STAR

Werkmann Horvat, Ana & Birtić, Matea & Gračanin-Yuksek, Martina. 2025. How does 'do so' do? Testing the applicability of the 'do so' test in Croatian. *Syntactic Theory and Research* 1.1.6. https://doi.org/10.16995/star.16974.

Open Library of Humanities

How does 'do so' do? Testing the applicability of the 'do so' test in Croatian

Ana Werkmann Horvat, University of Osijek, HR, awerkmannhorvat@ffos.hr

Matea Birtić, Institute for Croatian Language and Linguistics, HR, mbirtic@ihjj.hr

Martina Gračanin-Yuksek, Middle East Technical University, TR, mgy@alum.mit.edu

The argument–adjunct distinction is tacitly present in many grammatical theories and has been supported by psycholinguistic research on various languages. One way of testing this distinction is through the 'do so' test, which involves replacing a projection of the verb with the pro-form 'do so.' The current study undertakes two experiments, an acceptability judgment task and a self-paced reading task; in both, the 'do so' test is applied to Croatian verbs that can take theme NP arguments in both the accusative case and the instrumental case. We tested sentences in which 'do so' strands an accusative or instrumental theme of such a verb. We also tested similar sentences with instrumental-marked noun phrases denoting instruments and prepositional accusative phrases denoting direction. Our first goal in conducting the study was to investigate whether the 'do so' test, applied to Croatian data, can discriminate between VP-internal arguments and adjuncts when both are instrumental noun phrases (based on a number of diagnostics, we took instrumental-marked themes to be arguments and instrumental-marked instruments to be adjuncts). Our second goal was to assess whether the results of the 'do so' test in Croatian would be better explained by appealing to the structural status of the remnant constituent (the **argument–adjunct explanation**) or to the lexical requirements of the verb 'do,' paired with an identity requirement on the theta roles assigned to the constituent following 'do so' and its antecedent (the **lexical explanation**). The experimental results we obtained did not support either approach. In both experiments, instrumental arguments and instrumental adjuncts behaved the same (both yielded grammatical results when stranded by 'do so' substitution), while accusative themes behaved differently from prepositional accusative adjuncts (accusative themes could not be stranded by 'do so' substitution, but prepositional phrases could).

Keywords: arguments; adjuncts; 'do so' test; Croatian

1 Introduction

Establishing ways to detect differences between arguments and adjuncts in different languages is important because the two categories are at least tacitly present as universal categories in most grammatical analyses and across many theoretical frameworks (Tesnière 1959, Chomsky 1981, R. Kaplan & Bresnan 1982, Van Valin 2001, etc.). However, drawing a clear line between arguments and adjuncts is difficult even in a single language, let alone universally. As Koenig et al. 2003 puts it: "while most linguists agree that the distinction between arguments and adjuncts is real, no consensus currently exists as to its basis, the boundary between the two classes, or its role in grammar" (68). This has led to the emergence of grammatical approaches that either ignore the distinction (e.g., Vater 1978, Przepiórkowski 1999) or suggest that it is not binary (e.g., Grimshaw 1990, DeArmond & Hedberg 1998) or is a gradient property (e.g., Manning 2003, Forker 2014).

On the other hand, although there are some conflicting results, most experimental studies show that speakers process arguments and adjuncts differently (see Tutunjian & Boland 2008 for an overview). The difference has been demonstrated both in comprehension and (to a somewhat lesser degree) in production and in various modalities—in listening, reading, and visual world experiments—as well as from different perspectives: for instance, some studies are interested in the lexical representation of argument structure, while others focus on different attachment sites for arguments and adjuncts.

Through an eye tracking experiment that measured eye fixations on argument and adjunct prepositional phrases in sentences, Boland & Blodgett 2006 found that argument prepositional phrases were processed faster than adjunct prepositional phrases (as indicated by duration of first fixation, first pass reading times, and total reading times), suggesting that arguments and adjuncts have different statuses in the grammar. Similarly, Lee & Thompson 2011 documented an experiment in which participants' eye movements were recorded while they were orally producing sentences using verbs that were provided and nouns that were represented as pictures on the screen. The study found a higher processing cost associated with adjuncts than with arguments in both younger and older speakers, as well as in agrammatic aphasic speakers. Another study, Boland 2005, tracked participants' eye movements as they looked at pictures while listening to sentences that mentioned individuals

in the pictures. The focus of the study was on whether participants would cast anticipatory looks at pictures representing entities that were potential arguments or adjuncts of the verb before they were mentioned in the sentence. The study showed that potential arguments received more anticipatory looks than potential adjuncts, again suggesting different statuses for the two kinds of constituents.

Speer & Clifton 1998 and Kennison 2002 reached the same conclusion on the basis of both self-paced reading experiments and eye tracking experiments. Speer & Clifton's experiments used sentences in which the verb was followed by a noun phrase that was either an argument of the verb or an adjunct and was either plausible or implausible in the context of the sentence. Speer & Clifton found that arguments were processed faster than adjuncts in both plausibility conditions. This finding is consistent with the results of the later study, Kennison 2002, which measured how fast participants processed argument and adjunct noun phrases (*every play*/*every week*) following transitively and intransitively biased verbs (*read*/*perform*). Both in her self-paced reading experiment and in her eye tracking experiment, Kennison found that, at least following transitively biased verbs, argument noun phrases were read faster than adjunct noun phrases, suggesting an increased processing cost for adjuncts.

Importantly, there has recently been a rise in psycholinguistic studies that explore the privileged status of arguments with respect to adjuncts in languages other than English. For instance, on the basis of two eye tracking experiments, Akal 2017 showed that *wh* arguments of the verb in Turkish are processed faster than *wh* adjuncts, as evidenced by first fixation durations and total fixation durations, both of which were longer in the case of *wh* adjuncts than *wh* arguments. An interesting study that used recall as an experimental method, Chromý & Vojvodić 2024, explored the argument–adjunct distinction in Czech via four experiments, which were performed to assess immediate post-sentential recall of core sentence information conveyed by direct objects as well as of optional information conveyed by temporal or locative adjuncts. Participants were asked questions about the arguments and adjuncts in the sentences they had read. The study found that core information conveyed by direct objects was almost perfectly recalled while additional information conveyed by temporal and locative adjuncts was recalled with significantly lower accuracy rates.

In addition to these studies, Andreu et al. 2016 explored the role of verb semantics in rapidly anticipating upcoming information during real-time sentence comprehension

in four groups of L1 Spanish speakers: three groups of children, one of which had specific language impairment, and one group of adults. The visual world paradigm was used to compare anticipatory looks at themes, sources/goals, instruments (arguments), and locatives (adjuncts). The results showed that the proportion of looks at the theme, source/goal, and instrument referents was significantly higher than that of looks at locatives. This pattern was consistent across all four groups. These findings support the claims that verbs implicitly introduce their arguments as they are retrieved and that the status of arguments and adjuncts differs.

Based on the experimental results just reviewed, as well as the theoretical literature, we assume that the distinction between arguments and adjuncts is real.[1] We note, however, that the 'do so' replacement test (Lakoff & Ross 1976), which is often cited in the theoretical syntactic literature (both research texts and textbooks) as a means for distinguishing between complements[2] and adjuncts (e.g., Zwicky 1970, Gazdar et al. 1982, Zagona 1988, Bresnan 1994, Haegeman 1994, and Baxter 1999, to name just a few), has been almost completely neglected in the experimental literature.

In the 'do so' replacement test, (a part of) a VP is replaced by the pro-form 'do so,' which is anaphoric to (a part of) an antecedent VP, present in the same sentence or in the surrounding discourse. When the verb in the two VPs appears with a direct object, 'do so' can substitute for the entire VP, as shown in (1a), but not for the verb alone, as shown in (1b).

(1) a. *John ate an apple and Bill did so, too.*
[*did so* = ate an apple]

b. **John ate an apple and Bill did so a pear.*
[*did so* = ate]

On the other hand, the phrase *today* can survive the 'do so' replacement:

[1] A reviewer worries that the "cited [psycholinguistic] studies are not controlled in the ways necessary to suggest that speakers distinguish between arguments and adjuncts in the way intended in the standard grammatical sense." This is probably true, but we consider the results of the cited studies sufficient to further motivate exploring the argument–adjunct distinction experimentally.

[2] In the context of the 'do so' test, it is common for authors to primarily discuss complements, which are a specific type of argument (namely, internal arguments). This is because in the 'do so' test, the replaced element typically is a complement, rather than a different type of argument. Considering the particular emphasis on the 'do so' test in this paper, we will use both terms interchangeably, assuming their equivalence within the scope of this study.

(2) *John ate an apple yesterday and Bill did so today.*
[*did so* = ate an apple]

The one study we found that examines the 'do so' replacement test in an experimental setting is Kweon 2002, which investigated L2 acquisition of English argument structure by L1 Korean speakers. In the study, Kweon used a group of L1 English speakers for control purposes and collected their judgments on sentences that contained the pro-form *do so* followed by a prepositional phrase that was either a complement (with 'put'-type verbs), as in (3a), or an adjunct (with 'bathe'-type verbs), as in (3b). Kweon found that L1 English speakers rejected sentences like the one in (3a), in which *do so* is followed by an argument prepositional phrase, but accepted sentences like the one in (3b), where *do so* is followed by an adjunct prepositional phrase.[3]

(3) a. *Tom placed the book on the table. *And Sue did so on the chair.*
b. *Bill bathed the dog in the kitchen. And Martha did so in the bathroom.*
(Kweon 2002: 12)

Appealing to different structural statuses of the phrases stranded by the 'do so' substitution is a (standard) way to explain the contrast between (1b) and (2). In (1b), *a pear* is a direct object, the complement of the verb, therefore an argument. In (2), on the other hand, *today* is a time adverbial, therefore an adjunct. On this view, the 'do so' replacement test is sensitive to the argument/adjunct status of the stranded constituent. We refer to this kind of explanation of the contrast between (1b) and (2) as the **argument–adjunct explanation.**

On the other hand, the test has also been criticized for yielding results that are unexpected if the 'do so' pro-form replaces the projection of the verb as opposed to the verb alone (e.g., Miller 1992, Kehler & Ward 1999, Przepiórkowski 1999, Culicover & Jackendoff 2005, Houser 2010, McInnerney 2022a). The critics argue that the 'do so' replacement is not a test that discriminates between arguments and adjuncts, and they offer alternative explanations for the results that it yields. In this article, we consider one such alternative, proposed by Miller 1992 and adopted by

[3] L2 speakers, on the other hand, accepted both types of sentences. Kweon hypothesized that the Korean L2 English speakers had not acquired the argument structure of English verbs and were acting under the influence of their L1, in which the 'do so' test does not draw a sharp distinction between 'put'-type verbs and 'bathe'-type verbs.

Przepiórkowski 1999, McInnerney 2022a, and McInnerney 2022b, that accounts for the results of the 'do so' replacement by appealing to the lexical properties of the verb 'do' and the compatibility or incompatibility in theta roles of the constituent stranded by 'do so' and its antecedent. We refer to this explanation as **the lexical explanation.**

Our goal in this study is twofold.

In light of the lack of experimental treatment of the 'do so' replacement test, our primary goal is to experimentally examine whether the test discriminates between VP-internal arguments and adjuncts. To this end we applied the test to Croatian data in two experiments: an acceptability judgment task and a self-paced reading task.[4] In particular, we are interested in whether the test can discriminate between two argument–adjunct pairs. On the one hand, we examine whether the 'do so' test treats differently accusative noun phrases denoting themes (standardly considered arguments) and prepositional phrases headed by accusative-assigning prepositions and denoting direction/location (standardly considered adjuncts). On the other hand, we investigate whether the difference persists between instrumental noun phrases denoting themes (which we consider to be arguments) and instrumental noun phrases denoting instruments (which we consider to be adjuncts).[5] If the argument–adjunct explanation is correct, the 'do so' replacement should yield grammatical results in Croatian whenever the constituent following 'do so' is an adjunct and ungrammatical results whenever it is an argument, regardless of the syntactic makeup or case properties of the stranded constituent.

Our second goal is to assess whether the results of the 'do so' replacement test in Croatian can (also) be accounted for by the lexical explanation.

We will see that neither explanation can successfully account for the results we obtained but that the data fit the predictions of the argument–adjunct explanation somewhat better than the predictions of the lexical explanation.

The paper is organized as follows. In section 2 we outline our study and motivate our assumptions that Croatian instrumental noun phrases denoting themes are arguments and that those denoting instruments are adjuncts. In section 3 we present the details of how the argument–adjunct explanation and the lexical explanation

[4] Brač 2018 is one of the first sources that mentions the 'do so' substitution as a possible test for argumenthood in Croatian.

[5] For evidence that supports our assumptions about instrumental noun phrases, see section 2.

account for the results of the 'do so' replacement test, and we state their predictions regarding our experiments. Section 4 presents our experimental stimuli, procedures, and results, which we then discuss in section 5, both from the point of view of the argument–adjunct explanation and from that of the lexical explanation. Section 6 is the conclusion.

2 Instrumental noun phrases in Croatian: argument/adjunct status thereof

2.1 Assumptions made in the context of the present study

We conducted two experiments: an offline acceptability judgment task and an online self-paced reading task. In both, we tested whether instrumental-marked noun phrases behave differently on the 'do so' test when they are themes than when they are instruments. Our experimental sentences included Croatian verbs whose themes can appear both in the accusative case and in the instrumental case (e.g., *micati* 'move,' *tresti* 'shake,' *trzati* 'pull, jerk,' *rukovoditi* 'command'); most though not all of these verbs involve the movement of body parts or objects (Brač 2018: 128).[6]

We assumed that, despite the difference in case marking, instrumental-marked themes have the same status as accusative-marked themes: that of the complement of the verb, that is, that of an argument. We based this assumption on several semantic and syntactic criteria, discussed in detail in the next subsection: these noun phrases' semantic contribution to the event denoted by the verb and their behavior with respect to selection, omission, coordination, and multiple *wh* fronting. By all of these criteria, instrumental-marked themes behave like arguments.

On the other hand, we treated instrumental-marked phrases denoting instruments as adjuncts. This is not a trivial assumption. The status of instruments as arguments or adjuncts is highly controversial, with some treating them as arguments (e.g., Bresnan 1982, Schütze 1995), some as adjuncts (e.g., Van Valin & LaPolla 1997, Dowty 2003), and yet others as an in-between category (e.g., Rissman 2010, Needham & Toivonen 2011, Rissman et al. 2015, Barbu & Toivonen 2016, Barbu 2020). Finally, some distinguish two groups of instruments, those that are arguments and those that are adjuncts,

[6] In all we used 16 such verbs: *cimati* 'jerk,' *drmati* 'shake,' *klimati* 'nod,' *ljuljati* 'swing,' *lupati* 'bang,' *manipulirati* 'manipulate,' *micati* 'move,' *mrdati* 'move,' *njihati* 'swing,' *rukovoditi* 'command,' *tresti* 'shake,' *trzati* 'pull, *jerk*,' vitlati 'swirl,' *vrtjeti* 'spin,' *zabacivati* 'throw,' and *zakretati* 'turn.'

based on how necessary they are for the completion of an action or event (Croft 1991, Schlesinger 1995, Koenig et al. 2008, Rissman 2011, Belaj & Tanacković Faletar 2017, Brač 2018). In section 2.3, we present evidence from Croatian that instrumental noun phrases denoting instruments are adjuncts. The evidence comes from selection, omission, iterability, coordination, and multiple *wh* fronting. We will see that, on each of these diagnostics, instrumental noun phrases denoting instruments behave differently from both accusative themes and instrumental themes.

2.2 Evidence that instrumental themes are arguments

Our assumption that instrumental-marked themes are arguments is motivated by the intuition that they contribute the same semantic import to the event as their accusative counterparts do (at least in the sentences that we tested, where the verb can equally be followed by an accusative or instrumental noun phrase). A sentence with an accusative-marked theme and a corresponding sentence with an instrumental-marked theme are semantically equivalent: both denote the same event and are true under the exact same set of circumstances. Thus (4), for example, has the same translation regardless of case.

(4) *Petar drma koljeno/koljenom.*
Petar shakes knee.ACC/knee.INS
'Petar shakes/is shaking the knee.'

Since accusative-marked themes and instrumental-marked themes are both noun phrases and since either one can be used in the sentence without a change in the meaning of the sentence, it is reasonable to assume that both have the same status: that of an argument.

Furthermore, instrumental-marked themes appear with only about 80 verbs in Croatian (Katičić 2002, as cited in Brač 2018). These verbs fall into several semantic categories; different authors propose different classifications, with variation in the number of proposed categories and their semantic properties. Here we report the classification in Brač 2018, which divides these verbs into three classes: (i) verbs of control, command, and trade, (ii) verbs of dealing, playing, and bragging, and (iii) verbs of body part movement. The fact that instrumental-marked themes occur only with specific verbs suggests that they are subcategorized by these verbs, that is, that they are arguments (Borsley 1991, Koenig et al. 2003, Forker 2014).

Similarly, instrumental-marked themes are largely obligatory in a sentence. Of the 16 verbs used in our experiments that can appear with either an accusative theme or an instrumental theme, only one or two, *klimati* 'nod' and possibly *lupati* 'bang,' can appear without an object altogether; with all the others, the theme is obligatory. The fact that instrumental-marked themes (as well as accusative-marked themes) are not omissible also suggests that they are arguments rather than adjuncts (McInnerney 2022b).

Finally, instrumental-marked themes behave as arguments on two syntactic tests that can discriminate between arguments and adjuncts. The first one is coordination; it is commonly assumed that arguments can only be coordinated with arguments and adjuncts only with adjuncts (Radford 1988). An instrumental-marked theme can only be coordinated with another instrumental-marked theme, not with an instrumental noun phrase denoting manner (uncontroversially an adjunct):[7,8]

(5) a. $\text{INS}_{\text{theme}}$ & $\text{INS}_{\text{theme}}$: ✓
Sara velikom brzinom trza žicom i strunom.
Sara great.INS speed.INS strums wire.INS and string.INS
'Sara strums/is strumming the wire and the string at great speed.'

b. $\text{INS}_{\text{theme}}$ & $\text{INS}_{\text{manner}}$: ✗
**Sara trza žicom i velikom brzinom.*
Sara strums wire.INS and great.INS speed.INS
'*Sara strums/is strumming the wire and at great speed.'

Our last piece of evidence that instrumental-marked themes are arguments comes from multiple *wh* fronting. Bošković 1998 and 2010 show that multiple questions that contain only fronted adjunct *wh* phrases are ungrammatical in (Serbo-) Croatian:

---

[7] In the coordination test, we use instrumental noun phrases (instead of, for example, prepositional phrases) so as not to confound the judgments with a violation of the Law of the Coordination of Likes, according to which elements can be coordinated only if they are of the same syntactic category (Williams 1981). Since both manner-denoting instrumentals and theme-denoting instrumentals are noun phrases (with identical case marking), the impossibility of coordinating the two cannot stem from a violation of the Law of the Coordination of Likes. For the same reason, in (9, 10) and (11, 12) below we employ accusative and genitive adjunct noun phrases respectively.

[8] Likewise, an instrumental-marked theme cannot be coordinated with an instrumental-marked instrument:

(i) **Sara trza žicom i komadićem plastike.*
Sara strums wire.ins and piece.INS plastic.GEN
Intended: 'Sara strums/is strumming the wire and with a piece of plastic.'

This is consistent with our assumption that instruments are adjuncts. See section 2.3 for argumentation.

(6) *Zašto je kako istukao Petra?
why AUX how beaten Petar
'Why did he beat Petar how?'
(Bošković 2010: 38)

However, an instrumental-marked *wh* theme (čime 'what.INS') *can* be fronted with adjunct *wh* phrases (*kako* 'how,' *kada* 'when,' …), as in (7a), and also with an instrumental *wh* phrase denoting manner, as in (7b).[9]

(7) a. $INS_{theme,\ wh}$ $adjunct_{wh}$: ✓
Čime je kada/kako/zašto Sara trzala?
what.INS AUX when/how/why Sara strum
'What did Sara strum when/how/why?'

b. $INS_{theme,\ wh}$ $INS_{manner,\ wh}$: ✓
Čime je kakvom brzinom Sara trzala?
what.INS AUX what-kind-of.INS speed.INS Sara strum
'What did Sara strum at what speed?'

The grammaticality of (7a, b) suggests that instrumental themes are not adjuncts, that is, that they are arguments.[10] Thus, predictably, the addition of an accusative theme argument to (7a, b) results in ungrammaticality, as shown in (8). Here the presence of the theme argument žicu 'wire.ACC' forces the instrument interpretation of the fronted instrumental *wh* phrase, with the result that both of the fronted *wh* phrases are adjuncts, which is disallowed.

(8) a. $INS_{instrument,\ wh}$ $adjunct_{wh}$: ✗
Čime je kada/kako/zašto Sara trzala žicu?
what.INS AUX when/how/why Sara strum wire.ACC
Intended: 'With what did Sara strum the wire when/how/why?'

---

[9] A manner-denoting instrumental *wh* phrase cannot be fronted with another adjunct *wh* phrase:

(i) *Kakvom brzinom kada/zašto Sara trza žicom?
what-kind-of.INS speed.INS when/why Sara strums wire.INS
Intended: 'When/why does Sara strum the wire at what speed?'

This supports its status as an adjunct.

[10] The fact that the two *wh* phrases are separated by the clitic *je* 'be.3SG' does not mean that both have not been fronted. The placement of *je* 'be.3SG' is due to the robust requirement in Croatian that clitics be second in their clause (Browne 1974, Halpern 1992, Progovac 1996, and Cavar 1999, among others).

b. $INS_{instrument,\ wh}$ $INS_{manner,\ wh}$: ✗
*Čime je kakvom brzinom Sara trzala žicu?
what.INS AUX what-kind-of.INS speed.INS Sara strum wire.ACC
Intended: 'With what did Sara strum the wire at what speed?'

We believe that the discussion in this section supports our assumption that instrumental-marked themes are arguments. However, before we turn our attention to instrumental-marked instruments, we want to address a possible issue, raised by a reviewer, namely the difference in case between accusative-marked themes and instrumental-marked themes, which might call into question our expectation that the two will behave the same on the 'do so' replacement test. We believe that this expectation is warranted based on the fact that the behavior of instrumental-marked themes on the coordination test and the multiply fronted *wh* adjunct test is mirrored not only by accusative-marked themes but also by genitive-marked themes, as we will now illustrate.

Unsurprisingly, accusative-marked themes can be coordinated with accusative-marked themes but not with accusative noun phrases denoting frequency (uncontroversially adjuncts):

(9) a. ACC_{theme} & ACC_{theme}: ✓
Sara svaku večer trza žicu i strunu.
Sara every.ACC evening.ACC strums wire.ACC and string.ACC
'Sara strums/is strumming the wire and the string every evening.'

b. ACC_{theme} & $ACC_{frequency}$: ✗
**Sara trza žicu i svaku večer.*
Sara strums wire.ACC and every.ACC evening.ACC
'*Sara strums/is strumming the wire and every evening.'

On the other hand, an accusative-marked *wh* theme can front along with an adjunct *wh* phrase, including an accusative *wh* phrase denoting frequency:

(10) a. $ACC_{theme,\ wh}$ $adjunct_{wh}$: ✓
Što je kada/kako/zašto Sara trzala?
what.ACC AUX when/how/why Sara strum
'What did Sara strum when/how/why?'

b. $ACC_{theme,\ wh}$ $ACC_{frequency,\ wh}$: ✓
Što je koju večer Sara trzala?
what.ACC AUX which.ACC evening.ACC Sara strum
'What did Sara strum which evening?'

This is expected if one *wh* phrase is an argument and the other an adjunct.

The same behavior is replicated by genitive-marked themes. A genitive-marked theme can be coordinated with another genitive-marked theme but not with a genitive-marked noun phrase denoting frequency:

(11) a. GEN_{theme} & GEN_{theme}: ✓
Sara se svake večeri hvata knjige i olovke.
Sara REFL every.GEN evening.GEN reaches.for book.GEN and pencil.GEN
'Sara reaches for/is reaching for the book and the pencil every evening.'

b. GEN_{theme} & $GEN_{frequency}$: ✗
**Sara se hvata knjige i svake večeri.*
Sara REFL reaches book.GEN and every.GEN evening.GEN
'*Sara reaches for/is reaching for the book and every evening.'

As predicted, genitive-marked *wh* themes can front with *wh* adjuncts, including genitive-marked noun phrases denoting frequency:

(12) a. $GEN_{theme,\ wh}$ $adjunct_{wh}$: ✓
Čega se kada/kako/zašto Sara hvatala?
what.GEN REFL when/how/why Sara reach.for
'What did Sara reach for when/how/why?'

b. $GEN_{theme,\ wh}$ $GEN_{frequency,\ wh}$: ✓
Čega se koje večeri Sara hvatala?
what.GEN REFL which.GEN evening.GEN Sara reach.for
'What did Sara reach for which evening?'

The evidence regarding instrumental themes presented in this section allowed us to make the assumption that instrumental themes are arguments. Also, since instrumental-marked themes and themes bearing other cases (not only accusative but also genitive) behave the same with respect to coordination and multiple *wh* fronting, we believe that we are justified in our expectation that instrumental-marked themes

should pattern with accusative-marked themes on the 'do so' replacement test. Next, we motivate our assumption that instrumental-marked instruments are adjuncts.

2.3 Evidence that instrumental instruments are adjuncts

In comparison to instrumental-marked themes, which are possible only with a relatively small number of verbs in Croatian, instrumental-marked instruments are much freer in their distribution. They can modify almost all dynamic transitive verbs, as long as the event denoted by the verb admits an instrument. Instrumental-marked instruments can be present when the theme is omitted, but they can also always be omitted themselves. This is shown with a couple of examples:

(13) *Marko reže* (*kruh*) (*nožem*).
Marko cuts bread.ACC knife.INS
'Marko cuts/is cutting (bread) (with a knife).'

(14) *Leo ubija* (*svoje žrtve*) (*pištoljem*).
Leo kills self's victims.ACC gun.INS
'Leo kills/is killing (his victims) (with a gun).'

As a reviewer points out, the fact that instrumental-marked instruments are free in their distribution and that they are omissible is weak evidence that they are not arguments; after all, the same holds to a great degree of accusative-marked themes, which we assume to be arguments. However, instrumental-marked instruments do exhibit a number of properties that strengthen the claim that they are adjuncts.

First, we note the sentences in (15), where instrumental-marked instruments are used iteratively.[11] This should not be possible if they are arguments (Bresnan 1982, Pollard & Sag 1987, Forker 2014). Such sentences, if not perfect, are considerably

[11] Instruments can only be used iteratively when they do not both denote instruments of the same kind. For example, (i) is ungrammatical.

(i) **Goran je na koncu nožem izboo napadača bodežom.*
Goran AUX in.the.end knife.INS stabbed attacker.ACC dagger.INS
Intended: 'In the end, Goran stabbed the attacker with a knife with a dagger.'

This requirement holds of iterated adjuncts more generally. Iterated adjuncts cannot all be at the same level of specificity (Brunson 1993). If they are, the contribution of one contradicts the contribution of the other (Verspoor 1997), which leads to ungrammaticality.

better than those in (16), in which instrumental themes are stacked. This suggests that instruments are adjuncts, while themes are arguments.

(15) a. ?*Borna je* ***iglom*** *izvukao trn iz Vidovog oka* ***drhtavim prstima.***
Borna AUX needle.INS took.out thorn.ACC from Vid's eye shaky.INS fingers.INS
Literally: 'Borna took a thorn out of Vid's eye with a needle with his trembling fingers.'

b. ?*Sanja je* ***umornim očima*** *napokon* ***dalekozorom*** *ugledala neprijatelja.*
Sanja AUX tired.INS eyes.INS finally binoculars.INS spotted enemy.ACC
Literally: 'Sanja finally spotted the enemy with binoculars with her tired eyes.'

c. ?*Goran je na koncu* ***nožem*** *izboo napadača* ***izranjavanim rukama.***
Goran AUX in.the.end knife.INS stabbed attacker.ACC wounded.INS hands.INS
Literally: 'In the end, Goran stabbed the attacker with a knife with his wounded hands.'

(16) a. **Borna je na kraju* ***rukom*** *drmao* ***drhtavim prstima.***
Borna AUX on end hand.INS shook shaky.INS fingers.INS
'*In the end, Borna shook/was shaking his hand his trembling fingers.'

b. **Sanja je rastreseno* ***slomljenom nogom*** *ljuljala* ***stopalom.***
Sanja AUX absentmindedly broken.INS leg.INS swung foot.INS
'*Sanja absentmindedly swung/was swinging the broken leg the foot.'

c. **Goran je* ***dugačkim konopcem*** *vitlao* ***lasom.***
Goran AUX long.INS rope.INS swirled lasso.INS
'*Goran swirled/was swirling the long rope the lasso.'

We noted in section 2.2 (see (5b)) that instrumental-marked themes cannot be coordinated with instrumental noun phrases denoting manner (see also footnote 8). Instrumental-marked instruments do not resist such coordination:[12]

[12] Again, these examples might not be perfect, but they are much better than the one in (5b), in which an instrumental theme is coordinated with an instrumental noun phrase denoting manner.

(17) a. ?*Sara je trzala žicu komadićem plastike i velikom brzinom.*
Sara AUX strummed wire.ACC piece.INS plastic.GEN and great.INS speed.INS
'Sara strummed/was strumming the wire with a piece of plastic and at great speed.'

b. ?*Liječnik je pregledao pacijenta stetoskopom i laganim dodirima ruku.*
doctor AUX examined patient stethoscope.INS and light.INS touches.INS hands.GEN
'The doctor examined the patient with a stethoscope and by lightly touching them with their hands.'

Finally, an instrumental-marked *wh* instrument cannot be fronted together with another adjunct *wh* phrase:[13]

(18) $\text{INS}_{\text{instrument, } wh}$ adjunct_{wh}: ✗
a. **Čime je kako / kako je čime Sara trzala žicu?*
what.INS AUX how how AUX what.INS Sara strum wire.ACC
Intended: 'With what did Sara strum the wire how?'

b. **Čime je kada / kada je čime liječnik pregledao pacijenta?*
what.INS AUX when when AUX what.INS doctor examine patient.ACC
Intended: 'With what did the doctor examine the patient when?'

On the other hand, a multiple *wh* question with a *wh* argument (*wh* accusative theme) and an instrumental-marked *wh* phrase denoting an instrument is grammatical, even when both *wh* phrases are fronted:

[13] The examples in (18)—as well as all the comparable examples in section 2.2—become well-formed if one of the adjunct *wh* phrases is not fronted:

(i) a. *Čime je Sara kako trzala žicu?*
what.INS AUX Sara how strum wire.ACC
'With what did Sara strum the wire how?'

b. *Čime je liječnik kada pregledao pacijenta?*
what.INS AUX doctor when examine patient.ACC
'With what did the doctor examine the patient when?'

(19) $\text{INS}_{\text{instrument, } wh}$ $\text{ACC}_{\text{theme, } wh}$: ✓

Što	*je*	*čime*	/ *čime*	*je*	*što*	*Sara*	*trzala?*
what.ACC	AUX	what.INS	what.INS	AUX	what.ACC	Sara	strum

'What did Sara strum with what?'

Instrumental-marked instruments thus behave differently from instrumental-marked themes: they are less restricted in their distribution, can be iterated, and pattern with adjuncts on the coordination and *wh* adjunct fronting tests. Therefore, in our experiments we assumed that instrumental-marked instruments *are* adjuncts.

We thus expected instrumental-marked instruments to behave differently on the 'do so' test from both instrumental-marked themes and accusative-marked themes. On the other hand, we expected instrumental-marked themes to mirror the behavior of accusative-marked themes.

Our experimental items also included sentences involving prepositions that assign the accusative case (*u* 'in,' *na* 'on,' *za* 'for,' and *uz* 'with'). All the prepositional phrases were adjuncts in that they were not selected by the verb and mostly denoted direction (those introduced by *u* 'in' and *na* 'on'); one introduced a beneficiary (*za* 'for'), and one had comitative meaning (*uz* 'with'). We expected these to pattern with instrumental-marked instruments.

The fact that in Croatian, instrumental-marked noun phrases can denote both themes (which we take to be arguments) and instruments (which we take to be adjuncts) provides us with an opportunity to test how successful the 'do so' test is in discriminating between constituents that differ only in their argument/adjunct status while being categorically and prosodically identical (McInnerney 2022b).[14,15]

[14] Ideally, we would have liked to use accusative-marked noun phrases both as arguments and as adjuncts, without resorting to prepositions. However, in Croatian only a very limited number of accusative-marked noun phrases can be adjuncts (e.g., *cijeli tjedan* 'whole week'). At the same time, we were significantly constrained in the construction of the stimuli because of the fact that the number of verbs that can take both accusative-marked themes and instrumental-marked themes is also very limited. We were thus forced to complete the paradigm (instrumental: argument (theme), adjunct (instrument); accusative: argument (theme), **adjunct (direction)**) using prepositional phrases headed by prepositions that assign the accusative case (e.g., *u glavu* 'into head').

[15] A reviewer suggests that all instrumentals in Croatian are prepositional phrases (Milićev & Bešlin 2019), in which case any observed differences between instrumental- and accusative-marked themes might be due to the categorial difference. The reviewer goes on to say that "the alternating acc[usative]-inst[rumental] pattern largely resembles the NP/PP alternations found with similar verbs in English (*pull something* vs. *pull at something*) …" We have no answer to the possibility that all instrumental NPs are actually PPs (if that is the case, then this is precisely the confounding factor we were trying to avoid by testing instrumental-marked NPs). However, we disagree with the reviewer's statement that the accusative–instrumental

In the next section, we discuss the argument–adjunct and lexical explanations of the contrast that the 'do so' replacement test yields. We also state the predictions for our experiments of these two approaches to the 'do so' test.

3 Two approaches to the results of the 'do so' replacement test

3.1 The argument–adjunct explanation of the results of the 'do so' test

According to the argument–adjunct explanation of the results of the 'do so' test, the paradigm in (1) and (2), repeated here as (20) and (21), is explained in terms of the argument versus adjunct status of the stranded constituents (*a pear* in (20b), *today* in (21)).

(20) a. *John ate an apple and Bill did so, too.*
[*did so* = ate an apple]

b. **John ate an apple and Bill did so a pear.*
[*did so* = ate]

(21) *John ate an apple yesterday and Bill did so today.*
[*did so* = ate an apple]

The contrast between the ungrammatical example (20b) and the grammatical (21) suggests that the structure of VP is not flat but hierarchical (Lakoff & Ross 1976, Haegeman 1991, Haegeman 2006, Ackema 2015) and that complements and adjuncts occupy different structural positions in the hierarchy: the complement of the verb

---

alternation in Croatian resembles the conative alternation in English. The hallmark of the latter is the change in the semantics of the event. The change from NP to PP most prominently has a de-resultativizing effect: while (ia) entails that the target of shooting was hit, (ib) does not.

(i) a. *Sam shot the bear.*
b. *Sam shot at the bear.*
(Coon & Preminger 2017: 234)

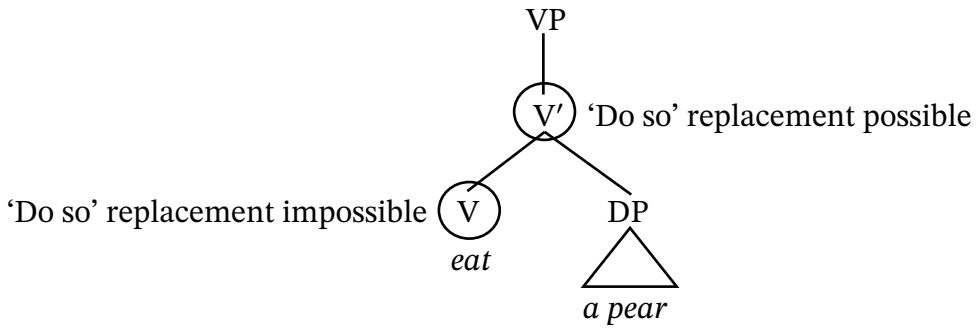
This is not the case for the accusative–instrumental alternation in Croatian; both (iia) and (iib) entail that the key has been turned.

(ii) a. *Branko je zakrenuo ključ.*
Branko aux turned key.ACC
'Branko turned the key.'

b. Branko *je zakrenuo ključem.*
Branko aux turned key.INS
'Branko turned the key.'

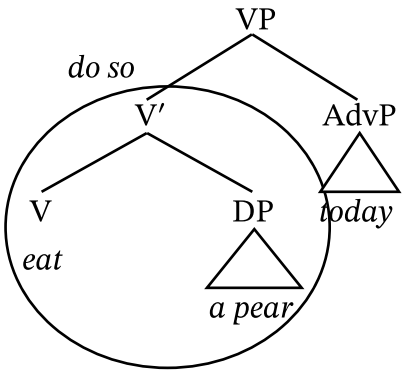
is the sister to the verbal head, while adjuncts occupy structurally more distant positions and are not included in the minimal projection of the verb. The pro-form *do so* must substitute minimally for the immediate projection of the verb, that is, the node labeled V′ in (22). The reason (20b) is ungrammatical is that there is no node in the structure that may be replaced by *do so* to the exclusion of the object *a pear*.

(22)



On the other hand, since the pro-form *do so* can replace a node (V′) that does not include an adjunct *today*, as shown in (23), example (21) is grammatical.

(23)



The results of the 'do so' test remain consistent when the verb appears with an adjunct but without a complement. This is a possibility with intransitive verbs and with optionally transitive verbs used intransitively. In these cases, as in the cases of obligatorily transitive verbs, the pro-form *do so* can strand an adjunct:

(24) a. *John swam yesterday and Bill did so today.*
[*did so* = swam]

b. *John ate yesterday and Bill did so today.*
[*did so* = ate]

While the judgments in (20), (21), and (24) are uncontroversial, there are many examples cited in the literature where 'do so' replacement is possible but should not be. Such examples, discussed in the next two subsections, have led researchers to question the status of the 'do so' replacement as a test for argumenthood and to propose alternative explanations for the observed contrasts. In section 3.2 we turn to issues with the 'do so' test reported in the literature. Then, in section 3.3, we explore one particular alternative to the argument–adjunct explanation of the 'do so' test results, which we are calling the lexical explanation.

3.2 Issues with the 'do so' replacement test

The 'do so' replacement test, as a diagnostic for the structure of the VP, relies on the pro-form 'do so' being a so-called **surface anaphor** (Hankamer & Sag 1976). Surface anaphors are anaphors that acquire meaning by replacing a chunk of structure that is already built. Conversely, **deep anaphors** are inserted into the structure in their own right and do not replace already built syntactic representations; such anaphors are identified with their antecedents through semantics/pragmatics, rather than through syntax.

Criticisms of the 'do so' test as an instrument for discriminating between arguments and adjuncts mostly rest on arguments against the view that 'do so' is a surface anaphor, in other words, against the view that 'do so' is inserted into the structure as a replacement for a node with an internal structure, identical to some linguistically present antecedent.

Surface anaphors require syntactic identity between the replaced VP and its antecedent. However, Kehler & Ward 1999, Przepiórkowski 1999, Culicover & Jackendoff 2005, and Houser 2010, among others, show that some examples of 'do so' anaphora are grammatical in the absence of syntactic identity between the antecedent VP and the VP replaced by 'do so.' Their examples include active–passive mismatches, as in (25a), causative–inchoative mismatches, as in (25b), discontinuous antecedents, as in (25c), and split antecedents, as in (25d).

(25) a. *Mary **was contacted** last night by the same man who had done so before.*
(J. Kaplan 1976: 250)

b. *Mary claimed that I **closed the door,** but it actually did so on its own.*
(Houser 2010: 20)

c. *Robin **slept** for twelve hours **in the bunkbed** and Leslie did so for eight hours.*
(Culicover & Jackendoff 2005: 125)

d. *Fortunately, the first person to **die in 1990** and the first couple to **file for divorce in 1990** were allowed to do so anonymously.*
(Kehler & Ward 1999: 248)

Another argument against 'do so' being a surface anaphor comes from *wh* extraction: such extraction out of the VP replaced by 'do so' should be possible if 'do so' is a surface anaphor (given that the replaced VP has internal structure from which constituents can in principle be extracted). This prediction is, however, not borne out:

(26) **I don't know which puppy you should adopt, but I know which one you shouldn't do so.*
(Houser 2010: 21)

Finally, Miller 1992 shows that 'do' in the 'do so' replacement does not undergo subject–auxiliary inversion:

(27) a. *John kicked Mary.*
b. **Did Peter so too?*
c. *Did Peter do so too?*
(Miller 1992: 96)

This suggests that 'do' in 'do so' is not an auxiliary verb but rather a main verb. This in turn argues against the view that it is a pro-form.

If 'do so' is a deep anaphor and the 'do so' replacement does not in fact involve any "replacement," then any reference to the structure on which 'do so' supposedly operates becomes impossible. This in turn casts doubt on any account that explains the results of the test by appealing to the structural status of the stranded constituent. Instead, the results of the 'do so' test must be explained by appealing to the only element that the elliptical VP contains, namely 'do so' itself. This is what the lexical explanation attempts to do.

3.3 The lexical explanation of the results of the 'do so' test

Miller 1992 argues that the grammaticality or ungrammaticality of 'do so' replacement has nothing to do with the status of the surviving constituent as an argument or an

adjunct. Instead, Miller proposes that whether a constituent can "survive the *do so* substitution" depends on whether the constituent is compatible with the selectional properties of the lexical verb *do* in *do so* in such a way that the thematic role of the surviving constituent (assigned by the lexical *do*) is also compatible with the thematic role of the constituent theta-marked by the antecedent verb.[16] For example, in (28), *do so* is followed by the PP *to many others*, headed by *to*, which is compatible with the lexical requirements of *do*. In addition, the thematic role of the PP *to many others* (the survivor of the *do so* substitution) can be identified as "that which is affected by the action of the verb" (Miller 1992: 97). This is compatible with the theme role of the DP *them all*, the complement of the verb *destroy* in the antecedent VP.

(28) *In any case he had decided to, because an eruption of hardcore, coke-and-speed-headed paranoia could destroy them all. It had done so to many others.*
(McInnerney 2022a: 6)

On the other hand, the example in (29) is ungrammatical because "main verb *do* never assigns a locative role to PP[*to*], and only a locative role is compatible with the role of the PP[*to*] after *go*" (Miller 1992: 97).

(29) ??*John went to Paris and Peter did so to Rome.*

The same reason underlies the ungrammaticality of (30): the lexical *do* never assigns a locative role to the following PP, while that is the role assigned to its antecedent by the verb *load* (Miller 1992: 97, fn. 22).

(30) **John loaded a sack onto the truck, and I did so onto the wagon.*
(Lakoff & Ross 1976: 106)

By treating *do* as a lexical verb with its own argument structure, the lexical explanation gives an elegant account of the results that the 'do so' test yields in examples where the syntactic identity between the antecedent VP and the 'do so' VP does not hold. However, this alone does not explain why it is impossible for *do so* to be followed by a VP-internal argument (like *a pear* in our example (20b)). To account for this, Miller 1992 proposes (97, fn. 22) that the subcategorization frame of the verb *do*, by stipulation, cannot accommodate both an accusative-marked argument

[16] This view is adopted by McInnerney 2022a and 2022b.

and the adverb *so*.[17] The fact that, on this explanation, the adverb *so* figures in the argument structure of *do* suggests that constituents that can follow *do so* must be compatible with *do*, regardless of whether they would traditionally be classified as arguments or adjuncts. We thus assume that the requirement of compatibility of the surviving constituent and the lexical verb *do* holds not only of possible arguments of *do* but also of adjuncts. This assumption readily accounts for the results of the 'do so' test reported by Kweon 2002, repeated here:

(31) a. *Tom placed the book on the table. *And Sue did so on the chair.* (Kweon 2002: 12)
b. *Bill bathed the dog in the kitchen. And Martha did so in the bathroom.*

The PP *on the chair* in (31a) is compatible with the verb *do* but only if it denotes location, which is different from the denotation of the PP *on the table* in the antecedent VP, which denotes direction. Since *do* cannot accommodate direction, (31a) is ungrammatical. The PP *in the bathroom*, on the other hand, denotes location and is thus compatible with the verb *do,* which results in the grammaticality of (31b).[18]

Now that we have presented the reasoning behind the way in which the argument–adjunct explanation and the lexical explanation account for the results of the 'do so' test, we turn to the predictions that each makes for our Croatian data.

3.4 Predictions for Croatian

The Croatian version of the pro-form that we used in the test was *to činiti* 'do it/that.'[19] This choice was based on several considerations. Our first concern was grammatical aspect. The verbs we used in the experimental items are all aspectually inherently imperfective and need a prefix to become perfective (e.g., *vrtjeti* 'spin.IPFV' ~ ***za****vrtjeti* 'spin.PFV'). Thus, we needed our pro-form to be imperfective too, which ruled out the

[17] The presence of *it* in *do it* has the same effect (Miller 1992: 97, fn. 22).

[18] A reviewer asks how the lexical explanation accounts for the Korean speakers' insensitivity to the directional interpretations of PPs in their L2 English. We believe that it does not. Kweon cites this insensitivity as an indication that the Korean L2 speakers of English had not acquired the argument structure of verbs that require a direction PP (*put*) and did not distinguish them from those that optionally take a location PP (*bathe*). It is then conceivable that these speakers had not acquired the argument structure of the verb *do* either, which would be compatible with their judgments. However, if L2 speakers' grammar contains non-native information about argument structure of relevant verbs, neither the argument–adjunct explanation nor the lexical explanation can be expected to account for the results.

[19] The pronominal element *to* 'it/that' in *to činiti* is analyzed by Progovac 1998 as an event pronominal.

verb *učiniti* 'do.PFV' and its more informal counterpart *napraviti* 'do.PFV.' Next, we wanted to keep the experimental items in the present tense, so as to avoid complications in the word order introduced by second-position auxiliary clitics, which necessarily accompany verbal participles in the past tense. Thus, the pro-form we settled on had to sound natural in the third person singular of the present tense. We judged that the pro-form *to činiti* 'do it/do that' (3SG: *to čini*) was more natural than the combination činiti isto 'do the same' (3SG: čini isto), proposed by Brač 2018. Finally, we chose not to use the more casual and less formal verb *raditi* 'do, work, make' because it is semantically richer than the more formal činiti 'do.'[20]

As stated before, we applied the 'do so' test to sentences built with verbs that allow their theme NP arguments to appear either in the accusative case or in the instrumental case. In section 2 we presented evidence for treating instrumental-marked themes as arguments and instrumental-marked instruments as adjuncts. In our study we were interested to see whether the 'do so' test would treat instrumental-marked themes on a par with accusative-marked themes (both arguments) and treat both differently from instrumental-marked instruments. Since accusative NPs do not typically figure as adjuncts, to complete the paradigm (as mentioned in footnote 14) we made use of PPs headed by prepositions that assign the accusative case to their complements, typically with the meaning of direction. We thus created the following paradigm of experimental items, yielding four experimental conditions.

(32)

	Condition
a. V + NP_{INS} (theme)	INS complement
b. V + NP_{ACC} (theme)	ACC complement
c. V + NP_{INS} (instrument)	INS adjunct
d. V + PP ($P + NP_{ACC}$)	ACC adjunct

Our expectations were as follows. If the 'do so' test discriminates between arguments and adjuncts (that is, if the argument–adjunct analysis is correct), as initially proposed by Lakoff & Ross 1976, then both instrumental-marked and accusative-marked themes should yield unacceptable results on the 'do so' test, while instrumental-marked instruments and direction-denoting PPs should both yield acceptable results.

[20] Przepiórkowski 1999 settles on *zrobić to* 'do it' as the closest Polish equivalent of English *do so*. Przepiórkowski opts for the verb *zrobić* 'do' over *uczynić* 'do' because of the perceived formality of the later.

Miller 1992's lexical explanation predicts that it should not be possible to strand instrumental-marked themes in the 'do so' test but that it should be possible to strand instrumental-marked instruments. This is because the verb činiti 'do' can only be followed by an instrumental-marked NP if the NP denotes an instrument or means:

(33) *Jan sve što čini, čini glavom/rukama.*
Jan everything that does does head.INS/hands.INS
✓ 'Whatever Jan does, he does it by using his head/his hands.'
'Whatever Jan does, he does it to his head/his hands.'

Since an instrumental-marked NP following *to činiti* 'do so' cannot bear a theme theta role (borne by the instrumental-marked NP in the antecedent), the outcome of the 'do so' substitution in the condition (32a) is predicted to be ungrammatical. By contrast, the outcome in the condition (32c), where the instrumental-marked NP denotes an instrument, is predicted to be grammatical.

Thus, the argument–adjunct explanation and the lexical explanation make the same predictions regarding instrumental-marked NPs. They also make the same predictions when it comes to accusative-marked NPs: the argument–adjunct explanation predicts these not to be strandable by 'do so' because they are arguments, and the lexical explanation predicts them not to be strandable because the main verb 'do,' by stipulation, cannot accommodate both an accusative-marked theme argument and the adverb 'so,' or more pertinently to our study, the pronoun 'it' (the Croatian version of which—*to*—figures in the pro-form *to činiti*).

The two analyses make different predictions, however, when it comes to the outcome of 'do so' replacement in the PP condition: the argument–adjunct explanation predicts the PPs to be acceptable after 'do so' (because they are adjuncts), but the lexical explanation predicts them to be unacceptable. This is because a direction-denoting PP cannot follow lexical činiti 'do' in Croatian:

(34) **Jan sve što čini, čini u glavu / u ruke.*
Jan everything that does does in head.INS in hands.ACC
Intended: 'Whatever Jan does, he does it into his head/his hands.'

The following is a summary of the predictions of the two explanations.

(35)		Condition	Argument–adjunct	Lexical
a.	V + NP_{INS} (theme)	INS complement	×	×
b.	V + NP_{ACC} (theme)	ACC complement	×	×
c.	V + NP_{INS} (instrument)	INS adjunct	✓	✓
d.	V + PP (P + NP_{ACC})	ACC adjunct	✓	×

A reviewer notes that our experiments would have provided a better testing ground for the two explanations if the predictions made by the two explanations differed with respect to the instrumental-marked themes, which, to the best of our knowledge, have not yet been investigated via the 'do so' test. While the reviewer certainly has a point, we believe that the diverging predictions in the PP condition still do tease the two explanations apart: since the presence of the adverb 'so' figures prominently in the lexical explanation's account of the inability of 'do so' to be followed by an accusative-marked theme, it follows that on this explanation the results of the 'do so' test depend on the semantic compatibility of 'do' not only with argument phrases that follow it but also with adjunct phrases. Thus, the fact that the two explanations make different predictions for direction-denoting PPs in our experiments can still make a contribution to assessing their performance.

We next turn to the description of our experiments.

4 The current study

We conducted two experiments: an acceptability judgment task and a self-paced reading task. The stimuli we constructed were identical for both tasks with one exception; we begin by describing them in detail.

4.1 Stimuli

In both experiments, the critical materials consisted of 64 sentences in Croatian, in a 2 × 2 design (type of constituent × case).[21] Each critical sentence involved coordination of clausal conjuncts in which the second conjunct contained the sequence *to čini* 'does so.' In the argument conditions, *to čini* 'does so' was followed by a noun, accusative-marked or instrumental-marked. In the adjunct conditions, *to*

[21] All stimuli, data, and R scripts used for analysis can be found in the Open Science Framework repository https://doi.org/10.17605/OSF.IO/YJB2C.

Type	Case	Experimental sentence	Gloss and translation
Adjunct	Accusative	*Elena tuče u rame, a Borna to čini u glavu usput vrišteći.*	Elena hits in shoulder.ACC and Borna that does in head.ACC while screaming 'Elena is hitting in the shoulder, and Borna is doing so in the head while screaming.'
Adjunct	Instrumental	*Elena tuče ramenom, a Borna to čini glavom usput vrišteći.*	Elena hits shoulder.INS and Borna that does head.INS while screaming 'Elena is hitting with the shoulder, and Borna is doing so with the head while screaming.'
Complement	Accusative	*Elena mrda rame, a Borna to čini glavu usput vrišteći.*	Elena moves shoulder.ACC and Borna that does head.ACC while screaming '*Borna is moving the shoulder, and Borna is doing so the head while screaming.'
Complement	Instrumental	*Elena mrda ramenom, a Borna to čini glavom usput vrišteći.*	Elena moves shoulder.INS and Borna that does head.INS while screaming '*Borna is moving the shoulder, and Borna is doing so the head while screaming.'
Filler		*Vojnici zarađuju puškom, a novinari perom usput putujući.*	Soldiers earn rifle.INS and journalists pen.INS while traveling 'Soldiers earn with a rifle, and journalists with a pen while traveling.'
Filler		*Mladen sluša radio, a Vid rješava križaljke usput čavrljajući.*	Mladen listens radio.ACC and Vid solves crossword.puzzles.ACC while chatting 'Mladen is listening to the radio, and Vid is solving crossword puzzles while chatting.'

Table 1: Example of a critical set and fillers. The *usput* 'while, at the same time' phrase was omitted in the acceptability judgment task.

čini 'does so' was followed either by an instrumental-marked noun or by a PP with an accusative-marked noun as the complement of the preposition. The critical region in all the experimental sentences was the (accusative- or instrumental-marked) noun. In the acceptability judgment task, the nouns were sentence-final. In the self-paced reading task, we added two additional words after the final noun so as to be able to measure the spillover effect after the critical part of the sentence and to avoid possible wrap-up effects on the critical region. The first of the two additional words was the same across all experimental items (*usput* 'while, at the same time').

In accord with the 2 × 2 design, the 64 critical stimuli were made up of 16 sets of four similar sentences differing critically in terms of the case and constituent type of the noun following *to čini* 'does so': accusative adjunct (PP),[22] instrumental adjunct (NP), accusative complement (NP), or instrumental complement (NP). See table 1 for an example of a critical set. One verb was used for the two adjunct sentences while another verb was used for the two complement sentences. For the complement sentences, we used 16 Croatian verbs belonging to a very restricted class of verbs whose theme can be either accusative or instrumental. All the verbs in the adjunct sentences either were intransitive (e.g., *bučiti* 'make noise') or could optionally take a theme and still be naturally followed by an adjunct phrase (e.g., *ubadati* 'pierce, sting'; the optional theme was omitted in our stimuli). The noun following the verb was lexically the same across the four sentences, but in the accusative adjunct condition, the noun was embedded in a prepositional phrase headed by a preposition that assigns the accusative case, such as *u* 'in,' *na* 'on,' or *za* 'for.'[23] In the analysis of the self-paced reading task, we analyzed reaction times in the noun region following the *to činiti* 'do so' phrase and in the spillover region, which contained the adverb *usput* 'while, at the same time.'

The stimuli were distributed into four lists using a Latin square design, so each participant only saw one of the four sentences from each set; thus, each participant saw 16 critical sentences (four sentences in each of the four type × case combinations). In each list, there were also 24 filler sentences, involving two clausal conjuncts in which

[22] For ease of exposition, we will refer to these adjunct constituents as accusative even though they actually involved an accusative noun embedded in a prepositional phrase.

[23] We acknowledge that using different prepositions is a limitation of the design. Due to the complexities of the process of creating the stimuli, unfortunately, it was not possible to use only one type of preposition in the accusative adjunct condition. In this case, we operated under the assumption that it was more important to keep the noun and the verb the same.

the second conjunct was either non-elliptical or contained a gapped structure (the verb was missing, but its complement was present). The fillers' structure resembled the structure of the critical sentences, so in the self-paced reading task they also contained the *usput* phrase.

While creating the stimuli, we asked for L1 speaker feedback via a small norming study in which we asked the participants about the acceptability of the tested verb–noun combinations (e.g., *trzati žicu/žicom* 'pull wire.ACC/wire.INS'). Each participant in the norming study saw a single verb only once across four lists (with either an accusative or instrumental noun). Each expression was seen by an average of 10 people. Based on the acceptability ratings, we revised several phrases and incorporated these revisions in the final stimuli.

4.2 Study 1: acceptability judgments

4.2.1 Participants

We recruited 120 L1 speakers of Croatian, aged between 18 and 30. The participants were all college students at a state university in Croatia. The study was approved via the ethical board of the faculty where the experiment took place.

4.2.2 Design and procedure

The stimuli used in this study were described in section 4.1 and are the same as the stimuli in the self-paced reading study, excluding the last two words of each sentence. Sentences within the four lists were randomized, and lists were randomly assigned to participants. There were 16 critical sentences and 24 fillers on each list. Since our assumption was that the four accusative complement sentences on each list were going to be judged ungrammatical and the rest grammatical, two thirds of the fillers (16 sentences) were ungrammatical as well, so that there were overall the same number of grammatical (20) and grammatical (20) sentences on each list.

The survey was presented using PsyToolkit (Stoet 2010, 2017); it took participants on average five minutes to complete the survey. The participants completed the survey either on their mobile phones or on desktop computers in a classroom at theUniversity of Osijek, Croatia, in the Faculty of Humanities and Social Sciences. For the acceptability ratings, we used a 1–5 acceptability Likert scale (1 = completely

unacceptable, 5 = completely acceptable). Following Schütze & Sprouse 2013, we instructed the participants that considerations such as how probable it is to actually encounter a sentence in real life should not affect their acceptability judgment of it, and we also offered an anchor sentence as an example of what is acceptable and what is not, to reduce variability in how the participants judged the sentences.

4.2.3 Results and analysis

Four participants were excluded from the analysis. We excluded three participants due to short finishing times (less than three minutes, which was deemed insufficient to supply meaningful judgments in the task). In addition to this, another participant was excluded because they were not following the instructions of the task; this was evident in their responses to filler sentences (their mean for acceptable sentences was 1.25, while the mean for acceptable sentences of the entire group was 4.54), and similar behavior was noticed for critical sentences (mean 1.31). We analyzed the data from the remaining 116 participants.

As suggested in Schütze & Sprouse 2013, the data were analyzed using a linear mixed effects model with acceptability ratings as the dependent variable; the software used was the Lme4 package, version 1.1-29 (Bates et al. 2015), in R, version 4.0.4 (R Core Team 2021). The fixed effects were type (complement/adjunct) and case (accusative/instrumental), with random effects of participant and item. We created a base model, which included an intercept and the two random factors (subject, item). In the second step, we then added a fixed effect of type. In the third step, we added a fixed effect of case, and in the fourth step, we added an interaction between type and case. The details of the four models can be found in table 2. These models were compared (stepwise, each model being compared to the next simplest one) to see which one was the best fit. Before the analysis, the ratings were *z* score–transformed (Schütze & Sprouse 2013).

The best fit model is the fourth one. Both the main effect of type and case are significant, and so is the interaction between type and case, as can be seen in table 2, model 3. To explore the significant interaction further, we ran a post hoc pairwise comparison test (Tukey method) using the Emmeans package, version 1.7.1-1 (Lenth 2021). The contrast between the instrumental adjuncts and complements

Fixed effects	Estimate	Standard error	Degrees of freedom	*t* value	Pr(> |*t*|)
Model 0					
(Intercept)	0.007	0.09	63.99	0.09	.92
Model 1					
Compared to model 0: $X^2(1) = 4.75$, $p = .02$	0.16	0.11	83.26	1.48	.14
Type = complement	−0.31	0.14	130.09	−2.19	.03
Model 2					
Compared to model 1: $X^2(1) = 72.93$, $p < .001$*					
(Intercept)	−0.40	0.07	64.56	−5.81	<.001
Type = complement	−0.24	0.08	73.00	−3.03	.003
Case = INS	1.06	0.08	73.00	13.44	<.001
Model 3 (best fit)					
Compared to model 2: $X^2(1) = 31.38$, $p < .001$*					
(Intercept)	−0.22	0.07	66.06	−3.30	.002
Type = complement	−0.66	0.09	66.27	−7.06	<.001
Case = INS	0.71	0.09	74.81	7.79	<.001
Type = complement : case = INS	0.80	0.13	70.21	6.11	<.001

Table 2: Comparison of linear mixed effects models for the acceptability rating task (fixed effect: rating, *z* score–transformed).

was not significant ($t(1,82.2) = -1.44$, $p = .15$). The contrast between the accusative adjuncts and complements was significant, with complements being judged as less acceptable than adjuncts ($t(1, 72.9) = 6.83$, $p < .0001$). See also figure 1 for details on raw rating means. The contrast between the accusative and instrumental adjuncts was significant ($t(1, 82.2) = -7.54$, $p < .0001$), with the latter being judged as more acceptable. Finally, the contrast between the accusative and instrumental complements was also significant ($t(1, 72.9) = -15.55$, $p < .0001$), with the latter being judged as more acceptable.

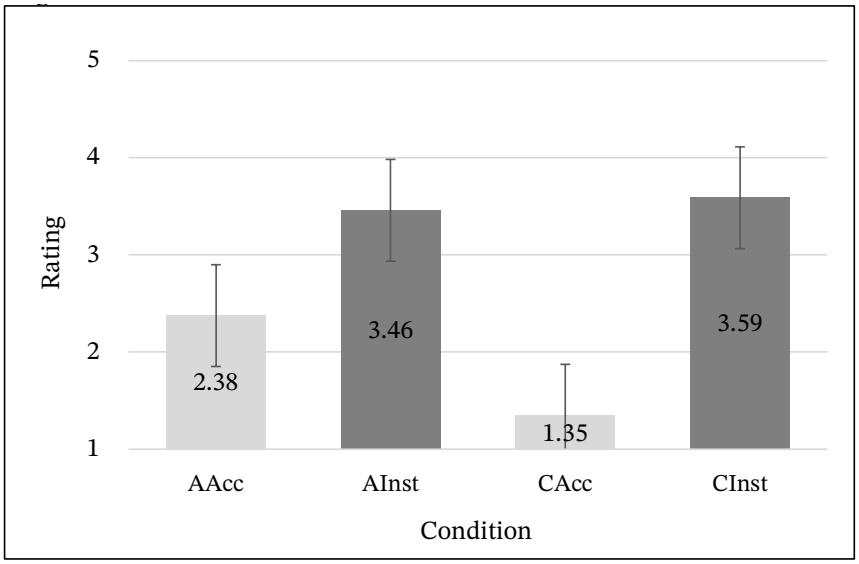


Figure 1: Raw rating values by condition. AAcc = accusative adjunct, AInst = instrumental adjunct, CAcc = accusative complement, CInst = instrumental complement.

To sum up, the results show that the accusative complements received the lowest ratings in the judgment task and that they are significantly less acceptable after 'do so' than both instrumental complements and accusative adjuncts. Crucially, the instrumental complement ratings received the highest ratings, these being comparable to the instrumental adjunct ratings.

4.3 Study 2: self-paced reading task

4.3.1 Participants

We recruited 48 L1 speakers of Croatian, aged between 18 and 30, with normal vision or corrected-to-normal vision and no known language/neurological/hearing disorder. All the participants were college students at a state university in Croatia. Their participation was rewarded with course credit. The study was approved by the ethical board of the faculty where the experiment took place. No participant in the acceptability judgment task was recruited for the self-paced reading task.

4.3.2 Design and procedure

The stimuli used in this study were the same as those used in the acceptability judgment study, described in section 4.1. The only difference in the self-paced reading task was the inclusion of the last two words in each sentence due to possible spillover effects.

The self-paced reading experiment was conducted in a computer lab at the University of Osijek, Croatia, in the Faculty of Humanities and Social Sciences. For the presentation of the experiment, we used PsyToolkit (Stoet 2010, 2017); the sentences were displayed on a desktop computer. Each participant saw 16 critical sentences and 24 filler sentences. The experiment lasted about 10 minutes, depending on each participant's speed. They used a keyboard to progress through each sentence at their own pace before answering the comprehension question. The mode of presentation was incremental: an additional word would appear every time the participant pressed the right arrow button, while the previous words remained on the screen. Once the sentence was completed, the next key press removed it from the screen and presented a statement about the sentence that the participant had just read. Participants had to answer whether this follow-up statement was true or false. (For example, after the sentence shown in table 3, 'Sara is pulling the wire, and Rita is doing so the tape while cursing,' the follow-up statement might have been 'Rita is pulling the wire,' which is false.) Each list had the same number of true and false follow-up statements, and the distribution of true and false follow-up statements across experimental conditions was equalized (for the four sentences presented for each condition, two of the follow-up statements were true and two were false).

4.3.3 Results and analysis

We analyzed the self-paced reading times. In both argument conditions as well as in the instrumental adjunct condition, there were 10 sentence regions; see the example in table 3. In the accusative adjunct condition, there were 12 regions, due to the presence of prepositions in both conjuncts. In addition, there was a yes–no answer region in each item. We analyzed how long participants took to read the critical region (the noun after 'do so,' region 8 in table 3) and the spillover region (*usput* 'while/at the same time,' region 9 in table 3). Regions before and after the critical and spillover regions were not analyzed. All fillers were excluded from the analysis. For both of the relevant regions, we excluded data points below 200 ms and above 3,000 ms. It was deemed that the words with very low reaction times were skipped rather than read, while the very high reaction times potentially represented times where participants were inattentive to the task. We also excluded three participants who performed below 50% accuracy on the true–false task. This relatively low threshold was set because the

sentences were quite complex and there was a lot of new information, so it may have been rather difficult to answer the question about the follow-up statement. In total, these two data scrubbing methods resulted in 9.98% data loss for the second noun region and 8.78% in the spillover region.

1	**2**	**3**	**4**	**5**	**6**	**7**	**8 (critical noun region)**	**9 (spillover)**	**10**
Sara	*trza*	*žicu,*	*a*	*Rita*	*to*	*čini*	*traku*	*usput*	*psujući.*
Sara	pulls	wire.ACC	and	Rita	that	does	tape.ACC	while	cursing
‘Sara is pulling the wire, and Rita is doing so the tape while cursing.’									

Table 3: Example of region division (accusative complement condition).

The data were analyzed using a linear mixed effects model for reaction times; again the software used was the Lme4 package, version 1.1-29 (Bates et al. 2015), in R, version 4.0.4 (R Core Team 2021). The fixed effects were case and type of constituent, with random effects of participant and item. For each of the two relevant regions, the critical region and the spillover region, we created a base model, which included an intercept and the two random factors (subject, item). In the second step, we added a fixed effect of type. In the third step, we added a fixed effect of case, and in the fourth step we added an interaction between type and case. These models are shown in table 4. For each region, the four models were compared (stepwise, each model being compared to the next simplest one) to see which one was the best fit. We thus report two best fit models, one for each region. In figures 2 and 3 we report the raw mean reading times for the critical region and the spillover region.

As can be seen in table 4, for the noun region following the ‘do so’ phrase, the best fit model was the one where both type and case were added but no interaction between them was added. In this region, it takes longer to read the instrumental case, while the fixed effect of type does not affect the dependent variable.

For the spillover region (which contained the adverb *usput* ‘while, at the same time’), the best fit model is the last one, with an interaction between the two fixed effects of type and case. To explore the significant interaction further, we ran a post hoc pairwise comparison test (Tukey method) using the Emmeans package,

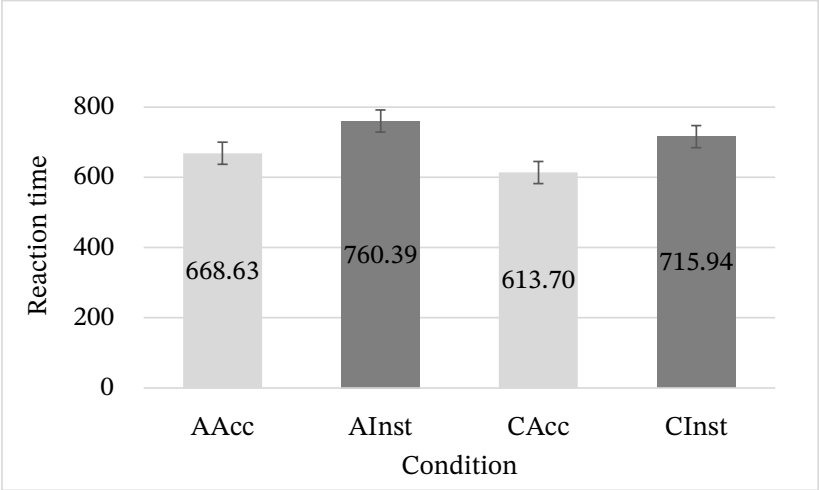


Figure 2: Raw mean reading times in the critical noun region by condition. AAcc = accusative adjunct, AInst = instrumental adjunct, CAcc = accusative complement, CInst = instrumental complement.

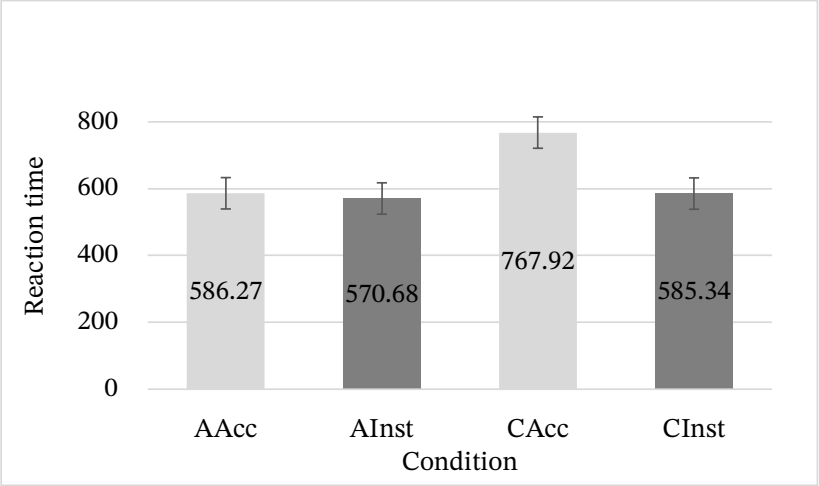


Figure 3: Raw mean reading times in the spillover region by condition. AAcc = accusative adjunct, AInst = instrumental adjunct, CAcc = accusative complement, CInst = instrumental complement.

version 1.7.1-1 (Lenth 2021). The contrast between the instrumental adjuncts and complements was not significant ($t(1, 636) = -0.37$, $p = .71$). The contrast between the accusative adjuncts and complements was significant, with complements taking longer to read ($t(1, 634) = -4.16$, $p < .0001$). See also figures 2 and 3 for details on reaction times. The contrast between the accusative and instrumental adjuncts was not significant ($t(1, 114) = 0.37$, $p = .71$). However, the contrast between the accusative and instrumental complements was significant ($t(1, 112) = 4.14$, $p = .0001$), with instrumental complements read faster than accusative complements.

Region	Fixed effects	Estimate	Standard error	Degrees of freedom	*t* value	Pr(> |*t*|)
Noun	Model 0					
	(Intercept)	689.00	35.90	42.61	19.2	<.001
	Model 1					
	Compared to model 0: $X^2(1) = 1.9$, $p = .17$	715.06	40.62	69.95	17.60	<.001
	Type = complement	−52.04	37.71	616.66	−1.38	.17
	Model 2 (best fit)					
	Compared to model 1: $X^2(1) = 5.51$, $p = .02$*					
	(Intercept)	666.29	44.27	72.91	15.05	<.001
	Type = complement	−52.44	37.70	617.00	−1.39	.16
	Case = INS	98.56	40.11	30.97	2.46	.02
	Model 3					
	Compared to model 2: $X^2(1) = 0.04$, $p = .84$					
	(Intercept)	669.94	48.03	100.61	13.95	<.001
	Type = complement	−59.79	53.16	618.62	−1.12	.26
	Case = INS	91.14	55.13	106.13	1.65	.10
	Type = complement : case = INS	14.80	75.48	618.62	0.20	.84
Spillover	Model 0					
	(Intercept)	627.36	24.79	36.91	25.31	<.001
	Model 1					
	Compared to model 0: $X^2(1) = 10.07$, $p = .002$					
	(Intercept)	578.28	29.32	72.47	19.72	<.001
	Type = complement	98.53	30.91	628.15	3.19	.002
	Model 2					
	Compared to model 1: $X^2(1) = 9.22$, $p = .002$					
	(Intercept)	627.36	32.10	144.01	19.54	<.001
	Type = complement	98.61	30.85	657.68	3.20	.001
	Case = INS	−98.04	30.85	657.08	−3.18	.002

(Contd.)

Region	Fixed effects	Estimate	Standard error	Degrees of freedom	*t* value	Pr(> |*t*|)
	Model 3 (best fit) Compared to model 2: $X^2(1) = 7.16$, $p = .007$					
	(Intercept)	586.57	35.42	203.99	16.56	<.001
	Type = complement	180.95	43.41	657.76	4.17	<.001
	Case = INS	−15.94	43.34	657.32	−0.37	.71
	Type = complement : case = INS	−164.64	61.38	657.55	−2.68	.007

Table 4: Comparison of linear mixed effects models for the self-paced reading task.

5 Discussion

The results of the acceptability judgment study show that there is a significant interaction between constituent type (argument/adjunct) and case (accusative/instrumental). The difference between the accusative complement and accusative adjunct sentences was significant, with the adjunct sentences being judged as more acceptable. The accusative complement sentences were also significantly different from the instrumental complement sentences, with the latter being judged as more acceptable. Finally, instrumental adjuncts were also judged as significantly more acceptable following *to činiti* 'do so' than accusative adjuncts. Crucially, however, there was no difference between instrumental arguments and instrumental adjuncts.

In the self-paced reading task, in the critical region (the noun after the 'do so' phrase), there was a significant main effect of case: instrumental nouns were read slower. This finding is most likely due to the fact that all instrumentals are longer than accusatives. In other words, the significant main effect may not be caused by case itself but rather by a variable that correlates with case, namely length. On the other hand, in this region there was no interaction between constituent type and case, which indicates that it was a word's case—which was conflated with its length—rather than its status as an argument or an adjunct that primarily influenced how people processed it. To avoid the confounding effect of length, we examined the spillover region, the region following the critical region. (The spillover region consisted of the adverb *usput* 'while, at the same time' in all conditions.) A significant interaction was found

in the spillover region. In this region, there were significantly longer reading times for accusative complement sentences than for either accusative adjunct sentences or instrumental complement sentences. Importantly for our purposes, instrumental phrases bearing the theme theta role, which we hypothesized were complements and would behave like accusative themes, received unexpectedly high ratings following the pro-form *to činiti* 'do so' in the acceptability judgment task, where they were judged as significantly more acceptable than accusative complements; consistently with this result, there was no significant difference between instrumental complements (instrument phrases denoting themes) and instrumental adjuncts (instrumental phrases denoting instruments). We obtained comparable results in the self-paced reading task: instrumental complements were read equally fast as instrumental adjuncts and significantly faster than accusative complements.

These results are unexpected for both the argument–adjunct explanation and the lexical explanation of the 'do so' test contrasts. The results refute the argument–adjunct explanation because, according to this explanation, the contrasts in grammaticality dovetail with the structural status of the constituent (argument vs. adjunct). Our results show that the 'do so' test fails to deliver the predicted ungrammatical results with instrumental-marked themes. The fact that arguments and adjuncts were distinguished in the accusative conditions can be explained by appealing to the difference in their categories: accusative arguments were NPs, while accusative adjuncts were PPs. Thus, the 'do so' replacement test revealed itself to be "[in]sensitive to properties which cross-cut categorial distinctions," which, according to McInnerney 2022a: 2, is a necessary condition for a diagnostic that successfully distinguishes between arguments and adjuncts. We thus conclude that the contrasts we obtained in our experiments cannot be explained by appealing to the argument versus adjunct status of the stranded constituent. This casts doubt on the 'do so' replacement as a reliable test for the argument–adjunct distinction.

The results we obtained in our experiments regarding the instrumental-marked NPs (themes and instruments) also argue against the lexical explanation. Recall that according to the lexical explanation, grammatical or ungrammatical results on the 'do so' test depend on (i) whether the stranded constituent is compatible with the lexical verb 'do' and (ii) whether the theta role of the stranded constituent is compatible with the theta role of its antecedent. This explanation, just like the argument–adjunct

explanation, predicted that the 'do so' test would yield ungrammatical results with stranded instrumental-marked themes. The reason behind this prediction is that the lexical verb činiti 'do' in Croatian can be followed by an instrumental-marked NP only if the NP denotes an instrument/means, not if it is the theme, as was shown in section 3.4 (see (34)). Consequently, items where *to činiti* 'do so' was followed by an instrumental NP were expected to be acceptable only when the antecedent denoted an instrument (i.e., in the adjunct condition), not when it denoted the theme (in the argument condition). Thus, the grammaticality of the items in which the stranded constituent was an instrumental-marked theme is surprising not only on the argument–adjunct approach to the 'do so' test but also on the lexical approach. The results we obtained in the self-paced reading task in the PP condition were also not predicted by the lexical explanation: since the lexical verb činiti 'do' does not tolerate PPs denoting direction (present in almost 90% of our experimental items), participants should have experienced processing difficulties whenever such a PP followed *to* činiti 'do so.' Our results, however, do not indicate that this was the case. Thus, the lexical explanation seems to have failed to predict our results in both the instrumental argument condition and the accusative adjunct condition.

Limited support for the lexical explanation is provided by the results of the acceptability judgment task: there, we found that instrumental-marked instrument NPs received significantly higher ratings than PPs containing accusative-marked NPs, as predicted by the lexical explanation. However, as stated above, the degradation disappeared in the self-paced reading task, where direction-denoting PPs were read equally fast as instrumental-marked instruments (and both significantly faster than the accusative-marked themes).

A discrepancy like the one we found with the adjunct PPs, which received relatively low ratings on the acceptability judgment task but did not elicit longer reading times in the self-paced reading task, is often found between offline and online techniques, with online tasks showing facilitation compared to offline difficulty. This discrepancy has been found in a number of studies in which authors collected both offline and online data and is not specific to the 'do so' replacement test. For example, De-Dios-Flores 2019 investigated whether L1 English speakers recognize that sentences featuring multiple negation (e.g., *The bills that no senators voted for have never become the law*) are grammatical, as opposed to sentences featuring double negation (e.g., **No authors that the critics recommended have never received acknowledgment for a*

best-selling novel). De-Dios-Flores found that speakers found grammatical multiple negation sentences degraded in all three experiments she conducted: a speeded acceptability judgment task, a self-paced reading task, and an offline acceptability rating task. However, multiple negation sentences received unexpectedly low ratings in the untimed offline task, with the rate of acceptance much lower than in the speeded acceptability judgment task, even though in the offline task, given ample time, participants could access a fully encoded final stage representation of the sentences. Similarly, Paolazzi et al. 2016 and 2017 tested for the complexity of passive sentences compared to active sentences with adult L1 English speakers with both an offline accuracy task and a self-paced reading task. The authors found that passives appeared to be more difficult than actives in offline measurements (lower accuracy, slowed decision making time) but not in the self-paced reading task, where passives were actually read faster than actives in several regions, starting from the verb of a passive sentence.[24] Such diverging results in online and offline tasks have been explained by appealing to the idea that language is interpreted via two different (although related) cognitive systems, described in Parker 2019 as a "system that contains the mental machinery for fast and efficient communication, traditionally referred to as the *parser*, and a slower backup system that defines the precise rules of the language and classifies grammaticality, traditionally referred to as the *grammar*" (2). Given that the parser is responsible for quick and efficient communication, it posits "good enough representations" (Ferreira et al. 2002, Ferreira & Patson 2007, Karimi & Ferreira 2016), which sometimes contain errors, while grammar is responsible for a finer-grained process that results in more detailed evaluations.

We suspect that the somewhat degraded status of accusative adjuncts in the offline task may have been a consequence of non-linguistic considerations related to the meaning that arose once the PP adjunct combined with the verb. All of our verb–PP adjunct combinations were grammatically acceptable (as revealed by the norming study); however, since in developing the stimuli we were heavily constrained by the inventory of verbs that can take both accusative- and instrument-marked themes, some of the combinations gave rise to meanings that were rather unusual. An example is given in (36), where the objects denoted by the accusative nouns (*ključ* 'key,' *nož* 'knife') are more likely to be used as instruments in an engraving event than as the surface for engraving.

[24] Ferreira 2003 was the first study that reported offline complexity of passives and object clefts compared to actives and subject clefts.

(36) *Lucija urezuje na ključ, a Lidija to čini na nož.*
Lucija engraves on key and Lidija that does on knife
'Lucija is engraving on a key, and Lidija is doing so on the knife.'

This kind of unexpected reading did not arise with instrumental adjuncts. The "strangeness" of PP adjuncts relative to instrumental adjuncts in our study may have been overlooked in the fast, online processing but did show up (to an extent) in the offline ratings.

Before we conclude, we would like to add a note regarding the status of instruments. The 'do so' test typically classifies instruments as adjuncts not only in English (Lakoff & Ross 1976, Hoffmann 2007) but also in Polish, a Slavic language (Przepiórkowski 1999). Przepiórkowski 1999 reports that instrumental-marked NPs denoting instruments can be stranded on the Polish equivalent of the 'do so' test:

(37) *Janek wbił gwóźdź młotkiem, a Tomek zrobił to siekierą.*
John drove nail hammer.INS and Tom did it axe.INS
'John drove the nail with a hammer, and Tom did so with an axe.'
(Przepiórkowski 1999: 313)

On the other hand, Polish instrumental-marked NPs denoting means are ungrammatical on the 'do so' test:

(38) **Janek przesłał zaproszenie pocztą, a Tomek zrobił to e-mailem.*
John sent invitation post.INS and Tom did it email.INS
Intended: 'John sent an invitation by post, and Tom did so by email.'
(Przepiórkowski 1999: 314)

Przepiórkowski cites this inconsistency of the 'do so' test results to argue against 'do so' being a test for argumenthood.[25] We did not test any instrumental NPs denoting means, so we cannot claim with certainty what results we would obtain, but to our ear, examples like (39) are grammatical in Croatian.

[25] Przepiórkowski shows that the 'do so' test also yields inconsistent results with goal clauses, in that it treats infinitival clauses as complements but clauses introduced by the complementizer *žeby* as adjuncts, and with benefactives and malefactives, in that it treats the former as complements and the latter as adjuncts.

(39) *Jan šalje pozivnice poštom, a Vid to čini e-mailom.*
Jan sends invitations post.INS and Vid it does email.INS
'Jan sends invitations by post, and Vid does so by email.'

Przepiórkowski also states that classification of instruments as adjuncts on the Polish version of the 'do so' test "is in accordance with the functional and syntactic-functional criteria [...], which are the most common criteria for the complement/adjunct distinction, but against the iterability test" (313). In section 2.3 we showed that instrumental-marked instruments in Croatian can be iterated, at least when they are not at the same level of specificity (a requirement that holds of iterating adjuncts more generally). It is possible, then, that cross-linguistically, instruments display different syntactic behavior. If the 'do so' test is sensitive to syntactic properties of stranded phrases, it might yield different results for different languages, depending on the syntactic behavior of the relevant constituents. We also showed that in Croatian, *wh* instruments do not front with other adjunct *wh* phrases. Thus, we have good evidence that, syntactically at least, instruments in Croatian are adjuncts. This means that in our experiments, the 'do so' test fared well on both adjunct conditions. The problem is that it admitted instrumental-marked themes, which we showed behave syntactically as arguments. It seems to us that it is possible that the 'do so' test in Croatian (and possibly more generally) makes only a one-directional prediction about the constituents that can or cannot be stranded on the test: the prediction is that any constituent that is *unacceptable* when stranded after 'do so' must be an argument. If the constituent is *acceptable* after 'do so,' then it may be either an argument or an adjunct. In other words, no adjunct should yield an ungrammatical result on the 'do so' test, as long as it more generally displays the syntactic behavior of an adjunct. Our results are compatible with such one-directional sensitivity of the 'do so' test, but the possibility should be investigated both theoretically and experimentally. We leave this for further research.

6 Conclusion

In this study we investigated whether the 'do so' replacement can be used as a test for distinguishing arguments from adjuncts, by applying it to Croatian instrumental-marked theme NPs and instrumental-marked instrument NPs. We found that the 'do

so' test does not discriminate between the two kinds of instrumental-marked NPs and yields grammatical results with both. We assumed that instrumental-marked themes are arguments while instrumental-marked instruments are adjuncts, based on a number of diagnostics with respect to which the two kinds of instrumental NPs exhibit different behavior. Under this assumption, it appears that the 'do so' test does not always yield ungrammatical results when the stranded constituent is an argument. This suggests that the 'do so' test is not a reliable diagnostic for distinguishing arguments from adjuncts. However, in our experiments, the test performed well in both adjunct conditions: it did not yield an ungrammatical result in either one. We thus speculated that the test more generally only classifies constituents that cannot be stranded by 'do so' as arguments, saying nothing about those that can be.

We were further interested in whether our results might support an alternative explanation of the contrasts yielded by the 'do so' replacement test proposed by Miller 1990 and adopted by Przepiórkowski 1999 and McInnerney 2022a and 2022b, which we called the lexical explanation. Again, we reached a negative result: the acceptability of instrumental-marked themes following *to činiti* 'do so' was not predicted on this analysis either. Similarly, on the lexical explanation PP adjuncts following 'do so' were expected to elicit longer reading times in the self-paced reading task, contrary to fact. The lexical explanation did predict the degradation of PP adjuncts containing an accusative-marked NP in the acceptability judgment task, but since that degradation was not found in the self-paced reading task, we concluded that the factors responsible for the diverging results of our offline and online experiments are those more generally found in experimental paradigms, where offline tasks often reveal lower acceptability ratings than online tasks.

To sum up, in our experiments the 'do so' test did not yield expected results. Furthermore, neither of the approaches to the 'do so' test that we considered predicted the results we obtained. While this is far from an ideal outcome, we do believe that our study revealed the need for experimental treatment not only of the 'do so' replacement test but also of other diagnostics for the argument–adjunct distinction, in order to arrive at a reliable set of tests (if such a set exists) for discriminating between arguments and adjuncts both within a single language and cross-linguistically.

Data-availability statement

The data that support the findings of this study are openly available in the Open Science Framework repository https://doi.org/10.17605/OSF.IO/YJB2C.

Funding information

This project was supported by two projects funded by the Croatian Science Foundation: Syntactic and Semantic Analysis of Arguments and Adjuncts in Croatian (IP-2019-04-7896) and Semantic-Syntactic Classification of Croatian Verbs (IP-2022-10-8074).

Acknowledgments

We wish to thank our participants for their contribution and our colleagues Goran Faletar, Branimir Belaj, Ana Mikić Čolić, and Goran Milić for their assistance with data collection.

Competing interests

We have no competing interests to declare.

Author contributions

The study was conceived and designed by AWH, MGY, and MB. The data were collected and analyzed by AWH. The results were discussed and interpreted by AWH, MGY, and MB. Finally, the article was drafted and edited by AWH, MGY, and MB. The order of authorship is not intended to indicate the relative importance of contributions.

References

Ackema, Peter. 2015. Arguments and adjuncts. In: Kiss, Tibor & Alexiadou, Artemis (editors). *Syntax—theory and analysis: an international handbook*. De Gruyter Mouton. 1.246–273. https://doi.org/10.1515/9783110377408.246.

Akal, Taylin. 2017. Wh-argument/adjunct asymmetry in sentence processing. *Dilbilim Araştırmaları Dergisi* 2.43–71.

Andreu, Llorenç & Sanz-Torrent, Mònica & Rodríguez-Ferreio, Javier. 2016. Do children with SLI use verbs to predict arguments and adjuncts: evidence from eye movements during listening. *Frontiers in Psychology* 6.1917. https://doi.org/10.3389/fpsyg.2015.01917.

Barbu, Roxana-Maria. 2020. *On the psycholinguistics of argumenthood*. Doctoral thesis. Carleton University.

Barbu, Roxana-Maria & Toivonen, Ida. 2016. Event participants and linguistic arguments. In: Papafragou, Anna & Grodner, Daniel J. & Mirman, Daniel & Trueswell, John (editors). *Proceedings of the 38th annual meeting of the Cognitive Sciences Society*. Swarthmore College. 1961–1966.

Bates, Douglas & Mächler, Martin & Bolker, Ben & Walker, Steve. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67.1–48. https://doi.org/10.18637/jss.v067.i01.

Baxter, David Paul. 1999. *English goal infinitives*. Doctoral thesis. University of Illinois at Urbana-Champaign.

Belaj, Branimir & Tanacković Faletar, Goran. 2017. *Imenska sintagma i sintaksa padeža* [The noun phrase and the syntax of cases]. Volume 1 of *Kognitivna gramatika hrvatskoga jezika*. Disput.

Boland, Julie E. 2005. Visual arguments. *Cognition* 95.3.237–274. https://doi.org/10.1016/j.cognition.2004.01.008.

Boland, Julie E. & Blodgett, Allison. 2006. Argument status and PP-attachment. *Journal of Psycholinguistic Research* 35.5.385–403. https://doi.org/10.1007/s10936-006-9021-z.

Borsley, Robert D. 1991. *Syntactic theory: a unified approach*. Edward Arnold.

Bošković, Željko. 1998. Wh-movement and wh-phrases in Slavic. Unpublished paper. Presented at the Comparative Slavic Morphosyntax Workshop, 5–7 June.

Bošković, Željko. 2010. Wh-movement and wh-phrases in Slavic. *Glossos* 10. https://slaviccenters.duke.edu/projects/glossos-journal/issues/issue-10.

Brač, Ivana. 2018. *Instrumental u hrvatskom jeziku* [The instrumental in the Croatian language]. Institute for the Croatian Language and Linguistics.

Bresnan, Joan. 1982. Polyadicity. In: Bresnan, Joan (editor). *The mental representation of grammatical relations*. MIT Press. 149–172.

Bresnan, Joan. 1994. Locative inversion and the architecture of universal grammar. *Language* 70.1.72–131.

Browne, Wayles. 1974. On the problem of enclitic placement in Serbo-Croatian. In: Brecht, Richard D. & Chvany, Catherine V. (editors). *Slavic transformational syntax*. University of Michigan. 36–52.

Brunson, Barbara. 1993. The instrumental role: argument or adjunct? *Toronto Working Papers in Linguistics* 12.1.13–25.

Cavar, Damir. 1999. *Aspects of the syntax-phonology interface: cliticization and related phenomena in Croatian*. Doctoral thesis. University of Potsdam.

Chomsky, Noam. 1981. *Lectures on government and binding*. Foris Publications.

Chromý, Jan & Vojvodić, Sonja. 2024. When and where did it happen? Systematic differences in recall of core and optional information. *Quarterly Journal of Experimental Psychology*. 77.1.111–132. https://doi.org/10.1177/17470218231159190.

Coon, Jessica & Preminger, Omer. 2017. Split ergativity is not about ergativity. In: Coon, Jessica & Massam, Diane & Travis, Lisa (editors). *The Oxford handbook of ergativity*. Oxford University Press. 226–252. https://doi.org/10.1093/oxfordhb/9780198739371.013.10.

Croft, William. 1991. *Syntactic categories and grammatical relations: the cognitive organization of information*. University of Chicago Press.

Culicover, Peter & Jackendoff, Ray. 2005. *Simpler syntax*. Oxford University Press.

DeArmond, Richard & Hedberg, Nancy. 1998. On complements and adjuncts. In: *Proceedings of the 1998 annual conference of the Canadian Linguistics Association*. Cahiers Linguistiques d'Ottawa. https://www.researchgate.net/publication/246998195_On_Complements_and_Adjuncts.

de-Dios-Flores, Iria. 2019. Processing sentences with multiple negations: grammatical structures that are perceived as unacceptable. *Frontiers in Psychology* 10.2346. https://doi.org/10.3389/fpsyg.2019.02346.

Dowty, David R. 2003. The dual analysis of adjuncts/complements in Categorial Grammar. In: Lang, Ewald & Maienborn, Claudia & Fabricius-Hansen, Cathrine (editors). *Modifying adjuncts*. Mouton de Gruyter. 33–66. https://doi.org/10.1515/9783110894646.33.

Ferreira, Fernanda. 2003. The misinterpretation of noncanonical sentences. *Cognitive Psychology* 47.2.164–203. https://doi.org/10.1016/s0010-0285(03)00005-7.

Ferreira, Fernanda & Bailey, Karl G. D. & Ferraro, Vittoria. 2002. Good-enough representations in language comprehension. *Current Directions in Psychological Science* 11.1.11–15. https://doi.org/10.1111/1467-8721.00158.

Ferreira, Fernanda & Patson, Nikole D. 2007. The "good enough" approach to language comprehension. *Language and Linguistics Compass* 1.1–2.71–83. https://doi.org/10.1111/j.1749-818x.2007.00007.x.

Forker, Diana. 2014. A canonical approach to the argument/adjunct distinction. *Linguistic Discovery* 12.2.27–40. https://doi.org/10.1349/PS1.1537-0852.A.444.

Gazdar, Gerald & Pullum, Geoffrey K. & Sag, Ivan. 1982. Auxiliaries and related phenomena in a restrictive theory of grammar. *Language* 58.3.591–638. https://doi.org/10.2307/413850.

Grimshaw, Jane. 1990. *Argument structure*. MIT Press.

Haegeman, Liliane. 1991. *Introduction to Government and Binding theory* (first edition). Blackwell Publishing.

Haegeman, Liliane. 1994. *Introduction to Government and Binding theory* (second edition). Blackwell Publishing.

Haegeman, Liliane. 2006. *Thinking syntactically: a guide to argumentation and analysis*. Blackwell Publishing.

Halpern, Aaron. 1992. *Topics in the placement and morphology of clitics*. Doctoral thesis. Stanford University.

Hankamer, Jorge & Sag, Ivan. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7.3.391–428.

Hoffmann, Thomas. 2007. Complements versus adjuncts? A Construction Grammar account of English prepositional phrases. *Occasional Papers in Language and Linguistics* 3.92–119.

Houser, Michael. 2010. *The syntax and semantics of* do so *anaphora*. Doctoral thesis. University of California, Berkeley.

Kaplan, Jeffrey P. 1976. *Pro-form reference to verb phrases in English*. Doctoral thesis. University of Pennsylvania.

Kaplan, Ronald M. & Bresnan, Joan. 1982. Lexical-functional grammar: a formal system for grammatical representation. In: Bresnan, Joan (editor). *The mental representation of grammatical relations*. MIT Press. 173–281.

Karimi, Hossein & Ferreira, Fernanda. 2016. Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology* 69.5.1013–1040. https://doi.org/10.1080/17470218.2015.1053951.

Katičić, Radoslav. 2002. *Sintaksa hrvatskoga književnog jezika* [The syntax of the Croatian literary language]. Croatian Academy of Sciences and Arts.

Kehler, Andrew & Ward, Gregory. 1999. On the semantics and pragmatics of "identifier so." In: Turner, Ken (editor). *The semantics/pragmatics interface from different points of view*. Elsevier. 233–256.

Kennison, Sheila M. 2002. Comprehending noun phrase arguments and adjuncts. *Journal of Psycholinguistic Research* 31.1.65–81. https://doi.org/10.1023/a:1014328321363.

Koenig, Jean-Pierre & Mauner, Gail & Bienvenue, Breton. 2003. Arguments for adjuncts. *Cognition* 89.2.67–103. https://doi.org/10.1016/S0010-0277(03)00082-9.

Koenig, Jean Pierre & Mauner, Gail & Bienvenue, Breton & Conklin, Kathy. 2008. What with? The anatomy of a (proto)-role. *Journal of Semantics* 25.2.175–220. https://doi.org/10.1093/jos/ffm013.

Kweon, Soo-Ok. 2002. Locative PPs in L2 English argument structure acquisition. *Korean Journal of English Language and Linguistics* 2.1–21.

Lakoff, George & Ross, John Robert. 1976. Why you can't *do so* into the sink. In: McCawley, James D. (editor). *Notes from the linguistic underground*. (*Syntax and Semantics* 7.) Academic Press. 101–131. https://doi.org/10.1163/9789004368859_008.

Lee, Jiyeon & Thompson, Cynthia K. 2011. Real-time production of unergative and unaccusative sentences in normal and agrammatic speakers: an eyetracking study. *Aphasiology* 25.6–7.813–825. https://doi.org/10.1080/02687038.2010.542563.

Lenth, Russell. 2021. Emmeans: estimated marginal means, aka least-squares means (version 1.6.0). Software (R package). https://doi.org/10.32614/CRAN.package.emmeans.

Manning, Christopher D. 2003. Probabilistic syntax. In: Bod, Rens & Hay, Jennifer & Jannedy, Stefanie (editors). *Probabilistic linguistics*. MIT Press. 289–314.

McInnerney, Andrew. 2022a. Against the argument/adjunct distinction. *University of Pennsylvania Working Papers in Linguistics* 28.1.13.

McInnerney, Andrew. 2022b. *The argument/adjunct distinction and the structure of prepositional phrases*. Doctoral thesis. University of Michigan.

Milićev, Tanja & Bešlin, Maša. 2019. Instrumental: why it is absent from the clitic system in Serbian/Croatian. In: Lopičić, Vesna & Mišić Ilić, Biljana (editors). *Jezik, književnost, teorija*. Faculty of Philosophy, University of Niš. 153–168.

Miller, Philip H. 1990. Pseudogapping and *do so* substitution. In: Ziolkowski, Michael & Noske, Manuela & Deaton, Karen (editors). *Papers from the 26th regional meeting of the Chicago Linguistic Society*. Chicago Linguistic Society. 1.293–305.

Miller, Philip H. 1992. *Clitics and constituents in phrase structure grammar*. Garland Publishing.

Needham, Stephanie & Toivonen, Ida. 2011. Derived arguments. In: Butt, Miriam & King, Tracy Holloway (editors). *Proceedings of the LFG11 conference*. CSLI Publications. 401–442.

Paolazzi, Caterina L. & Grillo, Nino & Alexiadou, Artemis & Santi, Andrea. 2016. Processing English passives: interaction with event structure, but no evidence for heuristics. Poster. Presented at the 29th CUNY Conference on Human Sentence Processing, 3–5 March.

Paolazzi, Caterina L. & Grillo, Nino & Alexiadou, Artemis & Santi, Andrea. 2017. Passives are not always more difficult than actives. Poster. Presented at the 23rd Architectures and Mechanisms for Language Processing conference, 7–9 September.

Parker, Dan. 2019. Two minds are not always better than one: modeling evidence for a single sentence analyzer. *Glossa* 4.64. https://doi.org/10.5334/gjgl.766.

Pollard, Carl & Sag, Ivan A. 1987. *Fundamentals*. Volume 1 of *Information-based syntax and semantics*. Center for the Study of Language and Information.

Progovac, Ljiljana. 1996. Clitics in Serbian/Croatian: Comp as the second position. In: Halpern, Aaron L. & Zwicky, Arnold M. (editors). *Approaching second: second position clitics and related phenomena*. CSLI Publications. 411–428.

Progovac, Ljiljana. 1998. Event pronominal *to*. *Journal of Slavic Linguistics* 6.1.3–39.

Przepiórkowski, Adam. 1999. *Case assignment and the complement/adjunct dichotomy: a non-configurational constraint-based approach*. Doctoral thesis. University of Tübingen.

Radford, Andrew. 1988. *Transformational grammar: a first course*. Cambridge University Press.

R Core Team. 2021. R: a language and environment for statistical computing. Software. R Foundation for Statistical Computing. https://www.R-project.org.

Rissman, Lilia. 2010. Instrumental *with*, locatum *with*, and the argument/adjunct distinction. *LSA Meeting Extended Abstracts* 1.23. https://doi.org/10.3765/exabs.v0i0.502.

Rissman, Lilia. 2011. Instrumental *with* and *use*: modality and implicature. In: Ashton, Neil & Chereches, Anca & Lutz, David (editors). *Proceedings of Semantics and Linguistic Theory 21*. Linguistic Society of America. 532–551.

Rissman, Lilia & Rawlins, Kyle & Landau, Barbara. 2015. Using instruments to understand argument structure: evidence for gradient representation. *Cognition* 142.266–290. https://doi.org/10.1016/j.cognition.2015.05.015.

Schlesinger, Izchak M. 1995. *Cognitive space and linguistic case*. Cambridge University Press.

Schütze, Carson T. 1995. PP attachment and argumenthood. In: Schütze, Carson T. & Ganger, Jennifer B. & Broihier, Kevin (editors). *Papers on language processing and acquisition*. (*MIT Working Papers in Linguistics* 26.) Department of Linguistics, Massachusetts Institute of Technology. 95–151.

Schütze, Carson T. & Sprouse, Jon. 2013. Judgment data. In: Podesva, Robert J. & Sharma, Devyani (editors). *Research methods in linguistics*. Cambridge University Press. 27–50.

Speer, Shari R. & Clifton, Charles, Jr. 1998. Plausibility and argument structure in sentence comprehension. *Memory and Cognition* 26.5.965–978. https://doi.org/10.3758/BF03201177.

Stoet, Gijsbert. 2010. PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods* 42.1096–1104. https://doi.org/10.3758/BRM.42.4.1096.

Stoet, Gijsbert. 2017. PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology* 44.1.24–31. https://doi.org/10.1177/0098628316677643.

Tesnière, Lucien. 1959. *Éléments de syntaxe structural* [Elements of structural syntax]. Libraire C. Klincksieck. (Published in English in 2015 by John Benjamins Publishing Company.)

Tutunjian, Damon & Boland, Julie E. 2008. Do we need a distinction between arguments and adjuncts? Evidence from psycholinguistic studies of comprehension. *Language and Linguistics Compass* 2.4.631–646. https://doi.org/10.1111/j.1749-818X.2008.00071.x.

Van Valin, Robert, Jr. 2001. *An introduction to syntax*. Cambridge University Press.

Van Valin, Robert, Jr. & LaPolla, Randy. 1997. *Syntax: structure, meaning and function*. Cambridge University Press.

Vater, Heinz. 1978. On the possibility of distinguishing between complements and adjuncts. In: Werner, Abraham (editor). *Valence, semantic case, and grammatical relations*. John Benjamins Publishing Company. 21–45.

Verspoor, Cornelia M. 1997. *Contextually-dependent lexical semantics*. Doctoral thesis. University of Edinburgh.

Williams, Edwin. 1981. Argument structure and morphology. *Linguistic Review* 1.1.81–114. https://doi.org/10.1515/tlir.1981.1.1.81.

Zagona, Karen. 1988. *Verb phrase syntax: a parametric study of English and Spanish*. (*Studies in Natural Language and Linguistic Theory* 13.) Kluwer Academic Publishers.

Zwicky, Arnold M. 1970. Auxiliary reduction in English. *Linguistic Inquiry* 1.3.323–336.