Slovenian clitics prefer to cliticize to the right

April 9, 2024

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 arguments 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 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 lean left and 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 jump.'/'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 big 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 and a discussion, and §5 concludes.

2 Proclitics vs. Enclitics

Slovenian clausal clitics are Wackernagel clitics forming a cluster which is (typically) located after the first syntactic constituent (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: p. 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: p. 192).

The 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) or, if prosodic requirements are satisfied, in any lower head (Bošković 2001, Marušič 2008). The theory of phasal spell-out (Chomsky 2001) predicts that this head (whether C or otherwise) should be spelled out with its specifier preceding it. 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 clitic head and its preceding specifier should get spelled out together 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 Algorithm of Kayne 1994). Thus, clitics should normally encliticize to the preceding constituent. It is hard to imagine a different way for clitics to be attached.

It seems, therefore, that phase theory predicts that Slovenian clitics should only be enclitics. Surprisingly, 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) puts forth further arguments suggesting that Slovenian clausal clitics are usually proclitics.

One of his arguments involves stressed 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). 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 (watch it).' (Orešnik 1984)

This sentence shows a genuine stressed clitic, with seemingly the same syntax as other clitic clusters.[2] We see no evidence of overt movement, nor is it possible that stressed *ga* is actually a full (non-clitic) pronoun. The full form of the third-person accusative masculine pronoun is *njéga* 'him', with stress on the first syllable; and *njéga* actually cannot replace the stressed *ga* in (4). 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 (watch it).'

According to Orešnik (1984), sentences like (4) and (5) provide arguments for a general preference for Slovenian clitics to attach to the right; if leftward attachment was the default, we would expect the *first* clitic in such cases to be stressed rather than the *last* one. Orešnik (1984) does not offer an account of these facts; he is only commenting on a descriptive generalization. But the fact that one of the clitics in (4)–(5) ends up carrying stress suggests a phonological repair strategy at work that assigns stress to one of the clitics because of a need for every prosodic word to have stress. Moreover, since there is a general preference for clitics to be proclitics, it is the last of the three clitics that is assigned stress, regardless of

[2]In principle it would be possible that the stressed clitic in (4) and (5) is syntactically in a different position than the unstressed clitics of (4)–(5). This possibility is given some plausibility by the fact that functionally similar-looking stressed clitics can also occur in examples such as (i), where they are clearly not part of the clitic cluster, being split off from the unstressed clitics by an adverb. The data also argue for the existence of a phonological mechanism that affects the pronounciation of clitics in Slovenian.

(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, there are reasons to believe that (i) and (4)–(5) are not comparable after all. Firstly, (i) has a verum focus interpretation, which does not seem to be the case with (4) and (5). And further, the fact that either a pronominal or the auxiliary clitic can receive verum focus makes it a little harder to imagine a common syntactic explanation for examples in (i). We leave (i) and similar examples for future work.

its syntactic status (see Franks 2016 for a proposal along these lines). We will not try to explain how this phonological repair strategy might work, but we should keep in mind that it does seem to exist.

However, (4) and (5) do not necessarily contradict the phase-based account, whose predictions for such sentences are unclear. If all three formally (but not prosodically) speaking clitic elements reside in the same syntactic head, they are all spelled out at the same time and are part of the same prosodic constituent. When the specifier is not prosodically empty, clitics are predicted to attach to it, but since there is no specifier onto which the clitics could prosodically lean in (4), the phase-based account alone has nothing to say about it. It is possible, for example, that a weak preference for clitics to attach to the right emerges *only* in cases where a repair strategy is needed; in the usual case, clitics are happy to attach leftward to a specifier.[3]

The evidence presented by Orešnik (1984), showing a general phonological preference for rightward clitic attachment in Slovenian, must be reconciled with the syntactic account. There are three options. The first, described above, is that clitics attach to the right as part of the phonological repair strategy but default to the left otherwise, as predicted by phase-based spellout. The second possibility is that the preference for clitics to attach to the right is strong enough to force a phonological repair altering the output of phase-based spellout *by default*, yielding a language-wide preference for rightward clitic attachment (and stretching the concept of phonological repair, which should presumably be limited to a small number of aberrant cases). The third 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 without the need for phonological repair. Regardless of which of these three options turns out to be correct, however, ascertaining the direction of cliticization in Slovenian is thus 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: p. 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.[4]

(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).

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

[4]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 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.

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. When clitics attach to the left (as enclitics), there is a prosodic word boundary after the cluster, between the cluster and the following word, but none before the cluster. On the other hand, when clitics attach to the right (as proclitics), the prosodic word boundary is located at the left edge of the cluster. While it is possible that the prosodic word boundary at either edge of the clitic cluster may also be the boundary of a larger (intermediate) prosodic domain, we cannot say this for sure ahead of time. Prosodic words do show some edge effects—for example, consonants tend to be articulated more strongly at the edges of prosodic domains, including prosodic words (see e.g. Fougeron & Keating 1997, Cho & Keating 2001, Cho et al. 2007)—but these effects can be quite small at low-level prosodic boundaries, especially given interaction with word stress, and may require articulatory data to detect. If such effects do exist, they do not rise to the level of Slovenian speakers' conscious perception.

To address these difficulties in detecting the presence of lower-level prosodic boundaries, in particular at the left or right edge of the clitic cluster, we developed a novel perception experiment described in §3.1.

Previous perception experiments use methodologies that are not applicable to our research question. For example, some 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.

Other studies (e.g. Scott 1982, Gollrad 2013, Petrone et al. 2017) rely on identifying differences in meaning implied by different prosodic parses. These experimental tasks are likewise unable to distinguish clitics' direction of attachment.

Some researchers (e.g. Cambier-Langeveld 1997) have also looked for prosodic boundaries—including low-level prosodic word boundaries—in speakers' production by measuring criteria like lengthening of pre-boundary syllables (cf. Wightman et al. 1992, Fougeron & Keating 1997, Tabain 2003). In the case of Slovenian clitics, it would be difficult to properly isolate the desired effect in a production study, which would have to compare differences in production among clitics—which are usually function words and heavily unstressed—and constituents in pre-clitic position, which are typically comprised of lexical words that can more easily bear stress.

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 developed a methodology intended to detect perceived prosodic boundaries: participants heard sentences with beeps inserted and marked the location where they heard the beep. We hypothesize that speakers are more likely to perceive beeps at prosodic boundaries. There are at least two reasons why this might be the case. First, beeps in the middle of a word or a prosodic constituent are intrusive, so hearers might naturally "snap" them to prosodic breaks instead, where they cause less of an interruption. Second, prosodic boundaries often include strengthening effects like pauses and phrase-final lengthening. If these boundaries are accompanied by the perception of a pause or a longer phrase-final syllable (whether or not these are actually present in the input), there will be more "perceptual time" for the beep to fall into, and beeps will be more likely to be perceived at and around prosodic boundaries. As is generally true for prosodic effects, beep perception effects should be stronger at larger prosodic boundaries: the edge of an intonational phrase, for example, should attract beeps to a greater extent than the edge of a prosodic word.

3.1.1 Hypotheses

The experiment tests three hypotheses, which are listed in (7). The first two hypotheses, in (7-a) and (7-b), are intended as baseline hypotheses testing the interpretation of the experimental paradigm: if these hypotheses are confirmed, we have reason to believe that beeps are more likely to be perceived at larger prosodic boundaries. This serves as the baseline for the third hypothesis, in (7-c), which is our main research question: if speakers are generally more likely to perceive beeps at the left edge of the clitic cluster than at the right edge, this would serve as evidence that Slovenian speakers generally place prosodic boundaries at the left edge of the cluster—that is, that they prefer to attach clitics to the right when possible.

(7) *Experimental hypotheses*
   a. Known prosodic word boundaries—for example, between adjectives and nouns they modify—should be perceptual attractors of beeps to a greater extent than syllable boundaries such as those within a two-syllable word.
   b. Larger prosodic boundaries—for example, at the beginning or end of a subordinate clause—should be perceptual attractors of beeps to a greater extent than low-level (prosodic word) and intermediate prosodic boundaries.
   c. The left edge of a clitic cluster will be a perceptual attractor of beeps to a greater extent than the right edge unless the cluster is directly followed by a larger prosodic boundary, in which case the right edge of the cluster will be a much greater perceptual attractor of beeps.

3.1.2 Perceptual attraction

The two explanations for perceptual attraction of beeps described in the introduction to this section predict different patterns for the distribution of beeps around a prosodic boundary. If hearers are snapping beeps to boundaries, we would expect a steady distribution of beeps around the boundary with a large spike at the boundary itself. If anything, there may be a dip in locations near the boundary, as hearers should be more likely to snap a beep to a boundary when it is very close. This distributional pattern, which we refer to as *absolute attraction*, is shown in Figure 1.

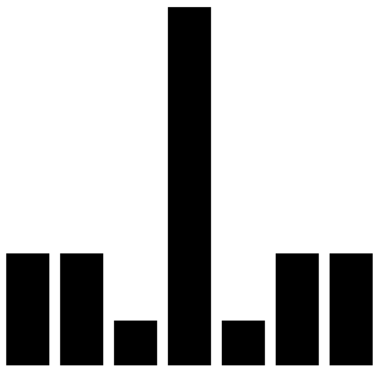

Figure 1: Expected distribution of perceived beeps around a prosodic boundary for absolute attraction

On the other hand, under the assumption that hearers are more likely to perceive beeps at boundaries because boundaries and the syllables near them are perceptually stronger, 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. This is because these syllables, too, are perceptually strengthened, so there is more perceptual space for the beeps to fall into. We call this distributional pattern *relative attraction*, and it is shown in Figure 2.

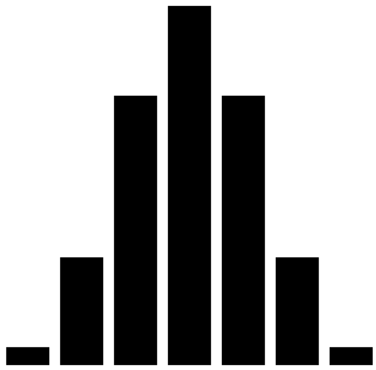

Figure 2: Expected distribution of perceived beeps around a prosodic boundary for relative attraction

There are at least two ways to quantify perceptual attraction of beeps, which show different sensitivities. First, we can measure the likelihood of a response (that is, the perceived location of a beep) at a particular boundary. This most clearly captures absolute attraction effects, in which responses at a boundary are much more likely than responses elsewhere. This measure should also capture relative attraction effects: responses are still more likely at a boundary than elsewhere, although the difference is not as great. However, this measure would have some trouble with a relative attraction effect centered for some reason *near* a boundary rather than precisely *at* one.

Alternately, we can measure the *distances* of responses from a particular boundary. This measurement is most sensitive to relative attraction effects, in which responses get more frequent as we get closer to a boundary. It should also be robust to slight deviations in the center of the distribution: if, for example, 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. Likewise, measuring distance also captures absolute attraction effects, in which a large number of responses are clustered a distance of 0 from the boundary.

In this paper, we make the conservative choice to present statistical models predicting the distance of responses from a particular point, because this measure shows greater sensitivity to a number of possible distributions centered at or near a given location. Models instead predicting whether a response is at a particular point are included in the supplementary materials; their results are broadly similar, with one substantive difference discussed in Appendix C.

3.1.3 Experimental task

In each trial, participants heard an audio file of a speaker reading one of the 30 sentences in Appendix A. The sentence had a beep inserted into it at some point. 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. An example is shown in Figure 3 for the sentence *Bankir mi bo hišo zastavil za kredit.* 'The banker will mortgage my house for a loan.' Spaces between words and punctuation (other than the period at the end of the sentence) are removed; syllable boundaries within words were placed according to the principles in Toporišič (1992: p. 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.[5] This sentence exemplifies one of the five experimental conditions, described below: the main clause clitics *mi bo* are preceded and followed by a disyllabic noun.

[5]Toporišič (1992: p. 377) gives the following examples of this principle of syllabification: *kav-ra* (here, *v* is pronounced [w]) 'rutabaga' , *prav-da* 'justice', *moj-ra* 'Moyra', *o-tka* 'stick', *sla-dka* 'sweet'.

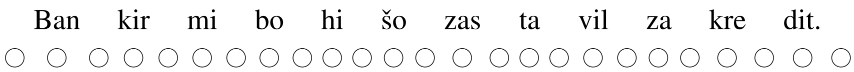

Figure 3: Sample experimental prompt, in which participants had to select where they heard the beep in the sentence *Bankir mi bo hišo zastavil za kredit.* ‘The banker will mortgage my house for a loan.’, with clitic cluster *mi bo* ‘me.DAT FUT.3SG’

3.1.4 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 conditions 1–3, the direction of cliticization is unknown, so the clitics are surrounded with ?, indicating that one boundary should contain a prosodic word boundary and the other, no such boundary, but it is not known which is which. In conditions 4–5, 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 4 or before the clitics in condition 5.

Condition	Sentence			
1	noun	? clitics ?	noun	| rest
2	modifier | noun	? clitics ?	noun	| rest
3	noun	? clitics ?	modifier | noun	| rest
4a	noun || relative clause ||	clitics	noun	| rest
4b	noun || relative clause ||	clitics	modifier | noun	| rest
5a	noun	clitics ||	adjunct clause	|| rest
5b	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 4 and 5 were divided into two subconditions with three sentences each (all experimental sentences are given in Appendix A). Each clitic cluster contained two clitics, a one-syllable pronominal clitic (e.g. the reflexive accusative clitic *se*) 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 1, 3, and 5a, while nouns with initial stress like *sáblje* ‘sabres’ appear after the clitic cluster in conditions 1, 2, and 4a. The modifier–noun pairs in conditions 2, 3, 4b, and 5b comprised two monosyllabic words, like *bel prt* ‘white tablecloth’ or *dva dni* ‘two days’. This configuration ensures that the noun and modifier–noun conditions do not differ in their stress pattern: in conditions 1–3, 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. In all cases, this *rest* includes the main verb (a verbal participle that combines with the clitic auxiliary). In conditions 1–3, the sentence always ends with the main verb followed by an additional phrase (e.g. *podarili šele včeraj* ‘give just yesterday’). In condition 4, the main verb can follow the same verb–phrase pattern (e.g. *dal za darilo* ‘give as a gift’), but can also stand on its own (e.g. *ukradli* ‘steal’) or be preceded by an adverb (e.g. *mogoče odplačal z delom* 'maybe repay with work'). In condition 5, the main verb is always followed by a phrase (e.g. *napisal tri listke* 'write three notes') and is also sometimes preceded by a phrase as well (e.g. *gotovo strgal hlače* 'surely rip pants').

Each sentence appeared with beeps in nine different locations: six target locations and three fillers. In the target conditions, beeps were placed at the midpoint of the syllable preceding the clitic cluster, the midpoint of the syllable following the clitic cluster, and four evenly spaced points in between. The filler beep locations varied by condition. In conditions 1–3, filler beeps were placed at the boundaries after each of the three elements following the clitic cluster: the second noun, the verb, and all other material (that is, one filler beep appeared at the very end of the sentence). In condition 4, fillers were placed at the beginning of the relative clause, the end of the clitic cluster inside the relative clause, and before the last word of the relative clause.[6] Finally, in condition 5, fillers were placed after the adjunct clause, before the verb, and after the verb.

Each participant saw 90 trials, divided internally into three blocks of thirty (the blocks were not separated with a break). Each block contained each sentence once; in each block, the six sentences in each condition appeared with four different target beeps and two different filler beeps. Participants saw each sentence with two target beeps and one filler beep (i.e. target 1, target 4, filler 1; target 2, target 5, filler 2; or target 3, target 6, filler 3). To maintain balance within a condition, the six sentences within each condition were divided into three groups of two each, so that each participant heard two sentences in each condition with beeps 1–4–filler 1, 2–5–filler 2, and 3–6–filler 3 (that is, each participant heard each of the nine beep locations twice for each of the five conditions). Sentences were assigned to groups randomly for each participant, so the number of times each sentence appeared with a given beep location could vary (minimum 9, maximum 23, mean 16.33, standard deviation 3.15). Thus, the only imbalance in the study design was at the sentence level; at the condition level, beep locations were balanced.

3.1.5 Predictions

In this section, we frame the experimental hypotheses in (7) in terms of comparisons between conditions. The comparison is framed in terms of perceptual attraction strength, which can be measured in two ways as described in §3.1.2.

(8) *Experimental predictions*

a. The boundary between the two syllables preceding the clitic cluster in condition 2 (a prosodic word boundary between a monosyllabic adjective and noun) should be a greater perceptual attractor than the two syllables preceding the clitic cluster in conditions 1 and 3 (between the syllables of a disyllabic noun).
The boundary between the two syllables following the clitic cluster in condition 3 (a prosodic word boundary between a monosyllabic adjective and noun) should be a greater perceptual attractor than the two syllables following the clitic cluster in conditions 1 and 2 (between the syllables of a disyllabic noun).

b. The left edge of the clitic cluster in condition 4 (a larger phrase boundary after a subordinate clause) should be a greater perceptual attractor than the left edge of the clitic cluster in conditions 1–3 (at most, an intermediate prosodic boundary) and condition 5 (no prosodic word boundary).
The right edge of the clitic cluster in condition 5 (a larger phrase boundary before a subordinate clause) should be a greater perceptual attractor than the right edge of the clitic cluster in conditions 1–3 (at most, an intermediate prosodic boundary) and condition 4 (no prosodic word boundary).

c. The left edge of the clitic cluster in conditions 1–3 will be a greater perceptual attractor than the right edge of the clitic cluster. In other words, the left edge of the clitic cluster in conditions 1–3 will be a perceptual attractor similar to (though weaker than) the left edge of

[6] One sentence in condition 4 only had one word in the relative clause after the clitic cluster; for this sentence, the second filler was placed in the middle of the relative clause clitic cluster.

the clitic cluster in condition 4, but the right edge of the clitic cluster in conditions 1–3 will not be a perceptual attractor like the right edge of the clitic cluster in condition 5.

These predictions are shown schematically in Table 2.

syllable	*prediction*	*hypothesis*
left edge of syllable preceding clitic cluster	$2 > 1, 3$	(7-a)
left edge of clitic cluster	$4 > 1, 2, 3 > 5$	(7-b), (7-c)
right edge of clitic cluster	$5 > 1, 2, 3, 4$	(7-b), (7-c)
right edge of syllable following clitic cluster	$3 > 1, 2$	(7-a)

Table 2: Experimental predictions by syllable and condition ($X > Y$ means that a given syllable boundary in condition X is expected to be a greater perceptual attractor than that syllable in condition Y)

We must be circumspect about the prediction in (8-c): a general preference for beeps to be perceived at the left edge of the clitic cluster over the right edge of the clitic cluster is not sufficient evidence that speakers prefer to attach clitics to the right. As shown in §4.1, participants show a general leftward shift in their perception of the beeps—although the actual location of the beeps is symmetrical about the clitic cluster, responses overwhelmingly precede their actual location. Given that the location of beeps is centered around the clitic cluster, this leads to a greater proportion of beeps being perceived at the left edge of the cluster than the right edge. Accordingly, we must use conditions 4 and 5 as benchmarks to determine the direction of clitic attachment in conditions 1–3.

3.2 Analysis

3.2.1 Coding syllable and beep locations

Responses of perceived beep locations are coded as numbers: the middle of the clitic cluster (the boundary between the two clitics) is coded as 0, and each syllable boundary is coded relative to this point. Thus, a response of the gap between the first clitic and the syllable preceding it (i.e., the left edge of the clitic cluster) is -1, the left edge of the syllable before it is -2, the right edge of the clitic cluster is 1, and so on. Responses of syllables are coded as half steps: a response of the first clitic is $-.5$, the second clitic is .5, the syllable after that is 1.5, and so on, as shown in Figure 4.

	Ban		kir		mi		bo		hi		šo		zas		ta		vil		za		kre		dit.
-3	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5	3	...										

Figure 4: Coding of responses for *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'

The actual locations of the target beeps are also coded as numbers. Target beeps 1 and 6 are centered at the midpoints of the syllables preceding and following the clitic cluster, respectively; these are coded as -1.5 and 1.5. 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 3.

As the example in Table 3 shows, the beeps in this sentence are evenly spaced (a tenth of a second 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 $-1.129 - (-1.5) = .371$, while for beeps 2 and 3, the difference is $-0.486 - (-1.129) = .643$. This is because of the difference in syllable lengths: one-tenth of a second is 37% the length of *bor*, but 79% the length of *ti*, so the distance between codes -2 and -1 is much greater than the distance between -1 and 0. This distortion reflects the experimental task (see Figure 3), in which each syllable was given a single button despite being of different lengths.

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

beep	*time (s)*	*code*
1	0.439	−1.500
2	0.539	−1.129
3	0.639	−0.486
4	0.739	0.283
5	0.839	1.014
6	0.939	1.500

Table 3: 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.'

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 left edge 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.

3.2.2 Statistical models

We test the predictions in Table 2 using mixed models centered around the four landmarks described there: the left edge of the syllable preceding the clitic cluster (coded as -2, as shown in Figure 4), the left edge of the clitic cluster (coded as -1), the right edge of the clitic cluster (coded as 1), and the right edge of the syllable following the clitic cluster (coded as 2). The models for the left and right edge of the cluster include trials from all 5 conditions, while the models for the edges of the syllable before and after the cluster only include trials from the first 3 conditions. As discussed in §3.1.2, the dependent variable of each presented regression is the distance between the response for a given trial and the landmark. For example, a response in the middle of the first clitic ($-.5$) is a distance of .5 from the left edge of the cluster, 1.5 from the left edge of the syllable preceding the cluster and the right edge of the cluster, and 2.5 from the right edge of the syllable following the cluster. The supplemental materials contain an additional set of analogous logistic regressions whose dependent variable measures whether or not a response for a given trial is at a given landmark. All models 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.

Each of the four regressions has the same predictors. First, we include random intercepts for sentence and participant. Next, we include condition, which is dummy-coded with a reference level dependent on the hypothesis for each landmark according to Table 2: condition 4 is the reference level for the regressions whose landmark is the left edge of the cluster; condition 5, for regressions focused on the right edge of the cluster, etc. 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). The regressions also include linear and quadratic terms for the location of the beep in a given trial, coded as described in §3.2.1. The quadratic term is intended to capture the fact that responses are expected to be close to the actual beeps, so the likelihood of a response at or near a landmark should increase as the beep location gets closer to the landmark. In theory, this distance relationship could be captured by a linear term measuring the absolute value of the distance between the landmark and the beep location. However, this is problematic, because, as mentioned in §3.1.5, participants generally perceived beeps as occurring somewhat earlier than they actually did, so the putative point from which beep distance should be measured is not the landmark itself but some point after the landmark. Including a quadratic term allows for the effect of beep location to be symmetric around an axis without needing to specify the location of this axis (see Appendix B for more details).

We removed 29 outlier responses that were more than 3 syllables away from the clitic cluster (earlier than -4 or after 4). 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 have been chosen randomly due to lack of attention on a particular trial); 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 of -3 or greater has an outsized effect for condition 4, since the other conditions only have two syllables before the cluster. Removing the outliers allows for better comparison across conditions. In this light, the question arises of why we put the cutoff at -4 instead of -3. As shown in Figure 5 below, condition 4 does have a few responses between -3 and -4 before tailing off fully. Thus, -4 seems to be the left edge of the distribution for "legitimate" responses, and it would be perhaps inappropriate to cut off part of the distribution. Moreover, excluding the few tokens between -4 and -3 turns a non-significant result (in §4.1.2.1) into a marginally significant one.[7] To avoid presenting a result whose significance is so contingent, we instead present the model with a more conservative removal of outliers.

Unless otherwise specified, 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 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.

All models and calculations are available in the supplementary materials.

3.3 Materials

The stimuli comprised recordings of thirty sentences, listed in Appendix A, 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 naturally; in the authors' subjective perception, there are no audible pauses on either side of clitic clusters except, sometimes, when adjacent to a subordinate clause (conditions 4 and 5). The resulting recordings were processed in version 6.3.17 of Praat (Boersma & Weenink 2023). First, ambient noise was removed. Some sentences were louder than others (as the speaker had the microphone at a slightly different distance from her mouth), so all were normalized to have an average intensity of 70 dB. Next, the first co-author (not a native Slovenian speaker) segmented the sentences into syllables. Ambiguity at the crucial junctures (within and near the clitic cluster, coded as -2 to 2) was reduced by ensuring that they had at least one non-glide consonant—that is, there were no clitic clusters like *mi je* 'me.DAT AUX.3SG'. Finally, beeps were inserted (by a Praat script) into the sentences centered at the locations described in §3.1.4. 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.

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

3.4 Participants

A total of 50 participants were recruited through Prolific. Participants who indicated that their first language was Slovenian were allowed to take the study. 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.

An additional 20 participants took an earlier version of the study with the same stimulus sentences and a different experimental task. These participants are not included in the results presented here, and participants in the first version of the study were not allowed to participate in the final version.

Each participant saw two trials introducing them to the task, where they were told where the beep would be before they heard it. This was followed by six more example trials (using stimuli with beeps in different locations from those used in the main portion of the experiment) and 90 trials in the main

[7] This same difference also reaches significance ($p = .022$) in the model with a binary dependent variable measuring whether or not a response was at the landmark (see §3.1.2 for discussion). This model is also available in the supplementary materials.

experiment, of which 30 were fillers—see §3.1.4 for more details. This left a total of $49 \cdot 60 = 2940$ target trials analyzed in §4.

4 Results and discussion

4.1 Results

4.1.1 Descriptive summary

Figure 5 shows responses for target stimuli (beeps 1–6) by condition. (A version that breaks down responses by both responses and beeps is available in the supplementary materials.) 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 4 has a thick black line at the left edge of the clitic cluster representing the larger prosodic boundary at the end of the relative clause; condition 5 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 conditions 1–4, and in between the adjective and noun in condition 2 (preceding the cluster) and 3 (following the cluster). Subconditions 4a and 4b are collapsed, as are 5a and 5b, so no prosodic word boundary is shown between the adjective and noun in conditions 4b and 5b. Finally, the clitic clusters themselves are delineated by dashed lines, indicating that the location of the prosodic word boundary is (in most cases) unknown. In conditions 1–3 and 5, 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 5 includes the area within 3 syllables of the clitic cluster; 29 outliers appearing earlier (possible in condition 4 only) or later (possible in all conditions) are not pictured.

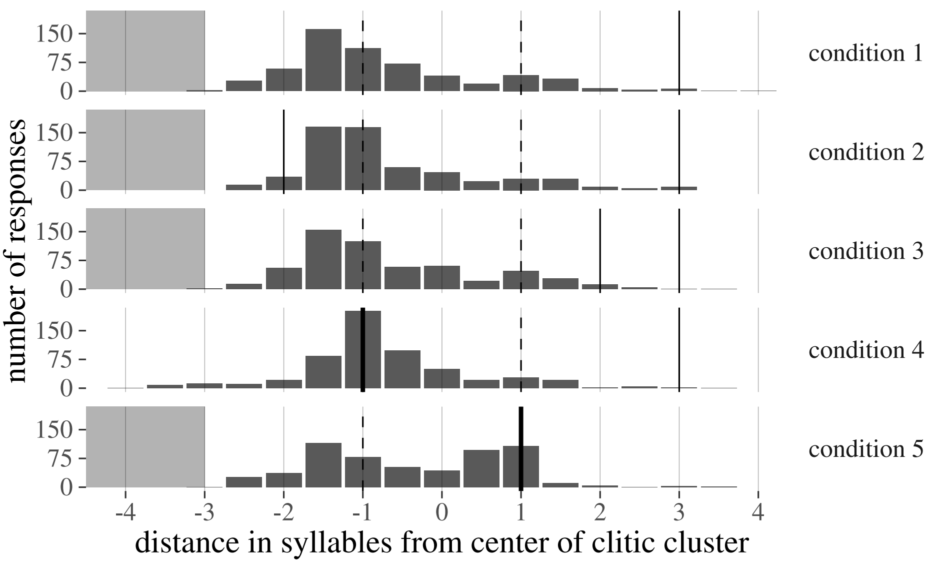

Figure 5: Experimental responses by condition, with expected prosodic boundaries according to Table 1 (with 29 outliers removed)

The pattern of responses in conditions 1–3 is quite similar: the greatest number of responses occurs at the left edge of the clitic cluster (position -1) and the preceding syllable (position -1.5), with a much smaller peak at the right edge of the clitic cluster (position 1). In condition 4, the peak is more heavily concentrated at the left edge of the cluster itself than on an adjacent syllable. Condition 5 shows quite a different pattern: there are two peaks of roughly equal size centered on the syllable preceding the left edge of the clitic cluster and on the right edge of the cluster (and the second clitic, which precedes it). These results follow the predictions in Table 2 for cluster edges: first, the larger prosodic boundaries (the left edge of the cluster in condition 4 and the right edge of the cluster in condition 5) have the greatest concentration of responses. Second, conditions 1–3 behave much more similarly to condition 4 than condition 5, showing many more responses on the left edge of the cluster than the right (suggesting that participants generally perceived the clitic cluster as attaching rightward). The large number of responses on the syllables adjacent to the cluster edge (especially the syllable preceding the left edge of the cluster) is unexpected.

Figure 5 also shows a slight difference between conditions 1–3 outside the cluster: there are fewer responses at position −2 for condition 2, which has a word boundary between adjective and noun there, than for the other two conditions, which have no word boundary there. This is the *opposite* of the pattern predicted in Table 2.

The asymmetry between left-favoring conditions 1–4 (which are heavily biased towards the left edge of the cluster) and right-favoring condition 5 (which has similar peaks on both sides of the cluster) is related to an overall leftward asymmetry in responses: as shown in Table 4 below, speakers tend to perceive beeps as occurring slightly earlier than they actually do. Given this, conditions 1–3 have a concentration of responses near the left edge of the clitic cluster that goes beyond what we would expect from the general leftward bias. The distribution of responses suggests that in conditions 1–3, as in condition 4, speakers attach clitics to the right, meaning that they place a prosodic boundary at the left edge of the cluster that serves as a perceptual attractor of beep responses.

In the baseline case, speakers tend to perceive beeps as occurring about one syllable earlier than they actually do. We can see this in Table 4, which shows the mean difference between a beep's actual and perceived location for each beep in conditions 1–3 (with no outliers removed). Also listed are the locations of each beep; the locations for beeps 2–5 are approximate, as these are interpolated from the locations of beep 1 and 6.

	beep	*location*	*mean leftward bias*
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	7	before main verb	0.91
	8	after main verb	1.04
	9	end of sentence	0.29

Table 4: Mean leftward bias (number of syllables by which a beep's perceived location precedes its actual location) by beep for conditions 1–3

The most neutral contexts are filler beeps 7 and 8, which are placed at prosodic word boundaries nowhere near the clitic cluster. Here the mean leftward bias is approximately 1, meaning that speakers perceived these beeps as occurring a full syllable earlier than they actually did. Target beeps 4–6 have similar biases, but the bias shrinks as we get closer to beep 1. This may be because the left edge of the clitic cluster is a greater perceptual attractor because it is a larger or particularly salient prosodic word boundary (made more salient by the experimental task), or because going further left places the beep too close to the beginning of the sentence. Filler beep 9 also has a smaller leftward bias because its location, at the very end of the sentence, is presumably easier to discern.

4.1.2 Statistical analysis

In this section, we demonstrate the statistical significance of the patterns described qualitatively in §4.1.1.

4.1.2.1 Landmark: left edge of the clitic cluster The model in Table 5 predicts the distance of a perceived beep from the left edge of the clitic cluster. This model has a marginal R^2 of .220; once the random effects are considered, the marginal R^2 goes up to .271. Here, and elsewhere, a *positive* coefficient represents *greater* distance from the left edge of the cluster, so *negative* effect sizes represent *stronger* attraction to that landmark.

Random effects	*variance*	*SD*		
Participant	0.03	.18		
Sentence	0.00	.07		
Residual	0.51	.72		
Fixed effects	*β coef*	*SE*	*t value*	*estimated p*
Intercept	**0.43**	**.05**	**8.32**	**<.0001**
Beep location				
linear	**0.28**	**.01**	**22.90**	**<.0001**
quadratic	**0.24**	**.02**	**15.94**	**<.0001**
Condition (default: 4)				
1	**0.19**	**.06**	**3.23**	**.0035**
2	0.09	.06	1.48	.1514
3	**0.15**	**.06**	**2.56**	**.0167**
5	**0.42**	**.06**	**7.28**	**<.0001**

Table 5: Mixed linear regression predicting the distance of experimental responses from the left edge of the clitic cluster (position -1)

The effect of condition is as expected: participants perceived beeps closer to the left edge of the clitic cluster in condition 4 than in any other condition; this difference was significant for all but condition 2. Comparison of the estimated marginal means further confirm that this landmark is significantly more likely for conditions 1–3 than condition 5 (estimated $p \leq .004$). None of the differences between conditions 1–3 are significant.

The quadratic effect of actual beep location is positive, indicating that the distance of responses from position -1 (at the left edge 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 $-.58$ (see Appendix B for an explanation of how this is calculated): that is, participants are most likely to perceive the beep at the left edge of the cluster (position -1) when it is actually at $-.58$—that is, .42 syllables later. This leftward bias is less than half of the general leftward bias of about one syllable described in §4.1.1, matching the smaller leftward bias for beeps located earlier in the sentence.

The random intercepts of participant and sentence in Table 5 are quite small. This suggests that participants were relatively uniform in their perception of beeps, and that there was little difference among individual stimulus sentences.

4.1.2.2 Landmark: right edge of the clitic cluster Table 6 predicts the distance of responses from the right edge of the cluster—that is, the strength of the right edge of the cluster as a perceptual attractor. This model is a substantially better fit than the equivalent model for the left edge in Table 5, with a marginal R^2 of .458 and a conditional R^2 of .554.

As expected, participants perceived beeps significantly closer to the right edge of the cluster in condition 5 than in any other condition. Comparison of the estimated marginal means confirms that this condition stands out from the others: conditions 1–4 were each significantly different from condition 5 (estimated $p < .001$) and not significantly different from one another (estimated $p \geq .603$).

The random intercept for participant has greater variance in this model than in the model for the left edge of the cluster in Table 5, suggesting that speakers showed more variability in how close they heard beeps to the right edge of the cluster than to the left. However, the greater variance may be due in part to the fact that responses near the right edge of the cluster were relatively few in number.

The linear and quadratic factors for beep location indicate that beeps are perceived as closest to

Random effects	*variance*	*SD*		
Participant	0.08	.28		
Sentence	0.01	.10		
Residual	0.41	.64		
Fixed effects	*β coef*	*SE*	*t value*	*estimated p*
Intercept	**1.35**	**.06**	**20.65**	**<.0001**
Beep location				
linear	**−0.58**	**.01**	**−52.72**	**<.0001**
quadratic	**0.05**	**.02**	**3.59**	**.0003**
Condition (default: 5)				
1	**0.48**	**.07**	**6.71**	**<.0001**
2	**0.39**	**.07**	**5.54**	**<.0001**
3	**0.37**	**.07**	**5.26**	**<.0001**
4	**0.42**	**.07**	**5.92**	**<.0001**

Table 6: Mixed linear regression predicting the distance of experimental responses from the right edge of the clitic cluster (position 1)

the right edge of the cluster (1) when they are at 5.93, nearly five syllables to the right. This is not a plausible result. However, the strangeness of this calculation is understandable given the overall leftward bias: if participants are perceiving beeps about one syllable early, as shown in Table 4, then the highest concentration of responses at the right edge of the cluster (position 1) is expected when the beep is located after the following syllable (position 2). However, the last target beep is only at the middle of the following syllable (position 1.5), so for the entire range in which participants heard target beeps (between positions -1.5 and 1.5), the frequency of responses at the right edge of the cluster is expected to monotonically increase. This can be captured with a linear factor. A peak of 5.93 is thus merely notional, and the inclusion of the quadratic effect thus reflects some information about beep location other than the symmetry around a point that motivated its use and proved meaningful in §4.1.2.1.

4.1.2.3 Landmark: left edge of the syllable preceding the clitic cluster In Table 7, we see the model predicting the distance of responses from the syllable boundary one before the left edge of the cluster. In this and the next section, we only compare conditions 1–3. In this case, condition 2—which serves as the baseline—has a prosodic word boundary (between monosyllabic adjective and noun) at this landmark, while conditions 1 and 3 do not (instead, the boundary falls in the middle of a disyllabic noun with stress on the second syllable). Thus, this landmark should be a greater perceptual attractor in condition 2 than in the others. This model has a marginal R^2 of .423 and a conditional R^2 of .532.

There is no significant effect of condition in this model; accordingly, adding the factor of condition to the model barely, and insignificantly, improves its fit ($F = .28$, $p = .759$). Moreover, the small, non-significant difference is in the opposite direction from our predictions: responses in condition 2 are, on average, *further* from the landmark than in conditions 1 and 3. (In the regression whose dependent variable is a binary choice measuring whether a response was at position -2 or not, the difference between conditions is significant, and similarly in the unexpected direction; this finding is discussed in Appendix C.)

The linear and quadratic factors for beep location in this model indicate that responses are closest to position -2 at a beep location of -2.76—that is, almost a syllable *earlier*. This goes against the overall leftward bias. Presumably, this quadratic term is somewhat nonsensical because the landmark is outside of the range of the beeps, so within the actual beep range, the likelihood of a response at this landmark is monotonically decreasing. Thus, as with Table 6, the quadratic term is capturing some property of the data other than non-monotonicity.

Overall, this landmark received relatively few responses, and the distance metric may show some

Random effects	*variance*	*SD*		
Participant	0.13	.35		
Sentence	0.01	.10		
Residual	0.58	.77		
Fixed effects	*β coef*	*SE*	*z value*	*p*
Intercept	**1.25**	**.08**	**16.17**	**<.0001**
Beep location				
linear	**0.66**	**.02**	**39.29**	**<.0001**
quadratic	**0.12**	**.02**	**5.59**	**<.0001**
Condition (default: 2)				
1	–0.05	.07	–0.74	.4715
3	–0.02	.07	–0.26	.8020

Table 7: Mixed logistic regression predicting the distance of experimental responses from the left edge of the syllable preceding the clitic cluster (position −2)

interference from another prosodic word boundary one syllable to the right at the left edge of the cluster.

4.1.2.4 Landmark: right edge of the syllable following the clitic cluster Table 8 shows the model predicting the distance of responses from the syllable boundary one to the right of the right edge of the cluster. Here, condition 3 (the baseline) has a prosodic word boundary separating two monosyllables, while conditions 1 and 2 have the boundary between the two syllables of a disyllabic noun. This landmark should thus be a greater perceptual attractor in condition 3 than in the other two conditions. The resulting model is a similarly good fit to the previous one (marginal $R^2 = .460$, conditional $R^2 = .583$). However, as described below, the model has no explanatory value when it comes to condition.

Random effects	*variance*	*SD*		
Participant	0.13	.37		
Sentence	0.01	.10		
Residual	0.49	.70		
Fixed effects	*β coef*	*SE*	*z value*	*p*
Intercept	**2.77**	**.08**	**36.24**	**<.0001**
Beep location				
linear	**–0.67**	**.02**	**–43.66**	**<.0001**
quadratic	**–0.08**	**.02**	**–4.12**	**<.0001**
Condition (default: 3)				
1	0.09	.07	1.21	.2446
2	0.00	.07	–0.03	.9753

Table 8: Mixed logistic regression predicting the distance of experimental responses from the right edge of the syllable following the clitic cluster (position 2)

This model has no significant effect of condition, and the factor of condition does not significantly improve the model's fit ($F = 1.00$, $p = .390$). This lack of result is not surprising, given that there were very few responses in the vicinity of this landmark (which is after the last beep location), as suggested by the very large positive intercept.

Unlike in the other models, the quadratic factor for beep location in this model is negative: according to the model, the distance from the landmark at position 2 is greatest when the beep is at −4.17 and decreases moving in both directions from there. This is obviously notional: not only are the actual beeps only between positions −1.5 and 1.5, but the beginning of the sentence is at position −3. Within the actual range of beep locations, the distance of response from the landmark is predicted to decrease monotonically as the beeps get closer to the landmark, which is sensible. As in the previous two models, then, the quadratic term is capturing a property of the data other than symmetry about a particular point.

4.2 Discussion

Table 9 compares the predictions from Table 2 with the experimental results presented in §4.1.2. In this section, we discuss these results with respect to the three experimental hypotheses shown in (7).

syllable	*predicted result*	*actual result*	*hypothesis*
left edge of syllable preceding clitic cluster	$2 > 1, 3$	$2 \not> 1, 3$	(7-a)
left edge of clitic cluster	$4 > 1, 2, 3 > 5$	$4 \begin{smallmatrix} > 1, 3 \\ \not> 2 \end{smallmatrix} > 5$	(7-b), (7-c)
right edge of clitic cluster	$5 > 1, 2, 3, 4$	$5 > 1, 2, 3, 4$	(7-b), (7-c)
right edge of syllable following clitic cluster	$3 > 1, 2$	$3 \not> 1, 2$	(7-a)

Table 9: Predicted and actual results by syllable and condition ($X > Y$ means that a given syllable boundary in condition X is a greater perceptual attractor than that syllable in condition Y, $X \not> Y$ means that there was no significant difference)

We will first address the main experimental hypothesis, (7-c), that Slovenian clitics attach by default to the right. Clitics must attach to the right in condition 4 (when they are preceded by a relative clause) and to the left in condition 5 (when they are followed by an adjunct clause); thus, this hypothesis states that the neutral conditions 1–3 should pattern with condition 4 and not with condition 5. This is what we see in the models in §4.1.2.1 and §4.1.2.2: condition 5 has a very different distribution of responses from the other four conditions, a difference that is statistically significant when looking at the distance of responses from either edge of the clitic cluster.

The remaining two hypotheses concerned the nature of the experimental task: we hypothesize that speakers will be more likely to perceive beeps at, and possibly near, syllable boundaries that comprise prosodic boundaries than syllable boundaries in the middle of a word. Like other prosodic phenomena, this effect should be cumulative: beeps should be more perceptually attracted to larger prosodic boundaries than to smaller ones.

The second hypothesis, (7-b), compared the high-level prosodic boundary between the clitic cluster and a subordinate clause to (presumably) lower-level prosodic boundaries between the clitic cluster and other main clause material. Given that clitics patterned as attaching to the right, as discussed above, the relevant comparison for this hypothesis is the left edge of the clitic cluster in condition 4 compared to conditions 1–3. In condition 4, this is the boundary between a relative clause and the clitic cluster, so it should be a greater perceptual attractor of beeps than conditions 1–3, where it is the boundary between the clitics and the preceding sentence-initial disyllabic phrase.

Figure 5 shows a clear difference between these conditions: condition 4 has a distribution clearly centered around the left edge of the cluster, with the sharp peak at the boundary. In conditions 1–3, the distribution is somewhat more diffuse, with a gentler peak on the syllable preceding the boundary and a relatively high number of responses at the boundary itself. This difference was partially confirmed by the model discussed in §4.1.2.1, which showed that responses were significantly closer to the left edge of the clitic cluster in condition 4 than in conditions 1 and 3; the difference between condition 4 and condition 2, though in the same direction, was not significant (although, as discussed in §3.2.2, our choice of model was conservative on this point). From these results, we conclude that our experimental paradigm largely succeeded in detecting differences in prosodic boundary strength: although conditions 1–4 all served as perceptual attractors of beeps, the attraction was more extreme and concentrated in condition 4 than in the others, especially conditions 1 and 3. This is the expected pattern given that the prosodic boundary in condition 4 is stronger than the prosodic boundary in conditions 1–3, which may be stronger than the prosodic word boundary between an adjective and a noun, but is not as strong as the boundary at the end of a subordinate clause.

A possible alternative explanation for the difference between condition 4 and conditions 1–3 is variation in the latter. If some speakers are, at least some of the time, attaching clitics to the left in conditions 1–3, then we would expect the effect on the left edge of the cluster to be weaker in those conditions than in condition 4, where all speakers must be placing a boundary. There are two reasons to doubt this alternative explanation. First, this would likewise predict that conditions 1–3 would have a greater (though likely still small) peak at the right edge of the cluster, which would be a perceptual attractor of beeps some of the time. While §3.2.2 seems to show a slightly lower rate at this edge for condition 4 than the others, this difference was insubstantial and not significant, as discussed in §4.1.2.2. Second, if the difference between the conditions was driven by variation in conditions 1–3, the peak at the left edge in these conditions should look the same as the peak in condition 4, only smaller. However, Figure 5 seems to show more of a qualitative difference: the peak in conditions 1–3 is more spread out across the left edge of the clitic cluster and the syllable preceding it. This is unexpected under the assumption that the boundary in condition 4 is of the same size as the boundary in conditions 1–3 when it exists; the boundaries seem to be different. We return to this qualitative difference at the end of this section.

The first hypothesis, (7-a), looked at the disyllabic phrases on either side of the clitic cluster. The sentence-initial phrase preceding the cluster was a disyllabic noun in conditions 1 and 3 and an adjective–noun pair in condition 2; thus, the boundary between the two syllables of this phrase should be a greater perceptual attractor of beeps in condition 2 than the others. Likewise, the middle of the phrase following the clitic cluster should attract more beeps in condition 3 (where this phrase comprises a monosyllabic adjective and noun) than in conditions 1 and 2 (where it comprises a disyllabic noun).

This hypothesis was not supported. Due to the overall leftward bias in beep perception (that is, participants generally perceived beeps as earlier than they were), there were very few responses to the right of the clitic cluster; although there were slightly more responses here in condition 3 (i.e., 12) than in the other two conditions (8 each), this is too few to say anything substantive. There were somewhat more responses in the middle of the phrase to the left of the cluster. If anything, participants gave visibly *fewer* responses at the middle of the pre-clitic phrase in condition 2, although this difference was not significant by our model (we discuss this difference in Appendix C).

There are several possibilities for this lack of result. Of course, one possibility is that our experimental paradigm cannot, in fact, detect prosodic boundaries at all. However, this interpretation is unlikely given the substantive results from the other hypotheses. It is more likely that the perceptual attractor effect of the low-level prosodic word boundary between an adjective and a noun is too weak, and drowned out by other factors, most notably the larger prosodic boundary at the edge of the clitic cluster one syllable to the right, and possibly also the beginning of the sentence one syllable to the left.

The overall picture presented by the results is that clitics prefer to attach to the right when possible, such that the left edge of the clitic cluster usually has a prosodic boundary—detected by an increased perception of beeps—in that location. This supports the arguments of Orešnik (1984) against the claims of Toporišič (2000) and Golden & Milojević Sheppard (2000), who assume that the Slovenian clitic cluster attaches to the left when possible. It also 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.

Moreover, not all prosodic boundaries behave the same: the higher-level prosodic boundary marking the edge of a subordinate clause attracts beeps more strongly than the (presumably) lower-level boundary between a sentence-initial phrase and the following prosodic unit including the attached clitics. Thus, this experiment serves both as evidence for default rightward attachment of Slovenian clitics and as a proof of concept for the ability of the experimental paradigm to detect prosodic boundaries. It also raises interesting questions for future research about the nature of the task and the prosodic boundary between the pre-clitic phrase and the rest of the sentence.

One open question is the extent to which the beep test can detect low-level prosodic boundaries. We were unable to distinguish between an adjective–noun phrase, which has a prosodic word boundary between its two syllables, and a disyllabic noun, which does not; however, any such effect was likely masked by larger boundaries in the vicinity. Thus, future work should use different stimulus sentences that give the beep test a better opportunity to detect low-level prosodic word boundaries.

Another important question is what exactly the beep test is detecting. In §3.1.2, we discussed two explanations for increased perception of beeps by prosodic boundaries. One 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 in adjacent syllables—if anything, there should be a lower rate of perceived beeps close to boundaries, since it is more convenient to align beeps to boundaries if the boundaries are closer. The other explanation is that prosodic boundaries and their surrounding syllables have increased (actual and perceptual) salience: pauses, phrase-final lengthening, phrase-initial strengthening, and similar (as mentioned at the beginning of §3). If boundaries increase the (actual or perceptual) "bigness" of the surrounding area, we would expect increased perception of beeps both at and near the boundary, since "bigger" or more salient syllables and boundaries have more space for beeps to fall into than others. In this case, we would expect the perceptual attraction effect to be more distributed in the syllable(s) adjacent to the boundary.

The latter pattern is what we see. Condition 4, which has the strongest boundary at the left edge of the cluster, shows a clear peak at the boundary, but the syllables adjacent to this boundary still have a substantially higher rate of responses than units further away. In conditions 1–3, moreover, the syllable preceding the left edge of the cluster has slightly more responses than the boundary itself. This distribution is unexpected, and may reflect the relative salience of the last syllable in intermediate prosodic phrases. For example, this result would suggest that phrase-final lengthening may be detectable in the last syllable of the first (pre-clitic) phrase in Slovenian sentences, which could perhaps be tested using a production study. Future work should also study the nature of the boundary separating clitics from the phrase that precedes them, which seems to be larger than the low-level prosodic word boundary internal to this phrase. Such work may also shed light on the theoretical issue discussed above—namely, that phasal spell-out predicts that encliticization should be the default pattern, contrary to our findings.

5 Conclusion

In this study, we presented a novel experimental paradigm for detecting 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 previous 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 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 is opposite from the predictions of phasal spell-out. Second, we hope that this experimental paradigm serves as a useful addition to the toolbox of perception studies detecting prosodic boundaries, and that future work both uses the paradigm for other questions of prosodic constituency and probes the nature of the experimental task itself in order to figure out what, exactly, is causing beeps to be perceived more often at and around prosodic boundaries.

Data

All 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

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.

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.

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.

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.

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.

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.

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.

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.

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.

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. Reprinted as Škrabec, Stanislav. 1994. Jezikoslovna dela 2. ed. by. Jože Toporišič. Nova Gorica: Frančiškanski samostan Kostanjevica. pp. 212–214.

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.

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.

Zeileis, Achim & Torsten Hothorn. 2002. Diagnostic Checking in Regression Relationships. *R News* 2(3). 7–10. https://CRAN.R-project.org/doc/Rnews/.

A Stimuli

Sentences used in the perception experiment described in §3.

Condition 1

noun	*clitics*		*noun*	*rest*	
Falot	mu	bo	kozo	pomolzel	s strojem.
crook.NOM	him.DAT	FUT.3SG	goat.ACC	milk	with machine
‘The crook will milk his goat with a machine.’					
Bombaš	jim	je	hišo	spremenil	v ruševino.
bomber.NOM	them.DAT	AUX.3SG	house.ACC	changed	in ruin
‘A bomber has turned their house into ruins.’					
Statist	nam	je	sablje	porinil	v roke.
extra.NOM	us.DAT	AUX.3SG	sabers.ACC	pushed	in hands
‘An extra has pushed sabers into our hands.’					
Pastir	ti	bo	krave	pripeljal	s planine.
shepherd.NOM	you.DAT	FUT.3SG	cows.ACC	bring	from mountain
‘A shepherd will bring your cows from the mountains.’					
Oštir	nam	je	žgance	zaračunal	dvojno.
housekeeper.NOM	us.DAT	AUX.3SG	maize.ACC	charged	double
‘The housekeeper has charged us double for maize.’					
Bankir	mi	bo	hišo	zastavil	za kredit.
banker.NOM	me.DAT	FUT.3SG	house.ACC	mortgage	for loan
‘The banker will mortgage my house for a loan.’					

Condition 2

adjective	*noun*	*clitics*		*noun*	*rest*	
Nov	stol	so	mu	fantje	podarili	šele včeraj.
new	chair.ACC	AUX.3PL	him.DAT	boys.NOM	gave	just yesterday
‘The boys gave him a new chair just yesterday.’						
Bel	prt	mi	bo	Petra	prinesla	za božič.
white	tablecloth.ACC	me.DAT	FUT.3SG	Petra.ACC	bring	for Christmas
‘Petra will bring me a white tablecloth for Christmas.’						

Fejst	fant	nam	bo	sobo	popleskal	zastonj.
good	guy.NOM	us.DAT	FUT.3SG	room.ACC	paint	for free

‘A good guy will paint our room for free.’

Lep	dan	vam	je	Metka	zagotovila	z izletom na Bled.
nice	day.ACC	you.DAT	AUX.3SG	Metka.NOM	guaranteed	with trip on Bled

‘Metka guaranteed you a nice day with a trip to Bled.’

Dolg	drog	mu	bo	Micka	pustila	za hišo.
long	stick.ACC	him.DAT	FUT.3SG	Micka.NOM	leave	behind house

‘Micka will leave the long pole behind his house.’

Suh	bor	ti	bo	danes	polomil	vrtno ograjo.
dry	pine.NOM	you.DAT	FUT.3SG	today	break	garden fence.ACC

‘A dry pine will break your garden fence today.’

Condition 3

noun	*clitics*		*adjective*	*noun*	*rest*	
Bandit	nam	je	ves	čas	kazal	nakradeno robo.
bandit.NOM	us.DAT	AUX.3SG	all	time	showed	stolen goods.ACC

‘The bandit kept showing us the stolen goods.’

Graščak	ji	bo	ves	rž	odvzel	brez razloga.
manor.NOM	her.DAT	FUT.3SG	all	rye.ACC	take	without reason

‘The manor will take away all the rye from her for no reason.’

Gardist	nam	je	več	let	razlagal	vojaško strategijo.
guardsman.NOM	us.DAT	AUX.3SG	many	years	explained	military strategy.ACC

‘For several years, the guardsman explained military strategy to us.’

Lingvist	nam	je	dva	dni	govoril	le o členkih.
linguist.NOM	us.DAT	AUX.3SG	two	days	talked	only about particles

‘For two days, the linguist talked to us only about particles.’

Voznik	jim	je	lep	dan	spremenil	v najgrše popoldne.
driver.NOM	them.DAT	AUX.3SG	nice	day.ACC	turned	in ugliest afternoon

‘The driver turned their beautiful day into the ugliest afternoon.’

Poljak	mi	bo	bel	trak	pobarval	na rdeče.
Pole.NOM	me.DAT	FUT.3SG	white	ribbon.ACC	color	on red

‘The Pole will paint my white ribbon red.’

Condition 4a

noun	*relative clause*	*clitics*		*noun*	*rest*
Kolo,	ki sem ga kupil včeraj,	so	mi	danes	ukradli.
bike.ACC	which AUX.1SG it.ACC bought yesterday	AUX.3PL	me.DAT	today	stole

‘The bike I bought yesterday was stolen today.’

Knjigo,	ki je izšla včeraj,	vam	je	Marko	dal za darilo.
book.ACC	which AUX.3SG published yesterday	you.DAT	AUX.3SG	Marko.NOM	gave for gift

‘Marko gave you the book that was published yesterday as a gift.’

Fantu,	ki ga je Micka predlagala,	se	bo	Tinka	zaupala.
boy.DAT	which him.ACC AUX.3SG Micka.NOM suggested	REFL.ACC	FUT.3SG	Tinka.NOM	trust

‘Tinka will trust the boy whom Micka proposed.’

Condition 4b

noun	*relative clause*	*clitics*		*adjective*	*noun*	*rest*
Punca,	ki je vedno oblečena v belo,	ti	bo	siv	pas	vrgla stran.

girl.NOM	which is.3SG always dressed in white	you.DAT	FUT.3SG	gray	belt.ACC	throw away	

‘A girl who always wears white will throw your gray belt away.’

Zvezek,	ki mi je padel na tla,	mi	bo	njen	stric	vrgel v koš.
notebook.ACC	which me.DAT AUX.3SG fell on ground	me.DAT	FUT.3SG	her	uncle.NOM	throw in trash

‘Her uncle will throw the notebook that I dropped on the floor in the trash.’

Denar,	ki nam ga je dolgoval,	nam	je	tvoj	brat	mogoče odplačal z delom.
money.ACC	which us.DAT it.ACC AUX.3SG owed	us.DAT	AUX.3SG	your	brother	maybe repaid with work

‘Your brother may have paid us the money he owed us with work.’

Condition 5a

noun	*clitics*		*adjunct clause*	*rest*
Dijak	mi	bo,	ko se bo začel pouk,	napisal tri listke.
student.NOM	me.DAT	FUT.3SG	when REFL.ACC FUT.3SG begin class.ACC	write three notes.ACC

‘When the class starts, the student will write me three notes.’

Dirkač	nam	je,	ker smo ga zelo lepo prosili,	pokazal svojo formulo.
racer.NOM	us.DAT	AUX.3SG	because AUX.1PL him.ACC very nicely asked	showed his formula.ACC

‘The racer, because we asked him very nicely, showed us his formula.’

Nečak	vam	bo,	četudi se s tem ne strinja,	odigral tri sonate.
nephew.NOM	you.DAT	FUT.3SG	even REFL.ACC with this NEG agree	play three sonatas.ACC

‘Your nephew will play three sonatas for you, even if he doesn’t agree.’

Condition 5b

adjective	*noun*	*clitics*		*adjunct clause*	*rest*
Dva	dni	sem	ti,	ko še nisem poznal razlogov,	res hotel odpustiti.
two	days	AUX.1SG	you.DAT	when yet NEG.AUX.1SG knew reasons	really wanted forgive.INF

‘For two days, when I didn’t yet know the reasons, I really wanted to forgive you.’

Vso	noč	nam	je,	kot da ni bilo že vsega dovolj,	Peter prepeval Čuke.
all	night	us.DAT	AUX.3SG	as that NEG AUX.3SG been already everything enough	Peter.NOM sang Čuki.ACC

‘All night long, as if everything else wasn’t enough, Peter sang Čuki to us.’

Tvoj	pes	mu	bo,	četudi ga ne bo niti pogledal,	gotovo strgal hlače.
your	dog.NOM	him.DAT	FUT.3SG	even him.ACC NEG FUT.3SG not-even look	surely rip pants.ACC

‘Your dog, even if he doesn’t even look at him, will surely rip his pants.’

B Interpreting the quadratic effect of beep location

The linear regression models in §4.1 include a linear and quadratic term for the actual location of the beep in the stimulus sentences. In this section, we show how to calculate the “optimal” beep location for a given model—that is, the beep location which is predicted to have the (usually) smallest value for the dependent variable, distance from a particular landmark.

The coefficients (effect sizes) for a regression model are used to predict the value of the dependent variable given the independent variable. The models in this paper have a term for the intercept, β_0; a linear term for beep location, β_1; a quadratic term for beep location, β_2, and terms for the effect of each condition i, β_{c_i}. (The baseline condition does not have a β_{c_i} term, since that is the default against which the other conditions are measured.) Thus, the predicted value y for the dependent variable for a given trial can be calculated from the model coefficients as follows, where x is the beep location and $I(c_i)$ is an indicator function equal to 1 when the trial is in condition c_i and 0 otherwise (that is, at most one of the β_{c_i} terms in the sum will be non-zero in any given trial; when the condition is the baseline, they will all be zero):

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \sum \beta_{c_i} I(c_i)$$

For example, the model predicting the distance of experimental responses from the left edge of the clitic cluster (-1) in Table 5, the equation is:

$$y = .43 + .28x + .24x^2 + .19 \cdot I(c_i) + .09 \cdot I(c_2) + .15 \cdot I(c_3) + .42 \cdot I(c_5)$$

The presence of a quadratic term means that the relationship between beep location, x, and the predicted distance from the landmark, y, will form a parabola, which is symmetrical about a single point, either a maximum or a minimum. In the case of Table 5, the quadratic term β_2 is positive, so the parabola has a minimum: there is a certain beep location at which responses are predicted to be closest to the left edge of the cluster. As Figure 6 shows, this minimum is somewhere between positions 0 and -1 (that is, the first clitic). The effect of condition will be to shift the parabola up or down, but not side-to-side: in our model, condition has no interaction with the optimal beep location.

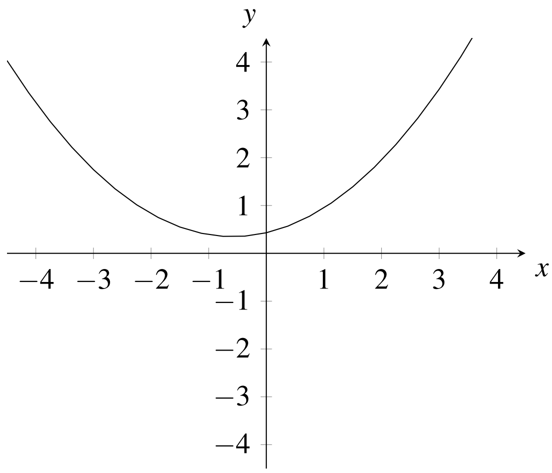

Figure 6: Relationship between beep location (x) and predicted distance of response from the left edge of the cluster (y) for the model in Table 5

To find this minimum, we must calculate the inflection point of the regression equation, which is the point at which the first derivative of the regression equation with respect to beep location is equal to 0:

$$\begin{aligned} \frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}(\beta_0 + \beta_1 x + \beta_2 x^2 + \sum \beta_{c_i} I(c_i)) = 0 \\ 0 &= \beta_1 + 2\beta_2 x \\ -2\beta_2 x &= \beta_1 \\ x &= -\frac{\beta_1}{2\beta_2} \end{aligned}$$

Thus, for the model in Table 5, the beep location at which responses are predicted to be at the minimum distance from the left edge of the clitic cluster is:

$$\begin{aligned} x &= -\frac{.28}{2 \cdot .24} \\ &= -.58 \end{aligned}$$

C An apparent absolute condition effect in the pre-clitic phrase

In §3.1.2, we described two ways to quantify perceptual attraction of beeps. In the models presented in §4.1, we predicted the distance of responses from a given landmark using mixed linear regression. The alternative (found in the supplemntary materials) is to use logistic regression to predict a binary outcome: whether or not a given response was *exactly at* a given landmark. The two calculations yield the same qualitative results, with one exception. The model in Table 7 finds no effect of condition on the distance of responses from the left edge of the syllable preceding the clitic cluster (-2)—that is, the boundary between the two syllables of the sentence-initial phrase before the clitic in conditions 1–3. In condition 2, this boundary separates an adjective from a noun, while in conditions 1 and 3, it is in the middle of a disyllabic noun. Accordingly, we predicted that this boundary should be a stronger perceptual attractor in condition 2 than in the other two conditions. This prediction was not confirmed.

The model in Table 10 predicts whether a given experimental response is at the syllable boundary between the two syllables of the pre-clitic phrase (position -2) in conditions 1–3. It uses the same predictors as the model in Table 7. In this case, a positive effect size means an increased likelihood of response at the landmark at position -2, so unlike the other models in the study, here a *positive* effect size indicates *greater* perceptual attraction. According to this model, the boundary between the two pre-clitic

Random effects	*variance*	*SD*		
Participant	0.70	.84		
Sentence	0.10	.31		
Fixed effects	*β coef*	*SE*	*z value*	*p*
Intercept	**–4.50**	**.35**	**–12.77**	**<.0001**
Beep location				
linear	**–1.38**	**.18**	**–7.78**	**<.0001**
quadratic	**0.34**	**.16**	**2.21**	**.0274**
Condition (default: 2)				
1	**0.79**	**.31**	**2.58**	**.0099**
3	**0.69**	**.31**	**2.22**	**.0262**

Table 10: Mixed logistic regression predicting whether an experimental response is at the left edge of the syllable preceding clitic cluster (position -2)

syllables is a significantly *worse* perceptual attractor in condition 2 than in the other two conditions—the opposite of the predicted result.

This result is not due to a direct effect of condition, but rather the conflation of condition with another factor: the *length* of the pre-clitic phrase. These phrases tend to be longer in condition 2, where they comprise two words, each with a stressed syllable, than in conditions 1 and 3, where the phrase is a single word with one stressed and one unstressed syllable. Indeed, Figure 7 shows a negative correlation between the length of the pre-clitic phrase and the number of responses at the syllable boundary in the middle of this phrase: the longer the phrase, the fewer responses at this landmark. As predicted, most of the longer phrases—which thus have fewer responses at the landmark—are in condition 2.

Taken together, the effects shown in Figure 7 derive the results in Table 10: responses are more likely between the pre-clitic syllables in condition 2 than in condition 1 and 3. Pre-clitic phrase length is a better, more direct predictor of the likelihood of a response at the left edge of the syllable before the clitic cluster (position -2) than condition. First of all, condition is moderately correlated with phrase length ($R^2 = .412$), so the effect of condition in Table 10 is plausibly an indirect effect of phrase length. Second, the equivalent of Table 10 that replaces condition with pre-clitic phrase length—that is, a model with random intercepts for participant and sentence, linear and quadratic fixed effects of beep location, and a fixed linear effect of phrase length—does significantly better than the condition-based model in

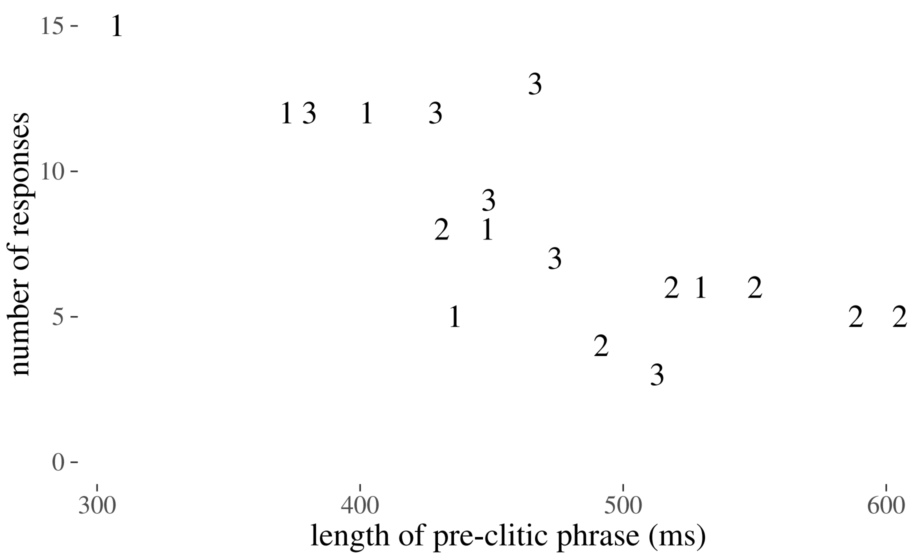

Figure 7: Experimental responses for individual sentences (labelled with their condition) at the left edge of the syllable before the cluster (-2), by length of sentence-initial pre-clitic phrase

Table 10 ($\chi^2 = 13.89$, $p < .001$).[8] We thus conclude that the relevant effect in predicting responses close to the beginning of the sentence is phrase length, not condition.

The pre-clitic phrase length effect shown in Figure 7 is likely related to the tendency of participants to perceive beeps earlier than their actual location. So far, we have discussed this leftward bias in terms of syllables (i.e. beeps were perceived, on average, about one syllable earlier than they actually occurred, though the difference was smaller towards the beginning of the sentence). However, it is possible that this leftward bias is dependent on actual time rather than, or in addition to, syllable timing. In that case, the shorter the pre-clitic phrase, the more beeps will be perceived in the region prior to their actual location (which is no earlier than position -1.5, the middle of the last syllable before the clitic cluster).

The negative correlation between pre-clitic phrase length and responses at the boundary between the syllables of this phrase shown in Figure 7 is not quite linear—in particular, the three sentences in condition 2 with the longest phrases have more responses than expected. This might be a slight effect of the prosodic word boundary, acting as a perceptual attractor of beeps at this location, though this effect is in general outweighed by the phrase length effect. Alternatively, it may simply be a floor effect—there will always be a few responses at this landmark for any sentence, so even the sentences with longer phrases will not go all the way down to 0 as would be expected given the trend in Figure 7. For now, we cannot report any reliable, substantive effect of condition for this landmark.

[8]Since the two models are not nested, an ANOVA is inappropriate, so we compare the two models using a likelihood ratio test with the `lrtest` function from version 0.9-40 of R's *lmtest* package (Zeileis & Hothorn 2002).