

UC Santa Cruz

UC Santa Cruz Electronic Theses and Dissertations

Title

The processing of direct discourse: When a subordinate speech act sticks around

Permalink

<https://escholarship.org/uc/item/9127m21n>

Author

Duff, John

Publication Date

2020

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike License, available at <https://creativecommons.org/licenses/by-nc-sa/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

**THE PROCESSING OF DIRECT DISCOURSE: WHEN A
SUBORDINATE SPEECH ACT STICKS AROUND**

A thesis submitted in partial satisfaction of the
requirements for the degree of

MASTER OF ARTS

in

LINGUISTICS

by

John Duff

September 2023

The Thesis of John Duff
is approved:

Professor Amanda Rysling, Chair

Professor Pranav Anand

Professor Jess H. K. Law

Peter Biehl
Vice Provost and Dean of Graduate Studies

Copyright © by

John Duff

2023

Table of Contents

List of Figures	iv
List of Tables	v
Abstract	vi
Acknowledgments	vii
1 Introduction	1
2 The processing of appositives	2
2.1 Relevant descriptive generalizations	3
2.2 Evidence for separation in processing	7
3 Direct discourse: The same and different	21
3.1 The basics of reported speech	21
3.2 Comparing relevant features of appositives and direct discourse	22
3.3 The processing of direct discourse: Perceptual simulation . .	25
3.4 Predictions	26
4 Experiment 1: Costs for complexity within reported speech	26
4.1 Method	27
4.2 Results	29
4.3 Discussion	30
5 Experiment 2: Comparing with appositives	31
5.1 Method	32
5.2 Results	33
5.3 Post-hoc comparison with Experiment 2	34
5.4 Discussion	36
6 Experiment 3: Dependencies across reported speech	36
6.1 Method	37
6.2 Results	38
6.3 Discussion	40
7 General discussion	42
7.1 Revisiting the comparability of DD and ARCs	42
7.2 On the prosodic boundaries of DD	45
7.3 Conclusions	47

List of Figures

1	Mean naturalness ratings by condition from Dillon et al. (2014), Expt. 3	10
2	Mean naturalness ratings by condition from Dillon et al. (2017), Expt. 1	13
3	Mean naturalness ratings by condition from Kroll and Wagers, Expt. 2	18
4	Mean naturalness ratings by condition from Expt. 1	29
5	Mean naturalness ratings by condition from Expt. 2	33
6	Mean naturalness ratings by condition from Expt. 3	39

List of Tables

1	Sample stimulus (excerpted) from Dillon et al. (2014), Expt. 3 . . .	9
2	Mean naturalness ratings by condition from Dillon et al. (2014), Expt. 3	9
3	Sample stimulus (excerpted) from Dillon et al. (2017), Expt. 1 . . .	12
4	Mean naturalness ratings by condition from Dillon et al. (2017), Expt. 1	13
5	Sample stimulus from Kroll and Wagers (2019), Expt. 2	17
6	Mean naturalness ratings by condition from Kroll and Wagers (2019), Expt. 2	17
7	Sample stimulus from Expt. 1	27
8	Mean naturalness ratings by condition for Expt. 1	29
9	Ordinal mixed-effects model fit to the data from Expt. 1	30
10	Sample stimulus from Expt. 2	32
11	Mean naturalness ratings by condition for Expt. 2	33
12	Ordinal mixed-effects model fit to the data from Expt. 2	34
13	Bayesian ordinal mixed-effects model fit over the data for Expts. 1-2	35
14	Sample stimulus from Expt. 3	38
15	Mean naturalness ratings by condition for Expt. 3	39
16	Bayesian ordinal mixed-effects model fit to the data from Expt. 3 . .	40

Abstract

The processing of direct discourse: When a subordinate speech act sticks
around

John Duff

The human processing of natural language occasionally reflects systematicies of the system's formal organization, often at the level of syntax. Data on the processing of appositive relative clauses (Dillon et al., 2014; Dillon et al., 2017) suggest that the processor may also be sensitive to the pragmatic organization of information within a sentence: independent speech acts embedded within a maximal speech act appear to be isolated from later parsing and decision making. In three experiments, I examine the processing of another structure which may create an embedded speech act, direct discourse speech reports. Despite existing evidence suggesting direct discourse is indeed processed independently from its superordinate context, the experiments reported here find no support for the prediction that it is isolated from downstream processes. Taking this as evidence against the strong hypothesis that subordinate speech acts serve as hard boundaries in processing, I outline possible alternative hypotheses.

Acknowledgments

The text of this thesis was originally prepared and defended as a qualifying paper during the course of my Ph.D. studies, in March 2020. The same experiments are presented with revised analyses, together with several follow-up experiments and improved discussion, in the article “Pragmatic representations and online comprehension: Lessons from direct discourse and causal adjuncts,” published in *Glossa Psycholinguistics* in 2023, credited to John Duff, Pranav Anand, Adrian Brasoveanu, and Amanda Rysling. Please read and cite the latter version.

1 Introduction

Over the past half-century, a central line of psycholinguistic research has seen the establishment of relationships between stored linguistic objects as one of the primary tasks of online comprehension. Today, this includes modern approaches which make precise quantitative predictions using explicit cognitive models (Lewis & Vasishth, 2005, e.g.). In studying the creation of these dependencies, psycholinguists have often sought to understand how the processor uses categories from the formal structure of a given language in selecting which of multiple hypotheses to entertain. We observe that obligatory syntactic configurations associated with certain dependencies are indeed observed by the parser. When constructing WH-dependencies, gap sites in correlation with a previously encountered filler are not projected within islands, syntactic constituents into which such dependencies cannot be formed (Stowe, 1986; Traxler & Pickering, 1996). When constructing anaphoric relationships upon encountering a reflexive, antecedents in ungrammatical positions within the structure are not considered (Dillon et al., 2013). The processor is able to exploit certain grammatical structure to efficiently construct a parse.

Recent findings presented in Dillon et al. (2014, 2017) suggest that the parser is not limited to syntactic information in its use of grammatical structure. From continued investigation of the processing of appositive relative clauses (ARCs), the authors suggest that when a sub-sentential syntactic unit constitutes an independent, subordinate speech act, it is less influential on later processing than when the same unit merely contributes to the superordinate speech act. This would suggest

an unprecedented role of discourse structure in online processing.

In this thesis, I seek to test the generalizability of this broad proposal by examining the comprehension of another construction which may host an independent, subordinate speech act, the direct discourse speech report. The status of direct discourse reports as distinct units in online processing is well-established (Stites et al., 2013, e.g.), yet in three experiments I demonstrate that it lacks both of the behavioral signatures observed for ARCs in Dillon et al. (2014, 2017). I use this evidence to reject the strong hypotheses made by Dillon and colleagues about the role of discourse structure in processing, and offer two alternative proposals to be investigated in future research.

The roadmap for this thesis is as follows: in Section 1, I will describe and provide background for the important findings of the recent work on appositives. In Section 2, I will compare and contrast direct discourse with the use of appositives, making the case that on some relevant dimensions it should resemble the findings recapitulated in Section 1. In Sections 3 through 5 I present three experiments demonstrating that these predictions are not met, and in Section 6 I consider the implications for this finding.

2 The processing of appositives

In order to motivate the experiments reported here, in this section I will provide a synopsis of the findings in the recent literature regarding the status of ARCs in online sentence processing. I begin first by presenting relevant background

on the nature of ARCs in English, the language of study in this entire examination.

2.1 Relevant descriptive generalizations

Linguists distinguish at least two major syntactic/semantic categories in the class of relative clauses (RCs) in English, exemplified in (1) and (2). RESTRICTIVE RELATIVE CLAUSES (RRCs) provide specifying information which restricts the domain established by their head noun alone (1) , while ARCs do not¹, serving only to elaborate on an already identifiable entity (2). In this way, they function identically to nominal parentheticals, also called nominal appositives (3).

- (1) **Restrictive RC:** Did the nurses who were hired in July receive a housing stipend that month?
- (2) **Appositive RC:** Did the nurses, who were hired in July, receive a housing stipend that month?
- (3) **Nominal Appositive:** Did the nurses, our latest hires, receive a housing stipend?

In the modern pragmatics literature, ARCs and nominal appositives have been argued to comprise a particularly marked way of contributing information to an ongoing discourse. Potts (2005) notes that these constructions, which he groups under the label of “supplements,” can be excised completely from an utterance while leaving the core, “at-issue,” meaning of the sentence intact. They are thus often referred to as “not-at-issue” (NAI) content.

¹Hence, their other name: non-restrictive relative clauses.

Perhaps the most widely foregrounded property of ARCs used to categorize them as NAI is their “nondeniability,” which is to say that discourses like (4) where an interlocutor attempts to reject the information contributed by an ARC are unacceptable (Potts, 2005; Tonhauser, 2012).

(4) A: The geologist, who is a fearless leader, started the descent.

B: # No (, she’s not a fearless leader).

Tonhauser (2012) also notes that information conveyed by an ARC does not felicitously answer a question in the same way as that information conveyed outside of an ARC. Example (5) is adapted from one of her examples in Guaran’i.

(5) A: Who stole my money?

a. B: That man, who lives in Paraguay, stole your money.

b. B: # That man, who stole your money, lives in Paraguay.

The content of ARCs is also projective, which is to say their contributions are not subject to scopal operators like negation or questionhood (Potts, 2005). In (6), clausal negation does not affect whether the speaker paid more than \$500 for the seat, just as in (7) the difference between declarative and interrogative mood does not affect the speaker’s commitment to the same fact. Note also that in the latter case, the ARC clearly falls outside of the alternatives generated by the question, such that an interlocutor may respond to the question positively independently of agreeing with the contribution of the ARC.

(6) a. I found out yesterday that the airline guaranteed my seat, which I paid more than \$500 for, on the flight to Chicago.

- b. I found out yesterday that the airline didn't guarantee my seat, which I paid more than \$500 for, on the flight to Chicago.
- (7) a. A: The airline guaranteed my seat, which I paid more than \$500 for, on the flight to Chicago.
- b. A: Did the airline guarantee my seat, which I paid more than \$500 for, on the flight to Chicago?
- B: Yes (, they did guarantee the seat), but you actually only paid \$300, remember?

Note that all of these patterns contrast with the conversational profile of RRCs, which can be denied (8), can felicitously serve to resolve a question under discussion (QUD) (9)², and are subject to scopal operators like questionhood 10.

- (8) A: The geologist who is a fearless leader started the descent.
- B: No, she's not a fearless leader.
- (9) A: Who stole my money?
- a. B: The man who lives in Paraguay stole your money.
- b. B: The man who stole your money lives in Paraguay.

- (10) a. A: The airline guaranteed a seat which is near the aisle for on the flight

²Note that information in neither an ARC nor an RRC can answer its corresponding polar question. It does seem like RRCs can answer some similar polar questions when their head noun serves as a witness to an existential.

- (1) A: Is the geologist a fearless leader?
B: # The geologist, who is a fearless leader, started the descent. / # The geologist who is a fearless leader started the descent.
- (2) A: Are there any geologists who are fearless leaders?
B: There's a geologist who is a fearless leader right over there.

to Chicago.

- b. A: Did the airline guarantee a seat which is near the aisle on the flight to Chicago?

B: # Yes (, they did guarantee a seat), but it's not near the aisle.

These properties have been captured in formal accounts like Potts (2005) by relegating ARC content to a secondary field of not-at-issue meaning conveyed by a given utterance. Other researchers have argued against a fully peripheral contribution for ARCs, because their separation from discourse phenomena is not complete: it is perfectly acceptable to introduce a discourse referent or satisfy a presupposition from within an ARC—for instance, observe that the referent for *that month* is just as easily established within an ARC as an RRC in (1-2). This leads those authors to posit ARCs as predicating the same common ground as normal speech acts, but doing so through a different set of discourse effects so as to still capture independence from QUD management (AnderBois et al., 2015; Murray, 2014).

A final and important quirk of ARCs is that there do appear to be certain utterance-final cases where ARCs can be denied and can felicitously resolve a question (AnderBois et al., 2015; Koev, 2013). Final ARCs seem optimal as answers to conjoined questions.

- (11) A: I met the geologist, who is a fearless leader.

B: No, she's not.

- (12) A: Who did you meet, and what were they like?

B: I met the geologist, who is a fearless leader.

Nevertheless, the scopal properties of such cases remain unchanged: they still appear to project beyond negation and interrogativity. For instance, in (13), the speaker does not appear to be questioning whether the reporter published the conversation.

- (13) Did the Secretary of State aggressively question that reporter, who published the conversation?

The syntactic analysis of ARCs also remains a contentious question. Some researchers follow the evidence of scopelessness, among other data, to accounts where appositives are high syntactic coordinates to their containing clauses (McCawley, 1988; Ross, 1967), while others suggest that they are integrated subordinates within the DP, minimally different from RRCs (de Vries, 2006; Jackendoff, 1977; Potts, 2005). While Dillon et al. (2018) provide some judgement data relevant to this debate, I do not address it further here.

An aside: We might accept the reality of an at-issue/not-at-issue division without accepting that the above phenomena (nondeniability, limited interaction with QUDs, projectivity) are all equally related to that division. Moreover, there are nuances to these properties that may not be well-served by using off-the-shelf tests like those provided by Tonhauser (2012). I address this below in the discussion of the discourse properties of direct and indirect discourse.

2.2 Evidence for separation in processing

Most importantly for our present purposes, appositives have some special status in processing, shown in offline acceptability judgments and eyetracking-while-

reading. I'll review the main findings in the below sections. These results support the generalization that content within appositives is downweighted in later computations. I'll argue that this pattern would be best explained by a hypothesis in which speech acts are a fundamental unit of parsing, such that linguistic material is always downweighted across a speech act boundary.

Differential impact of complexity

Dillon et al. (2014) report a series of naturalness judgement experiments on items with RRCs and ARCs. Convergent evidence across several of these experiments suggests that the complexity of an ARC is to some degree disregarded in acceptability judgements provided to the sentence which contains it, in a way which the complexity of an RRC is not. To present this finding most simply, I will focus on some of the conditions they examined in their Experiment 3.

Participants in the study rated 24 critical items in conditions like those depicted in Table 1, where the type of RC present in the item and the amount of complexity within it were manipulated. Complexity was in all cases provided by modifying the subject of the RC with a (sometimes prepositional) object relative clause (ORC). In addition, in this experiment the authors looked at critical RCs both modifying the subject and the object of the matrix sentence. I report here only the results for RCs modifying the subject.

The critical finding is that, as demonstrated numerically and graphically in Table 2 and Figure 1, complexifying material within an ARC is reflected by a smaller reduction in the naturalness judgement provided for a sentence than the very

same material when in a RRC. Dillon et al. (2014) report a difference of differences interaction, such that the effect of complexity is significantly different across different structure conditions. This difference of differences also held for stimuli where the RC modified the object, and was thus sentence-final.

Table 1: Sample stimulus (excerpted) from Dillon et al. (2014), Experiment 3.

Complexity	RRC	ARC
Short	That evil man who was on the cruise tried to intimidate the waitress.	That evil man, the one who was on the cruise, tried to intimidate the waitress.
Long	That evil man who was on the cruise Mary took to the Pacific Islands tried to intimidate the waitress.	That evil man, the one who was on the cruise Mary took to the Pacific Islands, tried to intimidate the waitress.

Table 2: Mean naturalness ratings (7 = very natural) (and standard error) by condition from Dillon et al. (2014), Experiment 3 (limited to subject relative clause conditions).

Complexity	RRC	ARC
Long	4.9 (0.15)	6 (0.09)

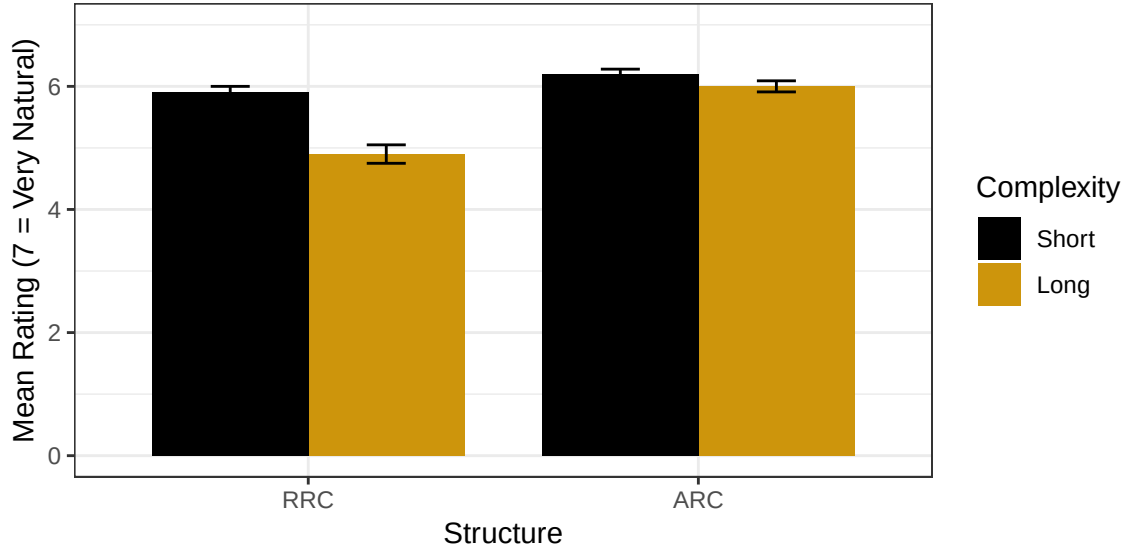


Figure 1: Mean naturalness ratings by condition from Dillon et al. (2014), Experiment 3 (limited to subject relative clause conditions). Error bars indicate ± 1 SE.

Complexity	RRC	ARC
Short	5.9 (0.1)	6.2 (0.08)

In order to interpret these results, we have to reflect on the linking hypotheses we are comfortable assuming for acceptability judgement data. A relatively conservative linking hypothesis would treat the acceptability judgement as a typical decision-making task, which ought to be prone to biases and differential treatment of evidence. Under this linking hypothesis, we are free to conclude that the crucial difference underlying the observed effects is within the decision-making process, such that the reflection and computation which produces a participant’s acceptability judgement treats material differently depending on its context within the sentence under consideration. Specifically, I suggest that the acceptability judge-

ment appears to downweight evidence of complexity encountered while processing an ARC.

We could instead assume the very strong linking hypothesis that these judgements directly reflect the cognitive difficulty participants were engaged in while reading a given sentence. In this case, we would have to conclude that the same material when present in an ARC was easier to process, or presents less difficulty to later processing,³ than when it is present in a RRC. I will offer discussion below on why this conclusion should be called into doubt given further data, but I will also note here that we can be skeptical of such a conclusion for reasons of conservativity alone. Psycholinguistics knows little about the acceptability judgement task, and the burden of proof lies on the argument that it is a direct reflection of cognitive difficulty.

Differential ease of filler-gap integration

Dillon et al. (2017) continue the investigation of the differences in processing ARCs and RRCs, but alter the critical manipulations to examine the acceptability and online processing of ARCs and RRCs which intervene in filler-gap dependencies.

In their Experiment 1, another naturalness judgement task, all critical items featured manipulation of structure, again between RRC and ARC, but also the presence of a filler-gap dependency which crosses them. A sample item is in Table 3,

³One can imagine a story here that ARCs interfere less than RRCs in subject-verb processing. This story would predict that when the RC is sentence final, the interaction would disappear, but this is not consistent with the Dillon et al. results.

with critical gaps in +Filler conditions indicated as “_____”. Results are displayed in Table 4 and Figure 2. The authors again observe a difference of differences reflecting a differential effect of some manipulated cost over the two structures, so that the cost results in a smaller naturalness penalty for the ARCs. The relevant difference between this finding and those reported above is that the manipulated cost was the presence of a filler-gap dependency crossing the critical RC, rather than the presence of complexifying material within the RC.

Table 3: Sample stimulus (excerpted) from Dillon et al.
(2017), Experiment 1.

Filler	RRC	ARC
-Filler	The butcher asked if the lady who bought Italian ham was cooking dinner for her guests.	The butcher asked if the lady, who bought Italian ham, was cooking dinner for her guests.
+Filler	The butcher asked who the lady who bought Italian ham was cooking dinner for _____.	The butcher asked who the lady, who bought Italian ham, was cooking dinner for _____.

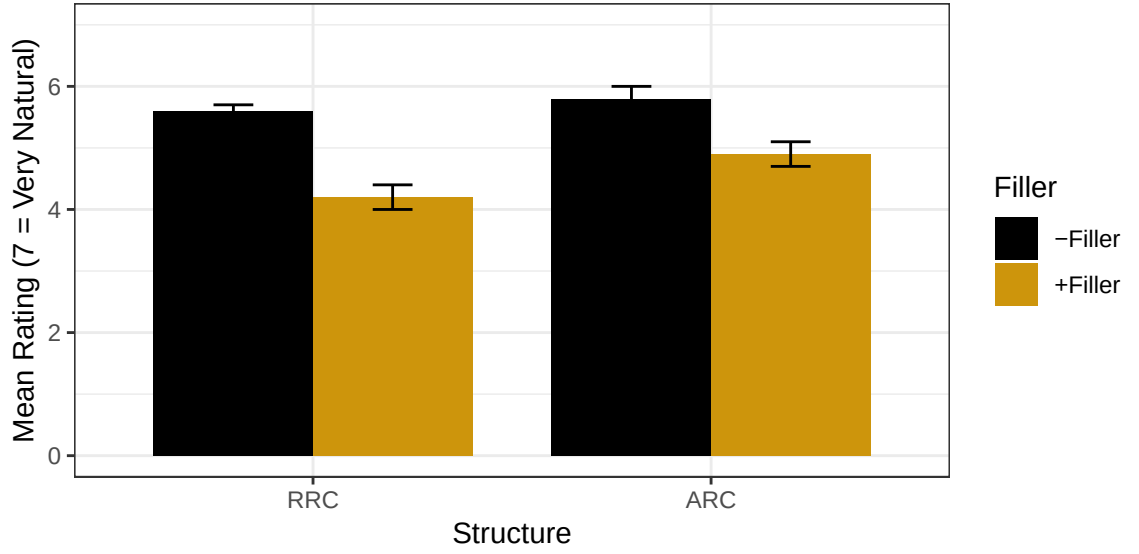


Figure 2: Mean naturalness ratings by condition from Dillon et al. (2017), Experiment 1. Error bars indicate ± 1 SE.

Table 4: Mean naturalness ratings (7 = very natural) (and SE) by condition from Dillon et al. (2017), Experiment 1.

Filler	RRC	ARC
-Filler	5.6 (0.1)	5.8 (0.2)
+Filler	4.2 (0.2)	4.9 (0.2)

Again, the linking hypothesis we adopt for the judgement task affects which conclusions we are able to draw from this effect. If we continue adopting the conservative linking hypothesis, we are free to consider that the effect is either one of differential online computational difficulty or a difference of evidence use in decision making.

The authors' Experiment 2 provides evidence for the former, an online

difference between items with RRCs and those with ARCs. Go-past times for the final region are longer in the +Filler conditions, where the region contains a gap site and thus is the site of filler integration: we can take this as a signature of the cognitive demands of filler retrieval and integration. Drawing on the large body of psycholinguistic work treating this task as one of cue-based retrieval, where one of the main causes of cognitive demand is the presence of interfering content, the authors suggest that the differences in reading time penalties are produced by differential interference. That is, the processor received more interference during filler retrieval when a RRC intervened between the filler and gap than when an ARC intervened. This likely underlies the acceptability judgement effect from Experiment 1 in this thesis.⁴

Discarding alternative hypotheses

In this section, I will briefly consider some approaches which may a priori seem attractive in their ability to provide an explanatory account for the data considered above. In each case, I'll discuss the evidence which leads me to reject that account.

Let us first revisit an alternative already touched upon in the above review of the data: a proposal rooted entirely in differences in online processing costs during reading of an RRC or ARC. This is proposed seriously in Dillon et al. (2014), as it is consistent with the judgement data offered there. Given the persistence of the

⁴Note that while we are also free to consider an evidence-weighting-in-decision-making explanation for these findings as well, it's not clear what that would look like. In what way would being able to downweight the contribution of the ARC in decision-making reduce the acceptability penalty associated with filler-gap processing?

difference of differences no matter whether critical RCs intervened in subject-verb dependencies or fell at the end of the sentence, the authors reject the possibility of a solely retrieval-based account. Instead, they suggest, the difference could be found in the RC region itself: perhaps complex ARCs are easier because they are processed using separate resources or a separate memory store. However, this would have predicted online effects at the RC region in the Dillon et al. (2017) eyetracking experiment, which are not observed. This, it seems, forces us towards a generalization about how this material is dealt with in downstream processing.

Dillon et al. (2014) also entertain that some subset of the results might be predicted by an account wherein, drawing on the notion that they are parenthetical and secondary content, ARCs simply receive less attention during online processing than RRCs. In some ways, this mechanically works similarly to the downweighting approach I foreground, in that there is a difference in the transmission of evidence to the judgement process, but here it would be a difference in evidence uptake rather than evidence use. Evidence from Experiment 2 in Dillon et al. (2014) suggests that this can't be the case: acceptability judgements are equally sensitive to simple errors of agreement in RRCs and ARCs. In fact, this might be a problem even for the downweighting approach, unless some distinction is made between error-detection and complexity-detection. An attentional account, and not a downweighting account, seems to require that even detection of common errors would be subdued within ARCs.

Even assuming that indeed ARCs somehow display differential persistence into following computation, the reasons why this is might be debated. A first stab

proposed by Dillon et al. (2014) is to pin the behavior of ARCs on their not-at-issue status. To test this directly, Kroll and Wagers (2019) capitalize on the apparent possibility of appositives which interact with the QUD as an opportunity to manipulate at-issueness independent of structure, testing the hypothesis that at-issueness in particular is driving the observed differential impact of complexity when within ARCs. In their Experiment 2 and 3, they examine whether providing a QUD relevant to the detail within an ARC, and thus, on some definitions, raising that content towards at-issueness, will eliminate the critical difference of differences between ARC and RRC structures. Experiment 2 does so with sentence-medial RCs, and Experiment 3 with sentence-final, also thus probing whether utterance-final position can be observed to have an effect parallel with its increase in apparent at-issueness.

I provide a sample item from Experiment 2,⁵ and the results obtained from their experiment, in Tables 5-6 and Figure 3. As apparent from the plots, the difference of differences interaction holds over both at-issueness conditions, and indeed this is the case as well for sentence-final appositives in Experiment 3.

⁵Each item was also manipulated for whether an explicit QUD made the content of the RC at-issue: for instance, in this case, the NAI question was “What is the bear wearing?” and the AI question was “Where is the bear standing and what is it wearing?”

Table 5: Sample stimulus from Kroll and Wagers (2019), Experiment 2.

Complexity	RRC	ARC
Short	The bear who is standing on the ball is wearing a hat.	The bear (who is standing on the ball) is wearing a hat.
Long	The bear who is standing on the ball the trainer rolled across the room is wearing a hat.	The bear (who is standing on the ball the trainer rolled across the room) is wearing a hat.

Table 6: Mean naturalness ratings (7 = very natural) (and standard error) by condition from Kroll and Wagers (2019), Experiment 2.

Complexity	AI	RRC	ARC
Short	NAI	6 (0.1)	6 (0.1)
Long	NAI	4.3 (0.1)	5.4 (0.1)
Short	AI	6.3 (0.1)	6.3 (0.1)
Long	AI	4.6 (0.1)	5.4 (0.1)

To the extent that we expect at-issueness to be a quality measureable

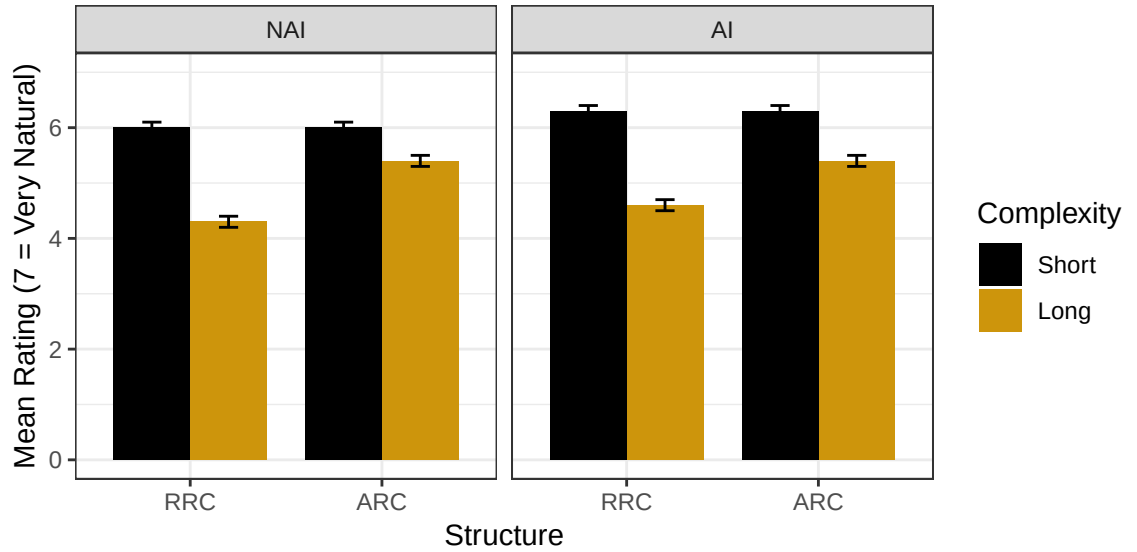


Figure 3: Mean naturalness ratings by condition from Kroll and Wagers, Experiment 2. Error bars indicate ± 1 SE.

directly from ability to interact with discourse processes like denial and QUD⁶, these results suggest that the hypothesis of these parsing consequences as based on at-issueness is not supported.

In sum, it does not seem possible to adequately form an explanatory account that derives the behavior of ARCs in these experiments from a lack of attention paid to ARCs, from any differences localized to the processing of the RC itself, from retrieval interference alone, or based on the treatment of not-at-issue information in processing.

Rather, the data is uniformly consistent with another hypothesis, the importance of speech acts as a minimal processing unit for some computation. I will cash this out as the ability for the parser to downweight subordinate speech acts when engaging in computations or retrievals involving the superordinate speech act.

⁶Note that Potts (2012) maintains exactly the opposite, that such appositives remain fundamentally not-at-issue despite their apparent foregrounding.

- (14) Hypothesis: Humans’ processing of natural language must compose at the speech act level to compute discourse effects. This level of organization is accessible to future processing, such that linguistic material which is part of a subordinate speech act may be downweighted or excluded during computation related to a superordinate speech act.

Here, the term *speech act* is meant to be interchangeable with the *discourse units* Dillon et al. (2017) refer to in their similar proposal. For my purposes, speech acts are the maximal units which serve to generate a single contiguous discourse effect, in a formalization such as Farkas and Bruce (2010). An assertion, even one which has embedded clauses or conjunction, comprises a single speech act, as it constitutes a single gesture of commitment and intent to update the common ground in a particular way. A simple polar question, too, comprises a single speech act, in which there is no (or trivial) commitment on the part of the speaker. As argued by Murray (2014) and AnderBois et al. (2015), ARCs also have a distinct life in discourse, such that they add to the common ground in a manner different from standard assertions. Consider, for instance, that the conversational profile of an appositive does not vary depending on the type of speech act it is embedded within. In a single sentence which contains an ARC, then, there are two speech acts: that of the matrix, and the secondary, syntactically subordinate update provided by the ARC.

It is non-controversial for most psycholinguists that the human processor is able to distinguish amongst content encountered in different sentences.⁷ It seems as

⁷Though we occasionally see cases of linguistic content which is not part of the sentence and

if most major semantic integration happens at the sentence level; this is how many authors explain frequently-observed sentence-final ‘wrap-up effects.’ The hypothesis in (14) rests, intuitively, on the proposal that at least one of the reasons for a sentence-level divide is the necessity of a speech-act-level divide.

We might also pin the above effects on something related to objects of the same size as speech acts without requiring speech acts themselves to be critical processing units. This is entertained most deeply in Kroll and Wagers (2019), who examine the proposal that it is the speech-act-sized prosodic unit, or prosodic features correlated with it, that generates the apparent downweighting of ARC content. They find that when RRCs are forced to be read (and thus implicitly prosodified) as a unit, the critical interaction disappears – that is, RRCs with such prosody are apparently downweighted just like ARCs. Nevertheless, the authors’ evidence is split, as manipulation of adverbs to force prosodic boundaries in another experiment do not have the same effect. I will revisit prosody below as a question of interest.

If indeed the hypothesis in (14) is true, we would observe the same downweighting of material within another subordinated speech act. Thus, the purpose of the experiments I will report here will be to examine the status of these effects for another test case, direct discourse speech reports.

indeed irrelevant for the processing task providing some amount of interference: see, for instance, Van Dyke and McElree (2006).

3 Direct discourse: The same and different

Natural language offers us another form of syntactically subordinated, discourse-independent speech act in the form of direct discourse (DD) speech reports. While DD and ARCs as constructions are vastly different in form and function, I will show here that they share many of the relevant properties discussed above: not only does DD appear similarly incapable of interacting with the QUD, but like ARCs it demonstrates interpretive freedom in all cases but discourse anaphora and presupposition. Furthermore, evidence from the perceptual simulation of DD in reading provides theory-independent support that DD must be read as an independent speech act.

3.1 The basics of reported speech

English permits two ways of expressing reported speech in the complement of a verb of speech like *say*, direct discourse (DD) (15) and indirect discourse (ID) (16).

(15) Rebecca said “The nurses in my ward are well-paid.”

(16) Rebecca said that the nurses in her ward were well-paid.

(17) a. * Who did Rebecca say “I saw yesterday”?

b. Who did Rebecca say she saw yesterday?

(18) a. Rebecca said “Eek!”

b. * Rebecca said that eek.

Like ARCs and RRCs, these alternatives are minimally different in notable ways. While ID is integrated in syntax to the extent that it permits binding and sequence of tenses phenomena (Cappelen & Lepore, 2017), and in some cases permits Wh-movement (17), DD maintains a non-integrated representation that simulates the reported utterance mimetically (Clark & Gerrig, 1990; Davidson, 2015). As such, DD even permits non-propositional and ungrammatical content, while ID only permits grammatical, propositional content in its capacity as transmission of the propositional content of the reported utterance (18).

3.2 Comparing relevant features of appositives and direct discourse

Like appositives, it seems possible to analyze direct discourse as an embedded speech act. In the same way, it fails to be opaque to some syntactic/semantic phenomena, and it clearly invokes a separate speech act, though the point is a little different from the independence of ARCs: ARCs seem to represent a separate, parallel speech act on the part of the speaker within the same discourse as their main assertion. Direct discourse contributes a demonstration (Clark & Gerrig, 1990; Davidson, 2015) as part of the same speech act (indeed, we don't observe the absence of scope effects), but in doing so calls forward a speech act performed at a different time by a different speaker. Critically: it certainly does not assert speaker commitment to its content in the same way as a matrix-level assertion.

Looking to compare the discourse status of DD and ARCs, we might turn to the tests for at-issueness derived from Potts (2005) and Tonhauser (2012), discussed above. These suggest that indeed, the propositional content of DD is absent from

the at-issue contribution of a sentence containing it. The propositional content of DD cannot felicitously be targeted by a polar response particle (19) and cannot felicitously answer a QUD (20) (except for in its function as an object, as in 21). Note that the latter is a point of disagreement in the literature: Bary and Maier (2018) suggest that DD may answer a QUD, though the authors provide only a judgement.

(19) A: Rebecca said "The nurses in my ward are well-paid."

B: No, they aren't.

(20) A: Are the nurses in this hospital well-paid?

#B: Rebecca said "The nurses in my ward are well-paid."

(21) A: What did Rebecca say?

B: Rebecca said "The nurses in my ward are well-paid."

The application of tests for projectivity also brings complications in this case. On one hand, it is clear that the propositional content of DD is not influenced by the negation (22) or interrogativity (23) of the matrix clause: in both cases the reported speech has the same content and illocutionary force. On the other hand, these scopal operators do influence whether the matrix speaker commits to the existence of the reported speech: in neither of these cases does the speaker assert that Rebecca actually made the depicted utterance.

(22) Rebecca didn't say "The nurses in my ward are well-paid."

(23) Did Rebecca say "The nurses in my ward are well paid"?

While we learn some relevant generalizations from the applications of these tests, I will note that they should not be taken as satisfactory demonstration that, say, DD contributes NAI information. For one, this isn't accurate: all that the diagnostics ever might have shown is whether certain aspects of DD, of which we've seen multiple, contribute any at-issue content.⁸ Secondly, the phenomena targeted by these diagnostics may be more diverse than testing interaction with the QUD: plenty of information in an utterance is inaccessible to polarity particles like *no*, and question-answer pairs can be incongruous for reasons unrelated to the AI/NAI divide.

What I will assert is that the apparent independence of the content of the reported speech in a token of DD supports the notion that it is a distinct discourse unit from the rest of the clause, in a way which bears some resemblance to the division between an ARC and its matrix.

Before moving forward, I will note that such independence, just as with ARCs, does not preclude DD from introducing discourse referents for discourse anaphora like *they* (24) and satisfying presuppositions like the one introduced by *too* (25).

(24) Rebecca said “The nurses in my ward are well-paid.” I checked, and they are.

(25) Rebecca said “The nurses in my ward are well-paid.” I checked, and mine are too.

⁸That is to say, just because some content fails to make an at-issue contribution does not mean that it makes a not-at-issue contribution.

3.3 The processing of direct discourse: Perceptual simulation

Outside of formal properties, existing research on the reading of DD has revealed one very salient feature it does not share with ID and presumably also not ARCs: the perceptual simulation of the represented speaker. Not only does activity in voice-selective areas of the auditory cortex increase during the reading of DD passages (Yao et al., 2011), but also first-pass and go-past reading times of DD are modulated by the described speed of the reported speech (Stites et al., 2013; Yao & Scheepers, 2011): even solely on the cue of a single pre-verbal temporal adverb indicating the reported speech act was originally delivered quickly, participants will increase the speed at which they read the passage (Stites et al., 2013). It is particularly notable that this effect is manifest only within passages of DD: indirect discourse passages otherwise identical do not show any effect of the temporal adverb.

The importance of such a finding is made clear when considering that readers are observed to perform the same perceptual simulation when reading attributed text (Alexander & Nygaard, 2008; Zhou & Christianson, 2016), that is, single simple utterances which were presented in a context which featured extra-stimulus information providing information about the speaker. It seems, then, that perceptual simulation provides insight into an online process in which DD tokens are processed alike with the speech acts they stand in for. This, alongside the formal properties outlined above, provides supportive evidence that DD, like ARCs, constitutes a unit which is treated independently from superordinate material, even during online

reading.

3.4 Predictions

Given that DD appears to also be the right kind of syntactically-embedded parallel discourse unit, and the strong hypothesis that all rightward speech act boundaries result in discounting from downstream processes, we predict that direct discourse processing will demonstrate the same offline judgement effects observed by Dillon et al. for ARCs. That is, complexity within a token of DD will result in a smaller penalty than the same complexity within a token of ID, and likewise, filler-gap dependencies spanning DD will result in a smaller penalty than the same dependency spanning a token of ID. These are exactly the predictions I test in the experiments reported below, the former in Experiments 1 and 2 and the latter in Experiment 3.

4 Experiment 1: Costs for complexity within reported speech

In Experiment 1, I conduct a straightforward extension of the 2x2 from Dillon et al. (2014) discussed above, looking at the effects of internal complexity within DD and ID. The hypothesis of speech act bounding predicts that a first-order difference of differences interaction between Structure and Complexity will obtain such that the penalty associated with Long conditions in ID is attenuated in DD.

4.1 Method

Materials

I constructed 32 items to be manipulated in a 2x2 contrast of Complexity [Short, Long] and Structure [ID, DD]. Just as in Dillon et al. (2014), ORCs were used to lengthen the critical regions, here speech report clauses. 24 of the 32 items feature identical ORCs to the items of Dillon et al. (2014), and the others were closely modeled on them, such that all complex conditions feature an ORC with an inanimate head extracted from object or prepositional object position, and a proper name subject. An example is given in Table 7. To adhere with prescriptive standards, a comma preceded instances of DD.

Table 7: Sample stimulus from Experiment 1.

Item	Complexity	ID	DD
1	Short	Evan said that the cruise departed three hours behind schedule.	Evan said, “The cruise departed three hours behind schedule.”
1	Long	Evan said that the cruise Mary took to the Pacific Islands departed three hours behind schedule.	Evan said, “The cruise Mary took to the Pacific Islands departed three hours behind schedule.”

Interleaved with these experimental items, participants saw the 72 fillers

used in Dillon et al. (2014) and Dillon et al. (2017), which include a mix of grammatical and ungrammatical (verb agreement and reflexive number mismatch, NPI illusion) cases.

Participants and procedure

48 native English-speaking participants with at least the equivalent of a high school diploma were recruited on Prolific, and paid in accordance with a \$12/hr wage. After completing a consent form, they were instructed that they would be rating sentences of English for their naturalness on a 7-point scale, using keys or the mouse. The “1” point was labelled “very unnatural,” and “7” was labelled, “very natural.” A few trials of guided practice identical to the procedure used in Dillon et al. (2014) and Dillon et al. (2017) instructed participants on how we expected them to use certain regions of the scale, before they proceeded to further practice. After the final practice item, participants answered questions in a short burn-in period to continue familiarizing them with the scale before any critical items were shown.

Critical items were Latin-squared into 4 forms with equal participation in each, before being randomized and interleaved with the 72 base fillers. Mandatory 10s breaks were interspersed throughout the experiment every 15 trials. At the end of the experiment, participants responded to a short survey about their language experience and their experience on the task, before receiving credit.

All elements of the procedure for other experiments were identical.

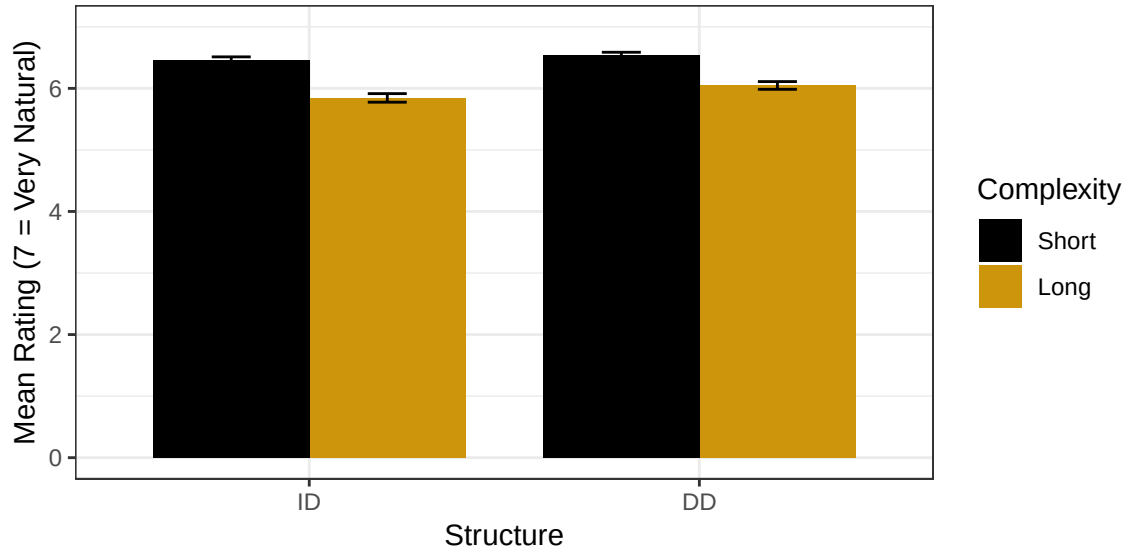


Figure 4: Mean naturalness ratings by condition from Experiment 1. Error bars indicate ± 1 SE.

4.2 Results

Ratings for trials which took longer than 30s were discarded, removing a total of 20 critical trials. Mean responses received are provided in Table 8 and Figure 4.

Table 8: Mean naturalness ratings (7 = very natural) by condition for Experiment 1.

Complexity	ID	DD
Short	6.46 (0.05)	6.54 (0.05)
Long	5.85 (0.07)	6.05 (0.06)

Statistical analysis was carried out by fitting an cumulative link mixed-effects model with Laplace approximation and a probit link, using the `ordinal`

package in R, with maximal random effects and treatment coded variables (treatment level specified in parentheses). Parameters for this model are given in Table 9. We observe a significant main effect of complexity ($z = -6.10, p < .001$), and no other significant effects. Most importantly, the predicted interaction of complexity and structure failed to reach significance ($z = 1.01, p = .31$).

Table 9: Ordinal mixed-effects model fit to the data from Experiment 1.

Effect	Estimate	SE	z	p
Structure (DD)	0.1	0.18	0.54	.59
Complexity (Long)	-1.14	0.19	-6.1	<.001
Structure $\times$ Complexity	0.18	0.18	1.01	.31

4.3 Discussion

In Experiment 1, we fail to find the predicted interaction such that the penalty for increased complexity within ID attenuates in DD. The results are inconsistent with the data on ARCs, and thus fail to support the hypothesis that all speech act boundaries should serve as meaningful divides in processing.

It is apparent that all critical items in Experiment 1 were rated very highly by participants. Given that there is a numeric trend towards the predicted interaction, we might wonder whether it escaped significance due to scale compression or a ceiling effect. I will note that two additional experiments on similar items not

reported here featured biases shifted toward this ceiling and away from this ceiling, but no significant interaction was observed in either.

5 Experiment 2: Comparing with appositives

The failure to observe the predicted interaction in Experiment 1 might call into question whether the original appositive results are themselves replicable with the increased power in the experiments reported in this thesis.⁹ The experiments in Dillon et al. (2014) feature 24 items and comparison between 8 cells, such that each participant only provided 3 observations per cell. In my design, which features 32 items and comparison between 4 cells, each participant is providing 8 observations per cell. This offers the ability to estimate the population parameters of interest with much more accuracy. I note that low power crucially can lead to an increased rate of both Type I and Type II errors, so it may be the case that increased power is revealing the original effect to be unreliable.

To ensure the security of the findings which this investigation is predicated on, in Experiment 2 I aim to replicate the critical finding of Dillon et al. (2014), and thus be able to argue more convincingly that Experiment 1 represents an informative null result.

⁹Though it is worth noting that Kroll and Wagers (2019) already have replicated the major finding with novel items and greater power than Dillon et al.

5.1 Method

Materials

24 items were taken directly from Dillon et al. (2014), and an additional 8 items were generated to the Dillon et al. template based on the 8 items created for Experiment 1. A sample item is provided in Table 10.

Table 10: Sample stimulus from Experiment 2.

Item	Complexity	RRC	ARC
1	Short	That man who was on the cruise tried to throw a waitress overboard.	That man, the one on the cruise, tried to throw a waitress overboard.
1	Long	That man who was on the cruise Mary took to the Pacific Islands tried to throw a waitress overboard.	That man, the one on the cruise Mary took to the Pacific Islands, tried to throw a waitress overboard.

Participants and procedure

Participants were recruited and completed the task as in Experiment 1.

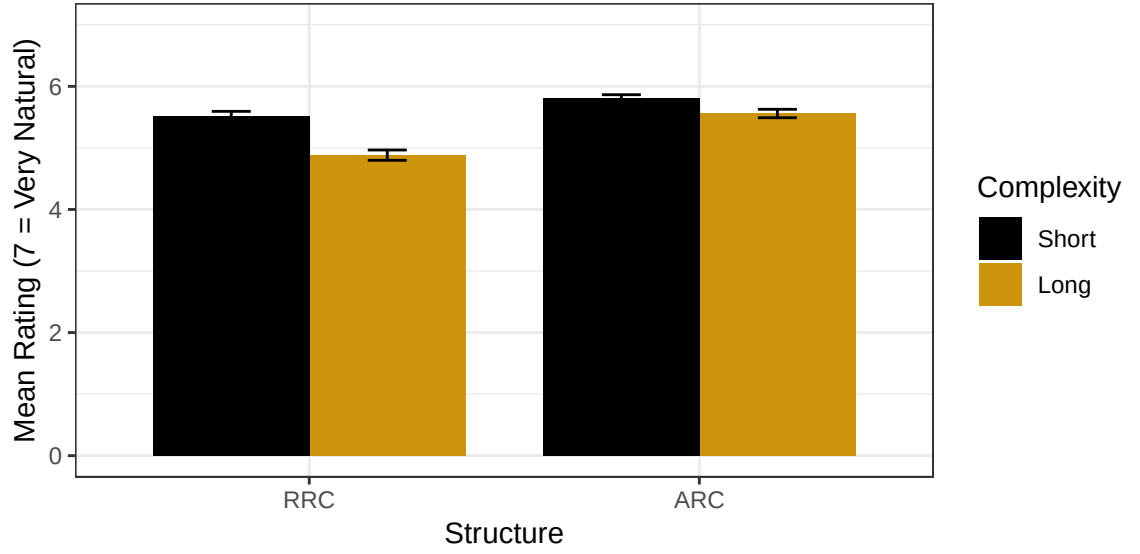


Figure 5: Mean naturalness ratings by condition from Experiment 2. Error bars indicate ± 1 SE.

5.2 Results

Ratings for trials which took longer than 30s were discarded, removing a total of 58 critical trials. Mean responses received are provided in Table 11 and Figure 5.

Table 11: Mean naturalness ratings (7 = very natural) by condition for Experiment 2.

Complexity	RRC	ARC
Short	5.52 (0.07)	5.8 (0.06)
Long	4.88 (0.08)	5.56 (0.07)

Statistical analysis was carried out as in Experiment 1. We observe a significant main effect of complexity ($z = -4.52$, $p < .001$), and a significant main

effect of structure, such that short ARCs were rated more natural than short RRCs ($z = 2.16, p = .031$). Most importantly, the predicted interaction of complexity and structure was borne out, such that the complexity penalty was smaller for ARCs ($z = 2.30, p = .021$).

Table 12: Ordinal mixed-effects model fit to the data from Experiment 2.

Effect	Estimate	SE	z	p
Structure (ARC)	0.36	0.17	2.16	.031
Complexity (Long)	-0.66	0.15	-4.52	<.001
Structure $\times$ Complexity	-0.34	0.15	2.3	.021

5.3 Post-hoc comparison with Experiment 2

Table 13 reports the parameters for a Bayesian ordinal mixed-effects model fit over data from Experiments 1 and 2 using Experiment as a two-way between-subjects factor with E1 as the treatment level and E2 as reference level. Posterior values for $\hat{\beta}$ and σ_{β} along with 95% credible intervals (CRIs) are provided. Parameters whose 95% CRI does not contain 0 are assumed to indicate reliable effects. For this analysis, no responses were removed. We find a significant main effect of Experiment such that E1 exhibits higher ratings than E2 (considering only short, unmarked structures), a small significant main effect preferring ARCs over RRCs (considering only short structures in E2), and a standard negative effect of com-

plexity such that complexity decreases ratings (considering only RRCs in E2). We also observe an interaction of experiment by complexity, such that the increase in ratings for E1 is slightly smaller when comparing the complex conditions, and a borderline significant interaction of complexity by structure, equivalent to the interaction found looking at E2 in isolation.¹⁰ Crucially, we fail to observe the three-way interaction that would demonstrate a significant replicating interaction in the RC experiment being eliminated in the DD experiment.

Table 13: Bayesian ordinal mixed-effects model fit over the data for Experiments 2 and 1 as a between-subjects manipulation.

Effect	Posterior $\hat{\beta}$	Posterior σ_{β}	95% CRI	95% CRI
			Lower	Upper
Experiment (E1)	1.13	0.09	0.95	1.32
Struc.	0.28	0.11	0.07	0.49
(ARC/DD)				
Complexity	-0.5	0.09	-0.68	-0.32
(Long)				
Expt. $\times$ Struc.	-0.13	0.13	-0.38	0.13
Expt. $\times$ Compl.	-0.3	0.12	-0.54	-0.06
Struc. $\times$ Compl.	0.22	0.11	0	0.44

¹⁰That is, it appears that in the un-excluded data set, the interaction in E2 is only borderline significant.

Effect	Posterior $\hat{\beta}$	Posterior σ_{β}	95% CRI	95% CRI
			Lower	Upper
E. $\times$ S. $\times$ C.	-0.17	0.17	-0.49	0.16

5.4 Discussion

We replicate the critical effect of Dillon et al. (2014) with our population and power, which may leave us with more confidence that the failure to find that effect in Experiment 1 with DD is an informative failure. However, a Bayesian-fit model over the two experiments does not find the Structure $\times$ Complexity parameters fit for the two experiments to be significantly different. Thus, while we may still have confidence that the measures and sample here had the ability to reveal the critical interaction in E1, and did not, it cannot be said for certain that the effect does not exist at the population level.

I take this at present as sufficient evidence to conclude a meaningful difference between the expected effect, given the observations with ARCs, and what we actually observe with DD. It seems that position within an instance of DD does not insulate complexity from affecting acceptability judgements given for the entire sentence.

6 Experiment 3: Dependencies across reported speech

Experiment 3 turns to extending the findings of Dillon et al. (2017), investigating discrepancies in filler-gap processing between ID and DD. This offers

another opportunity to test the predictions of the hypothesis that speech acts serve as meaningful segments for processing: this hypothesis predicts that resolving dependencies which cross DD will be less difficult than resolving dependencies which cross ID, by virtue of the processor’s ability to downweight DD material and thereby eliminate sources of interference.

6.1 Method

Materials

Similar to the construction of materials for Experiment 1, 32 items were created based on the qualities of the 24 items in Dillon et al. (2017). As in that study, each item was manipulated in a 2x2 cross of Structure [ID, DD] $\times$ Filler [-Filler, +Filler], where the presence of a filler-gap dependency crossing the critical speech report was manipulated by the type of embedded question used, between embedded content questions and embedded polar *if* questions. Speech reports were always contained within relative clauses modifying the subject of the embedded question. An example item is in Table 14.

Table 14: Sample stimulus from Experiment 3.

Item	Filler	ID	DD
1	-Filler	The butcher asked if the lady who said that she would like a nice big ham was cooking for a party.	The butcher asked if the lady who said, “I would like a nice big ham,” was cooking for a party.
1	+Filler	The butcher asked who the lady who said that she would like a nice big ham was cooking for.	The butcher asked who the lady who said, “I would like a nice big ham,” was cooking for.

Participants and procedure

Participants were recruited and participated as in Experiments 1 and 2.

6.2 Results

Mean responses received are provided in Table 15 and Figure 6. Statistical analysis was performed on the unexcluded data set using a Bayesian-fit ordinal model, reported as above in Table 16.

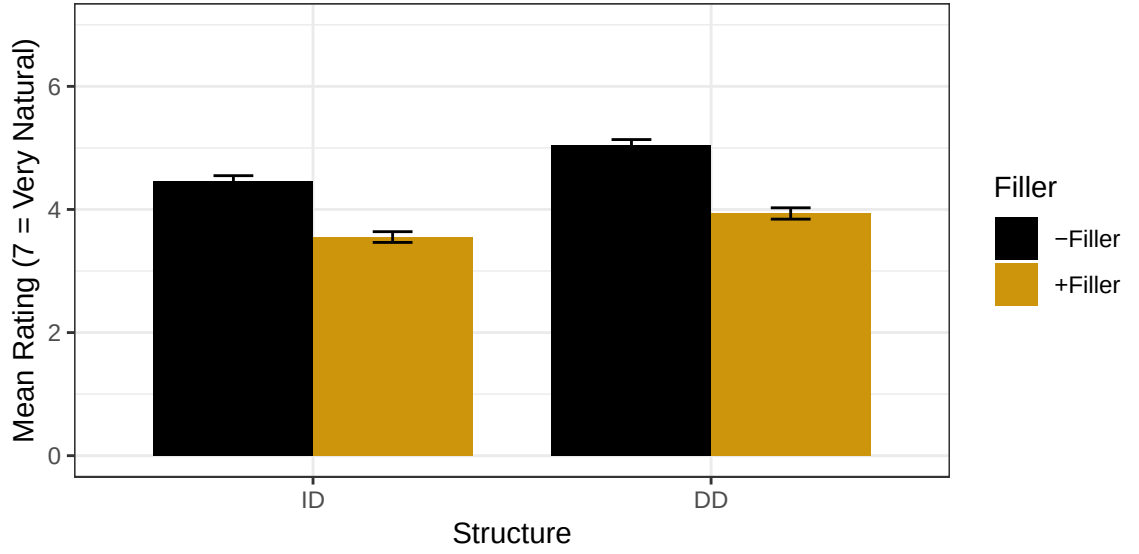


Figure 6: Mean naturalness ratings by condition from Experiment 3. Error bars indicate ± 1 SE.

Table 15: Mean naturalness ratings (7 = very natural) by condition for Experiment 3.

Filler	ID	DD
-Filler	4.46 (0.09)	5.05 (0.08)
+Filler	3.55 (0.09)	3.93 (0.09)

We observe a significant main effect (a penalty) of the presence of a filler-gap dependency ($\hat{\beta} = -0.7, 95\% \text{ CRI} = (-.99, -.41)$), and a significant main effect of structure such that DD is preferred to ID ($\hat{\beta} = 0.46, 95\% \text{ CRI} = (0.30, 0.62)$). The predicted interaction of filler and structure failed to reach significance ($\hat{\beta} = -0.19, 95\% \text{ CRI} = (-0.42, 0.03)$), this time trending numerically in the opposite direction from predicted.

Table 16: Bayesian ordinal mixed-effects model fit to the data
from Experiment 3.

Effect	Posterior $\hat{\beta}$	Posterior σ_{β}	95% CRI	
			Lower	Upper
Structure	0.46	0.08	0.3	0.62
(DD)				
Filler	-0.7	0.15	-0.99	-0.41
(+Filler)				
Struct. $\times$ F.	-0.19	0.11	-0.42	0.03

6.3 Discussion

In accordance with failures to find evidence for DD as a segmented processing unit akin to ARCs in Experiment 1, Experiment 3 fails again to find an interaction such that DD imposes less upon matrix computation than ID. Recall that in particular, this effect with ARCs was revealed in eyetracking to correspond to differential difficulty of filler integration, manifest as faster go-past times on the region containing the gap. My hypothesis predicts that DD, as a similarly subordinate speech act, should also contribute less difficulty - in both cases, the contents of the secondary speech act might be discounted such that they lead to less interference during retrieval of the filler. This is not observed: it is just as hard to retrieve fillers in the immediate context of DD and ID. We thus have no evidence that DD

engenders any separability from its matrix content. The consistency of these nulls lends to their weight.

7 General discussion

The predictions of the hypothesis that speech acts serve to delineate processing units have been unsupported by the data presented here. In two cases, no interaction was observed that indicated DD, as a case of a subordinate independent speech act, shared the properties of ARCs as observed with consistency in Dillon et al. (2014), Dillon et al. (2017), and Kroll and Wagers (2019).

This is particularly surprising given existing experimental work on the reading of DD, in which readers modulate reading rate just as they do in matrix-level attributed speech (Alexander & Nygaard, 2008; Stites et al., 2013; Yao & Scheepers, 2011; Zhou & Christianson, 2016). This would suggest that from early reading, the processor separates DD as the speech act of another individual, and subjects it to special treatment as a result.

I see two ways to interpret this result. Firstly, it may be the case that, despite the ultimate truth of the speech act hypothesis presented above, DD does not accurately represent an independent and separable speech act in the relevant manner. Alternately, perhaps the speech act hypothesis is incorrect, and should be replaced with a hypothesis targeting the influence of implicit prosody, as argued by Kroll and Wagers (2019). I outline these two conclusions individually below.

7.1 Revisiting the comparability of DD and ARCs

As mentioned above, presuming similarity between DD and ARCs is an unexpected comparison among many dimensions. Perhaps the argument that DD

is the same kind of subordinate speech act is not fully defensible. Indeed, as already noted, DD is more integrated with its matrix than ARCs on one measure: interactions with scope. ARCs are unaffected by negation, and always represent an assertion. The propositional content of DD is the same way, but the linguistic content is subject to the same scope context as the rest of the VP which embeds it, in ways which can suppress the actual existence of the DD as someone else's utterance.

- (26) The candidate said, "The decision of the American people is irrelevant."
- (27) The candidate didn't say, "The decision of the American people is irrelevant."
- (28) The candidate might say, "The decision of the American people is irrelevant."
- (29) The parody article claimed that the candidate said, "The decision of the American people is irrelevant."
- (30) Did the candidate say, "The decision of the American people is irrelevant"?

I have focused thus far on cases where DD is embedded in a positive assertion (26). In these cases, the sentence asserts a particular individual uttered some speech with given verbal content, and also re-presents that verbal content, such that the original words and message of the reported speech act are accessible. In such cases, a reader can be said to be tasked with representing two speech acts, one by the present speaker, and the reported act contained within it. However, in contexts with negation (27), modality (28), or further embedding (29), or within a question (30), the inference of the actual existence of this speech act evaporates. In such cases, this second embedded act must be reconstructed to understand the

sentence, but it would be a mistake to understand it as an actual speech act akin to the untarnished status of ARCs in these contexts.

This is because the verbal content of DD is thematically integrated with the verb of speech, and is thus at that level part of the matrix speech act. The reasoning about comparability with ARCs above mostly targeted the propositional content of the DD report, but perhaps this is critically not the relevant contributor to the effect observed in ARCs. That is, if the verbal content of the DD is undoubtedly part of the matrix speech act, even if its ultimate composition is not, we could imagine that this material is not downweighted in the same manner as the fully-independent material of an ARC.

Again, though, this is somewhat surprising, given that the verbal content of DD is apparently attributed and read accordingly very quickly. Even if the attributed utterance is part of the matrix speech act grammatically, it appears to be read as if fully independent.¹¹

If we did conclude that DD is not an appropriate instance of an independent, subordinate speech act, the hypothesis might stand as it does now, but we would not have a good way of evaluating it further with cases of sentence-internal secondary speech acts.

There is another way in which we might see DD as non-comparable to ARCs: the direction of information priority. The information conveyed by an ARC is clearly subordinate to the gist of the matrix sentence, but, despite the linguistic

¹¹I suppose one could, however, analyze the speed-modulation findings as something to do with imagining how a matrix-level speaker would represent that speech, rather than imagining the reported speech qua speech.

tests which reveal it as not-at-issue, it seems as if the information conveyed by DD is the most important part of many sentences which contain it. If this characterization were reliable, one might imagine that, even if it did represent the right kind of speech act unit to be weighted differently, DD would not be downweighted and the matrix would be, or indeed that DD would be upweighted.

From the evidence seen in Experiment 1, I think it is safe to conclude that DD is not upweighted compared to ID: there is no significant effect of structure on the effect of complexity, positive or negative. This still leaves open the possibility for a future experiment to investigate whether complexity in the matrix is treated differently between DD and ID.

However, we might a priori have some doubts about this comparison. In fact, the potential “main point” status of DD is a pattern most clearly demonstrated by ID: the original argument by Simons (2007) shows that in ID and other similar embedding predicates, the main at-issue content can sometimes be the embedded clause, and the matrix serves only as evidential color. This is exactly the behavior that DD does not demonstrate. It would seem that, if the effect of this kind of “main point” phenomenon were to be investigated, ID is not the ideal comparison, or if it is, we need a new rigorous way to define “main point” to the exclusion of at-issueness.

7.2 On the prosodic boundaries of DD

The intuition that implicit prosody may be critical here has arisen in Dillon et al. (2014) and most rigorously in Kroll and Wagers (2019). The argument is that,

rather than the speech act, it may be some unit of prosody roughly corresponding to a speech act that underlies the patterns of downweighting. ARCs and not RRCs ought to feature a difference in prosody such that ARCs comprise such a distinct unit and RRCs do not. This seems intuitively plausible.

Such a hypothesis doesn't require giving the processor knowledge of speech acts, and instead requires use of implicit prosody (Fodor, 2002). This may not be as controversial: psycholinguistic literature has demonstrated that even the default prosody readers assign to a string impacts their online parsing decisions (see Breen (2014) for a recent overview), and the idea that prosodic units are minimal areas of integration (see my earlier suggestion re: speech acts) has also been floated before (Frazier et al., 2006; Hirotsu et al., 2006).

If this is indeed to be a plausible explanation for the ARC and DD results together, it would need to be the case that DD does not exemplify the right kind of prosodic unit to be separate from the matrix clause. While no production study exists, to my knowledge, that compares the prosodification of minimally different strings of ID and DD, there is a large corpus literature comparing productions. The take-aways of this literature are somewhat mixed. Certainly, it is the case that recorded productions of DD regularly feature marked excursions from a speaker's typical speech rate and pitch range (Blackwell et al., 2015; Bolden, 2004; Estelles-Arguedas, 2015; Jansen et al., 2001; Kewitz & Couper-Kuhlen, 1999; Lampert, 2018), though such excursions are neither mandatory nor limited only to DD, for instance often showing up in vivid reporting of ID. A few studies have also found the use of more standard prosodic boundaries to mark the beginning and ending

of DD (Jansen et al., 2001), but other studies have remarked that such boundaries are not always clearly marked and can lead to ambiguities (Bolden, 2004; Hanote, 2015). In particular, Hanote (2015) notes that speakers in a corpus of English regularly produce at least short matrix clauses as pre-phrasal material adjoined to the prosodic unit of the DD itself. Perhaps there is something to be said about the lack of a systematic boundary found by this literature, although it must be compared to the boundary found in ARCs. If indeed DD is prosodically phrased with its matrix (pitch and rate excursions aside), this hypothesis would be quite tenable.

7.3 Conclusions

I have entertained the hypothesis in this thesis that the human processing of language may have knowledge of certain discourse units at its disposal, and may upweight and downweight these units for any given computational or memory task. The data on appositives from Dillon et al. (2014) and Dillon et al. (2017) are largely in line with such a hypothesis, and in particular a generalization that subordinate speech acts are downweighted, but experiments presented here on the processing and judgement of sentences containing direct and indirect discourse suggest that not all speech acts follow this pattern. I’ve entertained two alternate hypotheses: either direct discourse does not represent the right kind of structure to be a candidate for such a discourse unit, or prosodic units, which could differ between appositives and direct discourse, may be the underlying factor. A third possibility I have not entertained explicitly is that a general pattern shouldn’t be projected at all, and

the Dillon et al. (2014, 2017) results are instances of an effect limited only to appositives. In any event, the data here suggests that such a general hypothesis about the evidential weighting of speech act units is too broad a claim. Future work is tasked with deciding the most reasonable way to restrict it.

Bibliography

- Alexander, J. D., & Nygaard, L. C. (2008). Reading voices and hearing text: Talker-specific auditory imagery in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 446–459. <https://doi.org/10.1037/0096-1523.34.2.446>
- AnderBois, S., Brasoveanu, A., & Henderson, R. (2015). At-issue proposals and appositive impositions in discourse. *Journal of Semantics*, 32, 93–138. <https://doi.org/10.1093/jos/fft014>
- Bary, C., & Maier, E. (2018). *The landscape of speech reporting* (Manuscript).
- Blackwell, N. L., Perlman, M., & Fox Tree, J. E. (2015). Quotation as a multimodal construction. *Journal of Pragmatics*, 81, 1–7. <https://doi.org/10.1016/j.pragma.2015.03.004>
- Bolden, G. (2004). The quote and beyond: Defining boundaries of reported speech in conversational Russian. *Journal of Pragmatics*, 36, 1071–1118. <https://doi.org/10.1016/j.pragma.2003.10.015>
- Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass*, 8(2), 37–50.

- Cappelen, H., & Lepore, E. (2017). Quotation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017 ed.). Stanford University.
- Clark, H. H., & Gerrig, R. T. (1990). Quotations as demonstrations. *Language*, 66, 764–805. <https://doi.org/10.2307/414729>
- Davidson, K. (2015). Quotation, demonstration, and iconicity. *Linguistics and Philosophy*, 38, 477–520. <https://doi.org/10.1007/s10988-015-9180-1>
- de Vries, M. (2006). The syntax of appositive relativization: On specifying coordination, false free relatives, and promotion. *Linguistic Inquiry*, 37(2), 229–270.
- Dillon, B., Clifton, C., Jr, & Frazier, L. (2014). Pushed aside: Parentheticals, memory and processing. *Language, Cognition and Neuroscience*, 29(4), 483–498. <https://doi.org/10.1080/01690965.2013.866684>
- Dillon, B., Clifton, C., Jr, Sloggett, S., & Frazier, L. (2017). Appositives and their aftermath: Interference depends on at-issue vs. not-at-issue status. *Journal of Memory and Language*, 96, 93–109. <https://doi.org/10.1016/j.jml.2017.04.008>
- Dillon, B., Frazier, L., & Clifton, C. (2018). No longer an orphan: Evidence for appositive attachment from sentence comprehension. *Glossa*, 3(1), 1–22.
- Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, 69, 85–103. <https://doi.org/10.1016/j.jml.2013.04.003>

- Estelles-Arguedas, M. (2015). Expressing evidentiality through prosody? Prosodic voicing in reported speech in Spanish colloquial conversations. *Journal of Pragmatics*, 85, 138–154. <https://doi.org/10.1016/j.pragma.2015.04.012>
- Farkas, D. F., & Bruce, K. B. (2010). On reacting to assertions and polar questions. *Journal of Semantics*, 27, 81–118. <https://doi.org/10.1093/jos/ffp010>
- Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody 2002*.
- Frazier, L., Carlson, K., & Clifton, J., Charles. (2006). Prosodic phrasing is central to language comprehension. *TRENDS in Cognitive Sciences*, 10(6), 244–249. <https://doi.org/10.1016/j.tics.2006.04.002>
- Hanote, S. (2015). Are reporting clauses special cases of parentheticals? In S. Schneider, J. Glikman, & M. Avanzi (Eds.), *Parenthetical verbs* (pp. 257–285). De Gruyter. <https://doi.org/10.1515/9783110376142-011>
- Hirofani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, 54, 425–443.
- Jackendoff, R. (1977). *X'-syntax: A study of phrase structure*. MIT Press.
- Jansen, W., Gregory, M. L., & Brenier, J. M. (2001). Prosodic correlates of directly reported speech: Evidence from conversational speech. In *Prosody in speech recognition and understanding*. International Speech Communication Association.

- Klewitz, G., & Couper-Kuhlen, E. (1999). Quote-unquote? The role of prosody in the contextualization of reported speech sequences. *Pragmatics*, 9(4), 459–485. <https://doi.org/10.1075/prag.9.4.03kle>
- Koev, T. K. (2013). *Apposition and the structure of discourse* [Doctoral dissertation]. Rutgers University.
- Kroll, M., & Wagers, M. (2019). *Working memory resource allocation is not modulated by clausal discourse status* [Unpublished manuscript]. Department of Linguistics, UC Santa Cruz.
- Lampert, M. (2018). “Speaking” quotation marks: Toward a multimodal analysis of quoting verbatim in English. Peter Lang. <https://doi.org/10.3726/b13406>
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 375–419.
- McCawley, J. D. (1988). *The syntactic phenomena of English* (2nd). University of Chicago Press.
- Murray, S. E. (2014). Varieties of update. *Semantics and Pragmatics*, 7, 2. <https://doi.org/10.3765/sp.7.2>
- Potts, C. (2005). *The logic of conventional implicatures*. Oxford University Press.
- Potts, C. (2012). Conventional implicatures and expressive content. In C. Maienborn, K. von Stechow, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (pp. 2516–2536). De Gruyter. <https://doi.org/10.1515/9783110589849-017>
- Ross, J. R. (1967). *Constraints on variables in syntax* [Doctoral dissertation]. MIT.

- Simons, M. (2007). Observations on embedding verbs, evidentiality, and presupposition. *Lingua*, 117, 1034–1056. <https://doi.org/10.1016/j.lingua.2006.05.006>
- Stites, M. C., Luke, S. G., & Christianson, K. (2013). The psychologist said quickly, “Dialogue descriptions modulate reading speed!” *Memory and Cognition*, 41, 137–151. <https://doi.org/10.3758/s13421-012-0248-7>
- Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, 1(3), 227–245. <https://doi.org/10.1080/01690968608407062>
- Tonhauser, J. (2012). Diagnosing (not-)at-issue content. In *Proceedings of SULA 6* (pp. 239–254).
- Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35, 454–475. <https://doi.org/10.1006/jmla.1996.0025>
- Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55, 157–166.
- Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience*, 23(10), 3146–3152. https://doi.org/10.1162/jocn_a_00022
- Yao, B., & Scheepers, C. (2011). Contextual modulation of reading rate for direct versus indirect speech quotations. *Cognition*, 121, 447–453. <https://doi.org/10.1016/j.cognition.2011.08.007>

Zhou, P., & Christianson, K. (2016). I “hear” what you’re “saying”: Auditory perceptual simulation, reading speed, and reading comprehension. *Quarterly Journal of Experimental Psychology*, 69(5), 972–995. <https://doi.org/10.1080/17470218.2015.1018282>