening, and similar. If boundaries increase the (actual or perceptual) prominence of the surrounding area, we would expect increased perception of beeps both at and near the boundary, since larger or more salient syllables and boundaries have more space for beeps to fall into than others. In this case, we would expect a distribution with a gentler peak centered around the boundary.

The latter pattern is what we see. This is clearest in condition |C, which has a strong peak at the start of the cluster overall, but is also apparent from Figure 4, which shows results broken down by beep location: when the beep is sufficiently close to a prosodic boundary, we get the most beeps at the boundary and the second-most beeps on the syllable preceding or following it. When the beep falls closer to the edge of the cluster without a boundary, we get a bimodal pattern: some beeps still "jump" towards the boundary, while others settle for the nearer edge of the cluster without a prosodic word boundary. Finally, when the beep is perceived as sufficiently far from the boundary (that is, beep 1 in the test conditions), we get more responses at the syllable before the boundary than on the boundary itself. All of these patterns suggest that the syllables adjacent to (especially, preceding) a boundary enjoy some degree of perceptual prominence, though less than the boundary itself—in line with the predictions of relative attraction.

Next, we will discuss our results in light of the earlier click location studies. We have assumed, as argued by Fodor & Bever (1965) and others, that the perceived location of beeps is influenced by linguistic structure. However, Reber & Anderson (1970) and Reber (1973) argue that this perception is instead (largely if not entirely) influenced by three non-linguistic factors: response bias, attention, and memory. Can these factors provide an alternate explanation of our results? Let us go through each in turn. The response bias in question is that participants are more likely to say that they heard beeps at major boundaries overall. For example, participants are more likely to say they heard a click at major boundaries, *even when no click is actually audible* (Reber & Anderson 1970). In our study, we are using beep perception to figure out where the prosodic boundaries are, not how humans process sentences—as long as boundaries attract the perception of beeps, it does not matter whether they do so because of sentence processing or response bias. For our purposes, then, structurally sensitive response bias is a relevant linguistic factor as much as structurally sensitive processing.

The law of prior entry, proposed by Titchener (1908), states that objects that are the focus of attention are perceived earlier than those to which less attention is being paid (see also Spence & Parise 2010). Reber & Anderson (1970) and Reber (1973) find that beeps tend to be perceived *later than* their true location when they appear early in the sentence and *earlier than* their true location when they appear later in the sentence. In their interpretation, this is due to a shift in attentional priorities: early on, the sentence is new, so attention is focused there. As the sentence continues and becomes more semantically predictable, and as the wait for the click becomes longer and longer, attention instead shifts towards the click. The present study finds that beeps were consistently perceived earlier than their actual location, though the amount differed somewhat—a finding similar to Ladefoged & Broadbent (1960) and Bertelson & Tisseyre (1970), though not other studies. The law of prior entry, then, would indicate that speakers are consistently paying more attention to the beep than the sentence—a reasonable assumption given that speakers did not have to memorize the entire sentence they heard, they heard each sentence multiple times, and many of the sentences were similar in structure. On the other hand, this perceptual difference may be a task effect, in that most stimuli had a strong boundary pulling perception leftward. Another possible explanation is that speakers tend to synchronize regular pulses to syllables not at their onset but at the onset of their vowel (for an overview, see Rathcke forthcoming)—thus, beeps located at the onset of a syllable would be interpreted as preceding the syllable because the beep precedes the syllable's vowel.[10] At any rate, attention cannot explain the differences between our conditions—if anything, this factor would separate condition |C from the rest, since this condition has more content before the clitic cluster than the others.

Reber & Anderson (1970) and Reber (1973) also argue for an effect of memory: beeps later in the sentence are perceived more accurately than those earlier in the sentence, where there is more time to forget where the beep was heard. However, we are not concerned with accuracy of beep placement as such—in particular, we did not intentionally locate beeps at either edge of the clitic cluster—so much as direction of error. Accordingly, memory should not have a major effect on our results.

At first glance, our finding that surface prosodic structure determines beep perception goes against the findings of Bever et al. (1969), who argue that clicks are attracted to boundaries between constituents in deep structure (according to the transformational theory of syntax prevalent at the time), even when these do not align with surface boundaries. However, Chapin et al. (1972) criticize the syntactic analyses on which Bever et al. (1969) base their interpretation and instead provide evidence that clicks are attracted to major boundaries—which, for them, are the boundaries separating a sentence's subject from its predicate in standard English SVO sentences. This, presumably, is analogous structurally to the boundary we target in the test conditions, which separates the pre-clitic constituent from post-clitic material (with clitics, as we have shown, attaching to the former). Thus, our results are in line with those of Chapin et al. (1972) and others, like Geers (1978), arguing that surface structure plays a crucial role in click/beep perception.

Finally, we have argued that our results show sensitivity to prosodic boundaries. However, we can ask where participants' prosodic boundaries come from. One possibility is that they represent the natural prosody generated by their grammar for a given sentence. In this case, the results reflect participants' default patterns of cliticization, as we have been assuming. However, as discussed in §3.1.3, it is possible that our naturalistic stimuli have properties (such as intonation or syllable length) that cue listeners towards a particular prosodic parse, especially in the divergent condition C|. In this case, our results would instead reflect the *speaker's* patterns of cliticization and the fact that these are normal enough to be accommodated by listeners. In this case, our results are not qualitatively different, just quantitatively narrower: reflecting the grammar of at least one Slovenian speaker (who recorded the stimuli) rather than the grammar of approximately fifty (the participants). Future work could test this more thoroughly by presenting stimuli modified to remove possible prosodic cues, i.e. with flattened pitch and equally timed syllables. However, this would come at the cost of presenting unnatural stimuli.

6 Conclusion

In this study, we updated a classical experimental perception paradigm (Fodor & Bever 1965, Garrett et al. 1966, Bever et al. 1969, Reber & Anderson 1970, Chapin et al. 1972) to detect prosodic boundaries: beeps were inserted into stimulus sentences, and participants were asked to mark the location where they perceived the beep; hearers should be more likely to perceive beeps at and near prosodic boundaries. We used this paradigm to address an open question about the prosody of Slovenian sentences: although it was previously known that Slovenian "second position" clitics are able to attach either to the left or to the right when necessary, there are disputing claims about which direction they prefer to attach when both are possible. Unlike many more contemporary perception studies studying prosodic boundaries (e.g. Scott 1982, Gollrad 2013, Petrone et al. 2017), we could not rely on differences in meaning implied by prosodic parses, since the direction of cliticization has no difference in meaning and is difficult if not impossible to detect from acoustic input alone. Our beep-based experimental paradigm does not rely on assumptions about mapping between prosody and meaning, nor on detecting differences implicating

[10]We thank an anonymous reviewer for this suggestion.

prosodic parse in the input. We found strong evidence that Slovenian clitic clusters attach by default to the right, as argued by Orešnik (1984).

We hope that this study serves as impetus for future research in two directions. First, theoretical work is needed to properly account for this cliticization pattern, which at least at first sight seems to go against the predictions of phasal spell-out. Second, we hope that this study returns the paradigm of detecting foreign sounds in running speech to linguists' experimental toolbox, and that future work both uses the paradigm for other questions of prosodic constituency and revisits the previous work in this vein to figure out what, exactly, is causing beeps to be perceived more often at and around prosodic boundaries—whether it be processing constraints, response bias, or some other factor.

Data

All appendices, data, analysis code, experimental materials, calculations, and alternate models referenced in this study can be found in the supplemental materials at: https://osf.io/dgs2q/?view_only=1a9f605b3e3a4dcb931342cf413837ac.

References

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278.

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2018. Parsimonious Mixed Models. https://arxiv.org/abs/1506.04967.

Becker, Michael & Kathryn Flack Potts. 2011. The Emergence of the Unmarked. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology*, vol. III, chap. 58. Wiley. doi:10.1002/9781444335262.wbctp0058.

Bertelson, Paul & Françoise Tisseyre. 1970. Perceiving the Sequence of Speech and Non-Speech Stimuli. *Quarterly Journal of Experimental Psychology* 22(4). 653–662. doi:10.1080/14640747008401943.

Bever, T. G., J. R. Lackner & R. Kirk. 1969. The underlying structures of sentences are the primary units of immediate speech processing. *Perception & Psychophysics* 5(4). 225–234. doi:10.3758/bf03210545.

Boersma, Paul & David Weenink. 2023. Praat: doing phonetics by computer. Computer program. Version 6.3.17.

Bošković, Željko. 2001. *On the Nature of the Syntax–Phonology Interface: Cliticization and Related Phenomena* (North Holland Linguistic Series: Linguistic Variations 60). Amsterdam: Elsevier. doi:10.1163/9780585474250.

Browne, Wayles. 1974. On the problem of enclitic placement in Serbo-Croatian. In Richard Brecht & Catherine Chvany (eds.), *Slavic Transformational Syntax*, 36–52. Ann Arbor, MI: University of Michigan.

Browne, Wayles. 1975. Serbo-Croatian enclitics for English-speaking learners. In Rudolf Filipović (ed.), *Contrastive Analysis of English and Serbo-Croatian*, 105–134. Zagreb: Institute of Linguistics, University of Zagreb. Reprinted in Journal of Slavic Linguistics 12(1–2) (2004), pp. 249–283.

Cambier-Langeveld, Tina. 1997. The Domain of Final Lengthening in the Production of Dutch. *Linguistics in the Netherlands* 14(1). 13–24. doi:10.1075/avt.14.04cam.

Cambier-Langeveld, Tina, Marina Nespor & Vincent J. van Heuven. 1997. The domain of final lengthening in production and perception in Dutch. In *Proceedings 5th European Conference on Speech Communication and Technology (Eurospeech 1997)*, 931–934. ISCA. doi:10.21437/eurospeech.1997-329.

Chapin, Paul G., Timothy S. Smith & Adele A. Abrahamson. 1972. Two factors in perceptual segmentation of speech. *Journal of Verbal Learning and Verbal Behavior* 11(2). 164–173. doi:10.1016/s0022-5371(72)80073-2.

Cho, Taehong & Patricia A. Keating. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29(2). 155–190. doi:10.1006/jpho.2001.0131.

Cho, Taehong, James M. McQueen & Ethan A. Cox. 2007. Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics* 35(2). 210–243. doi:10.1016/j.wocn.2006.03.003.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A life in language*, 1–52. Cambridge, MA: MIT Press. doi:10.7551/mitpress/4056.003.0004.

Dvořák, Boštjan. 2007. Slovenian clitic pronouns and what is so special about them. *Slovenski jezik / Slovene Linguistic Studies* 6. 209–233. doi:10.17161/SLS.1808.4408.

Fodor, J. A. & T. G. Bever. 1965. The Psychological Reality of Linguistic Segments. *Journal of Verbal Learning and Verbal Behavior* 4(5). 414–420. doi:10.1016/s0022-5371(65)80081-0.

Fodor, Janet Dean. 2002. Prosodic disambiguation in silent reading. In *Proceedings of Nels 32, vol. 1*, 113–132. Amherst, MA: GLSA.

Fougeron, Cécile & Patricia A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101(6). 3728–3740. doi:10.1121/1.418332.

Franks, Steven. 2016. Clitics are/become minimal(ist). In Franc Marušič & Rok Žaucer (eds.), *Formal Studies in Slovenian Syntax: In honor of Janez Orešnik* (Linguistik Aktuell/Linguistics Today 236), 91–128. Amsterdam: John Benjamins. doi:10.1075/la.236.05fra.

Franks, Steven & Tracy Holloway King. 2000. *A Handbook of Slavic Clitics* Oxford Studies in Comparative Syntax. New York, Oxford: Oxford University Press.

Garrett, Merrill, Thomas Bever & Jerry Fodor. 1966. The active use of grammar in speech perception. *Perception & Psychophysics* 1. 30–32. doi:10.3758/bf03207817.

Geers, Ann E. 1978. Intonation Contour and Syntactic Structure as Predictors of Apparent Segmentation. *Journal of Experimental Psychology: Human Perception and Performance* 4(2). 273–283. doi:10.1037/0096-1523.4.2.273.

Golden, Marija & Milena Milojević Sheppard. 2000. Slovene pronominal clitics. In Frits H. Beukema & Marcel den Dikken (eds.), *Clitic phenomena in European languages*, 191–208. Amsterdam: John Benjamins. doi:10.1075/la.30.08gol.

Gollrad, Anja. 2013. *Prosodic cue weighting in sentence comprehension: processing German case ambiguous structures*: Universität Potsdam dissertation.

Gussenhoven, C. & A. C. M. Rietveld. 1992. Intonation contours, prosodic structure and preboundary lengthening. *Journal of Phonetics* 20(3). 283–303. doi:10.1016/s0095-4470(19)30636-9.

Kayne, Richard S. 1994. *The Antisymmetry of Syntax* (Linguistic Inquiry Monographs 25). Cambridge, MA: MIT Press.

Krivokapić, Jelena. 2007. *The planning, production, and perception of prosodic structure*: University of Southern California dissertation.

Krivokapić, Jelena & Dani Byrd. 2012. Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics* 40(3). 430–442. doi:10.1016/j.wocn.2012.02.011.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13). 1–26. doi:10.18637/jss.v082.i13.

Ladefoged, Peter & Donald E. Broadbent. 1960. Perception of Sequence in Auditory Events. *Quarterly Journal of Experimental Psychology* 12(3). 162–170. doi:10.1080/17470216008416720.

Lenth, Russell V. 2023. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://CRAN.R-project.org/package=emmeans. R package version 1.9.0.

Lüdecke, Daniel, Mattan S. Ben-Shachar, Indrajeet Patil, Philip Waggoner & Dominique Makowski. 2021. performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software* 6(60). 3139. doi:10.21105/joss.03139.

Marušič, Franc. 2008. Slovenian Clitics Have No Unique Syntactic Position. In Andrei Antonenko, John F. Bailyn & Christina Bethin (eds.), *Formal Approaches to Slavic Linguistics 16: The Stony Brook Meeting 2007*, 266–281. Ann Arbor, MI: Michigan Slavic Publications.

Matushansky, Ora. 2006. Head Movement in Linguistic Theory. *Linguistic Inquiry* 37(1). 69–109.

Orešnik, Janez. 1984. Slovenske breznaglasnice se vedejo predvsem kot proklitike [Slovenian clitics act primarily as proclitics]. *Jezik in slovstvo* 29(4). 129.

Petrone, Caterina, Hubert Truckenbrodt, Caroline Wellmann, Julia Holzgrefe-Lang, Isabell Wartenburger & Barbara Höhle. 2017. Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics* 61. 71–92. doi:10.1016/j.wocn.2017.01.002.

Poĺomská, Anna & Markéta Ziková. 2025. Phonology determined by syntax: enclisis in Czech. Talk presented at the 22nd Old-World Conference in Phonology, University of Amsterdam.

Progovac, Ljiljana. 1996. Clitics in Serbian/Croatian: Comp as the second position. In Aaron Halpern & Arnold M. Zwicky (eds.), *Approaching Second: Second Position Clitics and Related Phenomena*, 411–428. Stanford: CSLI Publications.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna. https://www.R-project.org/.

Radanović-Kocić, Vesna. 1988. *The Grammar of Serbo-Croatian Clitics: A Synchronic and Diachronic Perspective*. Urbana-Champaign, IL: University of Illinois dissertation.

Rathcke, Tamara. forthcoming. The P-Center Effect and the Domain of Beat Perception in Speech. In Lars Meyer & Antje Strauß (eds.), *Rhythms of Speech and Language: Physiology, Cognition, Culture*, Cambridge: Cambridge University Press.

Reber, Arthur S. 1973. Locating clicks in sentences: Left, center, and right. *Perception & Psychophysics* 13(1). 133–138. doi:10.3758/bf03207248.

Reber, Arthur S. & John R. Anderson. 1970. The perception of clicks in linguistic and nonlinguistic messages. *Perception & Psychophysics* 8(2). 81–89. doi:10.3758/bf03210179.

Schütze, Carson T. 1994. Serbo-Croatian Second Position Clitic Placement and the Phonology-Syntax Interface. In Andrew Carnie & Heidi Harley with Tony Bures (eds.), *Papers on phonology and morphology* (MIT Working Papers in Linguistics 21), 373–473. Cambridge, MA: Department of Linguistics and Philosophy, MIT.

Scott, Donia R. 1982. Duration as a cue to the perception of a phrase boundary. *The Journal of the Acoustical Society of America* 71(4). 996–1007. doi:10.1121/1.387581.

Simon, Anne-Catherine & George Christodoulides. 2016. Perception of Prosodic Boundaries by Naïve Listeners in French. In *Proceedings Speech Prosody 2016*, 1158–1162. doi:10.21437/SpeechProsody.2016-238.

Škrabec, Stanislav. 1895. Nekoliko slovenske slovnice za poskušnjo [A taste of Slovenian grammar]. *Cvetje z vertov sv. Frančiška* 14(4). n.p.

Spence, Charles & Cesare Parise. 2010. Prior-entry: A review. *Consciousness and Cognition* 19(1). 364–379. doi:10.1016/j.concog.2009.12.001.

Tabain, Marija. 2003. Effects of prosodic boundary on /aC/ sequences: Articulatory results. *Journal of the Acoustical Society of America* 113(5). 2834–2849. doi:10.1121/1.1564013.

Titchener, Edward Bradford. 1908. *Lectures on the Elementary Psychology of Feeling and Attention*. Macmillan.

Toporišič, Jože. 1992. *Enciklopedija slovenskega jezika [Encyclopedia of the Slovenian language]*. Ljubljana: Cankarjeva založba.

Toporišič, Jože. 2000. *Slovenska slovnica [Slovenian grammar]*. Maribor: Obzorja 4th edn.

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti J. Price. 1992. Segmental Durations In The Vicinity Of Prosodic Phrase Boundaries. *Journal of the Acoustical Society of America* 91(3). 1707–1717. doi:10.1121/1.402450.

Zehr, Jérémy & Florian Schwarz. 2022. PennController for Internet Based Experiments (IBEX). doi:10.17605/OSF.IO/MD832.