Title

Forward-looking Effects in Subject Pronoun Interpretation:What Comes Next Matters

Permalink

https://escholarship.org/uc/item/0jj941n5

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Authors

Song, Jina

Kaiser, Elsi

Publication Date

2020

Forward-looking Effects in Subject Pronoun Interpretation:
What Comes Next Matters

Jina Song (jinas@usc.edu)
Elsi Kaiser (emkaiser@usc.edu)
Department of Linguistics; University of Southern California
3601 Watt Way, GFS 301, Los Angeles, CA, 90089-1693, USA

Abstract

We report two experiments investigating how the interpretation of subject-position pronouns is guided by the referential structure of the pronoun-containing clause, and how this information interacts with information available in the clause that precedes the pronoun. Thus, we consider information that is available to the language processing system before the pronoun is encountered (pre-pronominal information), as well as information that comes after the pronoun (post-pronominal information). In particular, we test how implicit causality biases of verbs that precede the pronoun-containing clause interact with the referential structure of the pronoun-containing clause, i.e., whether or not the clause with the pronoun contains another ambiguous pronoun. We report two offline studies whose results reveal significant effects of both pre- and post-pronominal referential information on pronoun resolution: In addition to replicating effects of implicit causality biases observed in prior work, we also show that people's referential biases depend on whether the clause contains only a subject-position pronoun or also a second pronoun in object position.

Keywords: Pronouns; reference resolution; anaphor resolution; discourse processing; referential structure

Introduction

The question of pronoun interpretation is a fundamental challenge for theories of communication, and has been investigated from many perspectives (Crawley, Stevenson & Kleinman, 1990; Kehler, Kertz, Rohde, & Elman 2008 among others). Given that pronouns (e.g. *she, he, it*) are referentially underspecified, how do humans nevertheless interpret them so rapidly?

Many researchers assume that pronoun resolution is guided by the notion of salience, and agree that pronouns tend to be interpreted as referring to entities that are highly salient in comprehenders' mental models of discourse. This brings up the question of what influences how salient different entities are? Prior work suggests that multiple factors can contribute to an entity's salience and hence guide interpretation of subsequent pronouns. Strikingly, the factors investigated in most previous psycholinguistic work on pronoun resolution – such as the thematic and syntactic roles of potential antecedents, the linear position and discourse status of potential antecedents – tend to be pre-prenominal information; that is, information available to comprehenders *before* the pronoun is encountered (Caramazza, Grober, Garvey & Yate, 1977; Crawley et al., 1990; McKoon, Greene & Ratcliff, 1993 among many others).

However, less attention has been paid to whether and how post-pronominal factors—information available to the comprehenders *after* the pronoun is encountered—guide pronoun resolution. Although researchers acknowledge that information after the pronoun – in particular semantic cues – influences pronoun interpretation (e.g. Winograd 1972, see also Arnold 1998 and discussion below), the question of how the interpretation of subject-position pronouns is guided by post-pronominal information has not been the focus of systematic psycholinguistic investigation to the extent that pre-pronominal information has been.

Relatedly to our aims in the present paper, in an earlier paper Kaiser (2009) tested the influence of post-pronominal factors, specifically the referential properties of the rest of the sentence. Based on production data from sentence-continuation tasks, Kaiser shows that the interpretation of subject-position, sentence-initial pronouns is modulated by who is (or isn't) mentioned later in the same clause (see also Grosz, Joshi & Weinstein, 1995). Simply put, building on ideas from Grosz et al. (1995), Kaiser concludes that if a previously lower-salience referent is 'promoted' by being mentioned in subject-position with a pronoun, then mention of a previously higher-salience referent later in the same clause is avoided. In other words, if someone has been promoted, don't mention the guy who's been demoted.

In discussing the impact of post-pronominal information, Kaiser connects her work to Centering Theory (Grosz et al. 1995, discussed below). Another theory of pronoun resolution that allows for effects of post-pronominal information is the coherence-based approach (see e.g. Hobbs, 1979; Kehler et al., 2008). According to coherence-based approaches, the coherence relations between sentences (e.g. *result*, *explanation*) are a key driving force behind pronoun interpretation (see also Kehler et al., 2008; Kehler & Rohde, 2013 among others). They argue against views which focus only on surface-level heuristics such as subjecthood. Once we acknowledge the relation between discourse coherence and pronoun resolution, it becomes clear that it is not sufficient to focus only on the pre-pronominal factors.

In addition to the coherence-based approaches pioneered by Hobbs and Kehler, there is another line of research that also allows for effects of post-pronominal factors, namely Centering Theory (e.g., Grosz et al., 1995; Walker, Joshi & Prince 1998). According to Centering Theory, pronouns are resolved so that the transition from one sentence to the next is as coherent as possible. The coherence of transitions depends on (i) whether the *most central entity* from one utterance (the most significant discourse entity under discussion; in essence, *the topic*) is mentioned in the next

utterance and (ii) what its grammatical role is in the current utterance and as well as in the next utterance.

Centering Theory proposes three main types of coherence transition, with different levels of coherence: CONTINUE is the most coherent, RETAIN is the second most coherent and SHIFT is the least coherent. The different coherence levels of the three types of the transitions are related to the relative salience of discourse entities, called *centers*. There are three types of centers: forward-looking centers (Cf), preferred centers (Cp) and backward-looking centers (Cb).

(1) Grammatical obliqueness hierarchy in English
Subjects > Direct Objects > Indirect Objects > Adjuncts

The set of forward-looking centers (set of Cfs) is an ordered *list of all the discourse entities in the current utterance*. Languages can differ in how Cfs are ranked. In English, Cfs are assumed to be ordered by grammatical role, such that the subject (the least oblique element) is ranked the highest (see (1)). The term highest ranked Cf is called the *preferred center (Cp)*. The term *backward center (Cb)* refers to the most central entity that is mentioned in both the previous and the current utterance. The *Cb* is defined as follows: The highest ranked Cf of the preceding utterance *n* that is mentioned in the current utterance *n+1* is the Cb of the current utterance.

Among the three transitions, CONTINUE is the most coherent because the Cb of the current utterance is the same as that of the previous one (or undefined), and is also the Cp of the current utterance. RETAIN is less coherent, because the Cb of the current utterance is different from its Cp, even though the Cb of the current utterance is the same as the previous one (or undefined). We discuss examples in (4-6). Lastly, SHIFT is least coherent because the Cb is different from the Cp in the current utterance and, moreover, the Cb of the previous utterance is changed in the current utterance.

In essence, according to this approach, the less the central entity changes, the more coherent the discourse. Thus, when resolving anaphoric expressions, the interpretation that results in the most coherent transition is preferred over others.

Because Centering transitions consider entities mentioned anywhere in the sentence, it makes predictions relevant for our aims of testing whether the cognitive salience of referents is modulated only by information available *before* the pronoun or also by information available *after* the pronoun.

Research Questions

In the present study, the research questions in (2) are addressed by investigating how a subject-position pronoun in the subsequent clause is resolved in an ambiguous context.

(2) I. Is the interpretation of subject-position pronouns affected by post-pronominal information – specifically, the referential properties of the rest of the clause, as predicted by Centering Theory – in addition to pre-pronominal factors (verb type)?

II. How do pre-pronominal and post-pronominal factors interact in guiding reference resolution?

We test contexts like (3), with two potential antecedents (subject, object) preceding the ambiguous pronoun *he*.

(3) Henry {surprised / respected} Kevin because he *verbed* { Ø / him / Tom}.

As for pre-pronominal factors, we manipulated verb semantics by using **Implicit Causality (IC) verbs** in the first clause. Implicit causality verbs in explanation contexts (e.g. *Henry surprised/respected Kevin <u>because</u> he*) are known to trigger an expectation that the subject pronoun refers to the preceding subject (IC1 verbs, e.g. *surprise*) or to the preceding object (IC2 verbs, e.g. *respect*) (e.g. Caramazza et al., 1977). In our study, to ensure uniformity of thematic roles, all IC1 verbs were Stimulus-Experiencer (SE) verbs, and all IC2 verbs were Experiencer-Stimulus (ES) verbs.

As for post-pronominal factors, we manipulated the **referential structure** of the second (pronoun-containing) clause. The second clause only contained **one pronoun** (intransitive: *he verbed*), a **pronoun and a new name** (transitive: *he verbed Tom*) or **two pronoun**s (transitive: *he verbed him*). Most prior work has focused on subject-position pronouns, so we use the term 'post-pronominal' for information not available until *after the subject-position pronoun.* Crucially, the Two-Pronoun condition is the only one in which both the subject and the object from the preceding sentence are mentioned in the continuation. In the One-Pronoun condition and the Pronoun+Name condition, only one of the referents from the preceding sentence is mentioned in the continuation.

The studies reported here build on Kaiser (2009) but crucially go beyond that earlier work by (i) investigating the interplay between pre-pronominal information in the form of verbs' implicit causality biases (in the preceding clause) and post-pronominal referential information (in the pronoun containing clause) and (ii) by using a comprehension task (rather than a sentence-completion task as Kaiser used) which allows us to directly test people's pronoun interpretation preferences in specific referential configurations (One-Pronoun, Two-Pronoun, Pronoun+Name).

Predictions

We consider three competing predictions. First, if pronoun resolution is **guided only by pre-pronominal factors**, we expect that only the type of verb in the preceding clause will have an effect: Subject-position pronouns should be more likely to be interpreted as referring to object antecedents when preceded by a clause with an IC2 verb than when preceded by a clause with an IC1 verb, regardless of subsequent referential structure (*he verbed/he verbed him/he verbed Tom*). Furthermore, if pronoun resolution is not sensitive to post-pronominal information (here, the referential structure of the pronoun-containing clause), the

strength of the IC verb effect should be equally strong in all three referential structure conditions.

Second, if pronoun resolution is guided *only* by **post-pronominal factors** – in our case, the referential structure of the pronoun-containing clause – we expect to see clear differences between (i) the intransitive One-Pronoun conditions and the transitive Pronoun+Name conditions on the one hand (conditions where only one entity from the preceding sentence is mentioned), and (ii) the transitive Two-Pronoun condition on the other hand (where both entities from the preceding sentence are mentioned).

Specifically, we predict that a subject-position pronoun in the second clause is more likely to refer to the subject of the preceding clause when *both* subject and object of the preceding clause are mentioned (Two-Pronoun condition), compared to when only one of the antecedents is mentioned in the second clause (One-Pronoun/Pronoun+Name). This prediction is derived from a core intuition that is part of Centering Theory. Although our studies should not be construed as an evaluation or test of Centering Theory – which is a rich framework with many more dimensions that we discuss here – we want to acknowledge that Centering Theory provides the foundation for our predictions regarding the effects of referential structure. Let us consider these predictions in more detail:

When both antecedents from the previous clause (4) are mentioned (Two-Pronoun condition), the pronouns can be resolved in two ways: In (5a), *he* refers to the subject *Henry* and *him* refers to the object *Kevin*. In (5b), *he* refers to the object *Kevin* and *him* refers to the subject *Henry*. In Centering terms, the interpretation in (5a) yields a CONTINUE transition, which is preferred over (5b) which is a RETAIN. In (5b), the preferred center of Clause 1 (Cp: *Henry*) is changed in Clause 2 (Cp: *Kevin*), but it is unchanged between clauses in (5a). (The configuration in (5a) also fits with well-known parallelism effects, e.g. Chambers & Smyth, (1998).)

(4) Clause 1
Henry {surprised/respected} Kevin.
[Cf : Henry, Kevin   Cp : Henry   Cb :∅]

(5) Clause 2 – both antecedents are mentioned.
    a. He*Henry* daxed him*Kevin*. – **CONTINUE <= preferred**
    [Cf : Henry, Kevin   Cp : Henry   Cb : Henry]
    b. He*Kevin* daxed him*Henry*. - RETAIN
    [Cf : Henry, Kevin   Cp : Kevin   Cb : Henry]

(6) Clause 2 – only one antecedent is mentioned.
    a. He*Henry* daxed (Tom). - **CONTINUE**
    [Cf : Henry   Cp : Henry   Cb : Henry]
    b. He*Kevin* daxed (Tom). - **CONTINUE**
    [Cf : Kevin   Cp : Kevin   Cb : Kevin]

On the other hand, when only one of the antecedents in the first clause is mentioned in the second clause, as in the Pronoun+Name and One-Pronoun conditions, the two interpretations shown in (6) – where the subject pronoun *he*

refers to either the subject antecedent *Henry* or the object antecedent *Kevin* – are *equally preferred*. Since the backward center is the most significant discourse entity under discussion in both the preceding and subsequent clauses, in (6), it can be either the preceding subject or the preceding object depending on which entity is realized in Clause 2. For this reason, the backward center (Cb: *Henry* or *Kevin*) is the same as the preferred discourse entity of Clause 2 (Cp: *Henry* or *Kevin*) in both interpretations of (6). Thus, we do not predict a bias toward either antecedent for the pronoun *he*.

The third possibility is that pronoun resolution is guided **by both pre-pronominal and post-pronominal factors**. If so, we predict that the IC verb effect (pre-pronominal) will be modulated by referential structure (post-pronominal). We predict that with *IC1 verbs*, there will be a stronger preference to interpret a subject position pronoun as referring to a subject antecedent in the Two-Pronoun condition than in the One-Pronoun or Pronoun+Name conditions, because this yields a coherent interpretation in terms of the event structure of the sentence with the IC1 verbs as well as in terms of the discourse transition (in the case that all of the antecedents are mentioned by the pronouns). However, with *IC2 verbs,* we may find a weaker verb-driven object bias in the Two-Pronoun condition, because it conflicts with the subject bias stemming from the preference for a CONTINUE transition.

Experiment 1

Method

Participants Forty-five native English-speaking adults participated. We excluded five participants because they were not self-identified U.S.-born native speakers of English or did not pass catch trials. All reported normal or corrected-to-normal vision and hearing.

Materials and Design We tested pre- and post-pronominal factors on pronoun resolution with a 2 × 2 design – verb type vs. referential structure type, as in example (7). Pre-pronominally, verb semantics was manipulated by using IC verbs in the first clause. 48 IC verbs were adopted from the verb lists of Hartshorne & Snedeker (2013) and Ferstl et al. (2011) – 24 SE verbs (i.e., IC1) and 24 ES verbs (i.e., IC2). We controlled the strength of each verb's IC bias, so that the degree of bias towards the subject or the object antecedent was between 60% to 80% (SE (IC1) verbs: Sbj-bias M=67.4%, SD=13.6 and ES (IC2) verbs: Obj-bias M=77.2%, SD=10.9). The two clauses in each target item are connected with the connector *because*. This is because prior work showed that the IC verb type effects only occur in explanation frames (Rohde & Kehler, 2008 inter alia).

Post-pronominally, we manipulated referential structure as mentioned above. We tested the configurations in (7): The second clause is transitive and contains two pronouns (7b, Two-Pronoun condition) or the second clause is intransitive and contains only one pronoun (7a, One-Pronoun condition).

In the second clause, we used 24 different nonce verbs (e.g. *daxed, zoobed, frobbed)* in target items to exclude additional effects of verbal semantics. Each nonce verb was only used

once. Our decision to use nonce words was partly inspired by Hartshorne & Snedeker's (2013) influential paper on IC verb biases, which used the nonce noun *dax* and successfully obtained meaningful results. (As we discuss below, our findings corroborate that meaningful results about pronoun interpretation patterns can be obtained using nonce words.)

(7) a. Henry {surprised (IC1) / respected (IC2)} Kevin because he daxed him.   [Two-Pronoun]
   b. Henry {surprised (IC1) / respected (IC2)} Kevin because he daxed.   [One-Pronoun]

This study had 24 targets and 36 fillers. Targets and fillers were intermixed and presented using a Latin-Square design.

Procedure We developed a picture-writing task where participants typed in the names of the characters involved in an event depicted with a picture. We conducted the study via a web-based interface (Qualtrics and Amazon Mechanical Turk). Every item consisted of simultaneous presentation of (i) the critical sentence and (ii) a schematized picture that depicts the event of the underlined part of the sentence (Figure 1). In targets, the second clause was underlined.
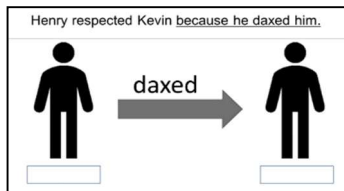

Figure 1a: Sample Two-Pronoun stimulus.   Figure 1b: Sample One-Pronoun stimulus.

Target pictures consisted of one or two person-shaped stick figures that represented the event participants, and a nonce word in the middle that represented the action denoted by the verb of the second clause. For pictures with two stick figures, participants were told that the figure at the origin of the arrow is the initiator of the action (e.g., Agent), and the one at the end of the arrow undergoes the action (e.g., Theme). Thus, the nonce verb expresses a transitive action. In pictures with only one stick figure, participants were told that the stick figure is the person doing the action (e.g., Agent), so the nonce verb expresses an intransitive action.

Participants were instructed to type the name of each person in the text box under each figure so that the picture matches the underlined part of the sentence. This reveals how people interpret the pronouns, e.g. whether they interpret the subject-position pronoun as referring to the subject or object.

Results and Discussion

Figure 2 shows the proportion of trials on which participants interpreted the subject-position pronoun as referring to the preceding object (in all conditions averaged over participants), as indicated by the names they typed in the

boxes. The labels below each bar indicate referential structure type of the second clause and the verb type of the first clause.
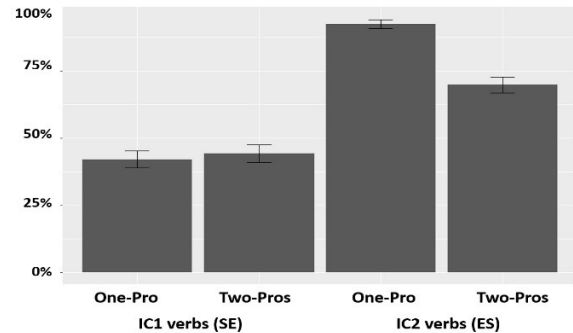

Figure 2 (Experiment 1): Proportion of trials where the subject-position pronoun refers to the preceding object.

Let us first consider whether the *pre-pronominal factor* (IC verb type) influences interpretation of the following subject-position pronoun. As Figure 2 shows, we find a clear IC verb effect in the predicted direction: IC2 verbs markedly increase the proportion of object interpretations compared to the IC1 verbs. In the One-Pronoun condition, participants interpret the subject position pronoun as referring to the preceding object 92.1% of the time with the IC2 verbs, but 42.1% of the time with the IC1 verbs. In the Two-Pronoun condition, participants interpret the subject position pronoun as referring to the preceding object 69.6% of the time with the IC2 verbs but 43.3% with the IC1 verbs.

Let us now consider whether pronoun resolution is also guided by the *post-pronominal factor* (referential structure). As can be seen in Figure 2, referential structure has an effect with IC2 verbs but not with IC1 verbs. With IC2 verbs, the presence of a second pronoun in object position (Two-Pronoun condition) markedly decreases the proportion of object interpretations (69.6%), relative to the One-Pronoun condition (92.1%). However, with IC1 verbs, the presence of an object-position pronoun does not decrease the proportion of object interpretations relative to the One-Pronoun condition (43.3% vs 42.1%). In other words, IC1 verbs pattern the same way regardless of subsequent referential structure, but (unexpectedly) IC2 verbs do not.

To assess these effects statistically, we used a generalized linear mixed logit model (glmer) using R (R Development Core Team, 2017). We find main effects of IC verb type ($z=4.27$, $p<.001$,) and referential structure type ($z=11.81$, $p<.001$). This result confirms that both pre-pronominal verb semantics and post-pronominal referential structure information guide pronoun resolution. Moreover, we found a Verb type × Referential structure type interaction ($z=0.38$, $p<.001$), indicating that the referential structure effect (i.e., stronger subject preference with an object-position pronoun than without) was stronger with IC2 verbs than IC1 verbs. This is confirmed by planned comparisons on the proportion of object choices, which reveal a significant effect of referential structure with IC2 verbs ($z=6.08$, $p<.001$) but not with IC1 verbs ($z=-0.78$, $p>0.4$).

Taken together, these results suggest that reference resolution is guided both by pre-pronominal verb type cues (in the preceding clause) and by differences in referential structure that occur *after* the subject-position pronoun – but effects of referential structure only arise with IC1 verbs, not IC2 verbs. We return to this in the general discussion.

However, the design of Experiment 1 contains a potential confound: the Two-Pronoun condition has *transitive* nonce verbs in the second clause, while the One-pronoun condition uses *intransitive* nonce verbs. As a result, the differences between the conditions could be due to transitivity rather than the presence/absence of an object-position pronoun. For example, perhaps the effects are related to differences in the semantics of subjects of transitive vs. intransitive verbs (see e.g. Hopper & Thompson, 1980), rather than anything to do with referential structure *per se*. To provide a more direct test of the effects of referential structure, Experiment 2 compared transitive Two-Pronoun clauses (*he verbed him*) to transitive Pronoun+Name clauses (*he verbed Tom*).

Experiment 2

Method

Participants Forty-eight English-speaking adults participated. We excluded four participants because they were not self-identified U.S.-born native speakers of English, did not complete all the questions or consistently gave random answers. Four more participants were excluded to match the number of participants in each list. All participants reported normal or corrected-to-normal vision and hearing.

Materials and Design Experiment 2 used the same materials as Experiment 1, except for the way in which referential structure was manipulated: In Experiment 2, we compared sentences with two pronouns (8a) to sentences with a pronoun and a name (8b), i.e., now both sentences were transitive and differed only in the referential form of the object-position element (pronoun/name). This avoids potential confounds due to verb transitivity. The name always referred to a third person who was not mentioned in the first clause.

(8) a. Henry {surprised (IC1) / respected (IC2)} Kevin because he daxed him.   [Two-Pronoun]
    b. Henry {surprised (IC1) / respected (IC2)} Kevin because he daxed Tom.   [Pronoun+Name]

Procedure We used the same task as in Experiment 1. All targets now depicted two characters (Figure 1a), as all nonce verbs in the second clause were interpersonal transitive verbs.

Results and Discussion

Figure 3 shows show the average proportion of trials on which the subject-position pronoun was interpreted as referring to the preceding object.

We again see that the pre-pronominal factor (IC biases of verbs) had a clear influence on how people interpret the subject-position pronoun, just as in Experiment 1.
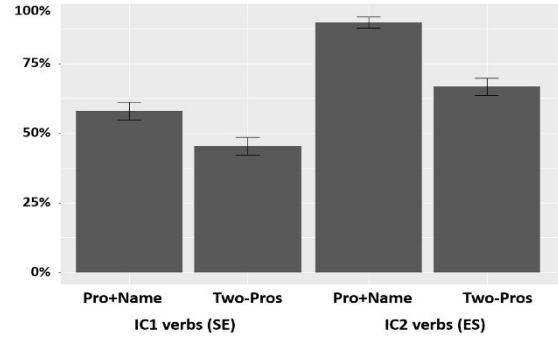

Figure 3 (Experiment 2): Proportion of trials where the subject-position pronoun refers to the preceding object.

As Figure 3 shows, IC2 verbs markedly increase the proportion of object interpretations compared to IC1 verbs. In the Pronoun+Name condition, the subject position pronoun was interpreted as referring to the preceding object 87.5% of the time with IC2 verbs, but 56.7% of the time with IC1 verbs. In the Two-Pronoun condition, the subject position pronoun was interpreted as referring to the preceding object 66.7% of the time with IC2 verbs but 45.4% of the time with IC1 verbs.

The post-pronominal factor (referential structure) has a more uniform effect in Experiment 2 than Experiment 1. As shown in Figure 3, the presence of an anaphoric expression in the object position (Two-Pronoun condition) in the second clause decreases the proportion of object interpretations with both IC1 and IC2 verbs, compared to the configuration with a (non-anaphoric) name in object position (Pronoun+Name condition): With IC1 verbs, Two-Pronoun conditions elicit 45.5% object interpretations and Pronoun+Name conditions elicit 56.7% object interpretations. With IC2 verbs, Two-Pronoun conditions elicit 66.7% object interpretations, and Pronoun+Name conditions elicit 87.5%.

To assess the statistical significance of these patterns, a generalized linear mixed logit model was used (glmer, R). There were statistically significant main effects of IC verb type ($z=-9.18$, $p<.001$), and referential structure type ($z=-6.07$, $p<.001$). This result confirms that both the *pre*-pronominal verb semantics and the *post*-pronominal referential structure information guide the interpretation of subject-position pronouns. Moreover, we found a statistically significant Verb type × Referential structure type interaction ($z=-3.02$, $p<.01$), indicating that the referential structure effect (i.e., a stronger subject preference with an object-position pronoun than without it) was stronger with IC2 verbs than with IC1 verbs. Nevertheless, though the effect is stronger with IC2 verbs, planned comparisons show that both IC1 and IC2 verbs show significant effects of referential structure (IC1: $z=-2.76$, $p<.01$, IC2: $z=-6.15$, $p<.001$).

Taken together, the results of Experiment 2 echo Experiment 1 in revealing effects of both pre- and post-pronominal factors. However, unlike Experiment 1 where we saw an unexpected absence of referential structure effects with IC1 verbs, in Experiment 2 we find significant effects of referential structure with *both* verb types, although the effects

are not as strong with IC1 verbs as with IC2 verbs (as shown by the significant verb type x referential structure interaction). We discuss potential reasons for the IC1/IC2 asymmetry in the next section.

Furthermore, Experiment 2 also shows that the referential structure effects in Experiment 1 cannot be reduced to simple transitivity differences, because all conditions in Experiment 2 used transitive verbs in the second clause.

General Discussion

Pronoun interpretation is known to be guided by information available to the language processing system *before* the pronoun is encountered. We conducted two offline studies to shed light on how pronoun resolution is guided by information available *after* a subject-position pronoun is encountered. Our studies tested effects of post-pronominal information (specifically, referential structure) and its interplay with pre-pronominal information (verb semantics).

Pre-pronominally, we manipulated verb semantics by using implicit causality (IC) verbs. We found a significant IC verb effect in both in Experiments 1 and 2: Subject-position pronouns showed a stronger preference for object antecedents in conditions that used IC2 verbs in the preceding clause as compared to IC1 verbs. This fits with previous work on IC verbs (e.g., Caramazza et al. 1977; McKoon et al. 1993; Ferstl et al., 2011, Hartshorne & Snedeker, 2013).

These results confirm that (i) the picture-writing task we developed works as expected and (ii) corroborate prior findings that humans use pre-pronominal information (here, the verb in the preceding clause) for pronoun resolution. Thus, although use of nonce verbs in the second clause (intentionally) strips away part of the semantics of the sentence and may strike people as unusual, we still see the predicted effects of verb implicit causality in the preceding clause. Our results regarding verb IC biases replicate what other researchers have found with a variety of other methods. We take this as evidence that interpretable results can be obtained in an untimed task using nonce verbs (see also Hartshorne & Snedeker 2013 for other work on pronoun resolution using nonce words).

Crucially, our results also show that, in addition to information from the *preceding* clause, properties of the *pronoun-containing clause itself* – specifically, the referential structure of the clause beyond the subject-position pronoun – guides pronoun resolution. We compared (i) configurations where the critical pronoun-containing clause only refers back to one of the preceding referents (*Henry respected/surprised Kevin because he daxed* in Experiment 1 or *...he daxed Tom* in Experiment 2) to (ii) configurations where the critical pronoun-containing sentence contains two pronouns and thus mentions both preceding referents (*Henry respected/surprised Kevin because he daxed him*).[1] Thus, the referential structure of the remaining material that follows the subject-position pronoun differs in these two configurations; hence, we call it post-pronominal information.

We find effects of the pronoun-containing clause's overall referential structure on pronoun resolution with IC2 verbs in both experiments, and with IC1 verbs as well in Experiment 2: A subject-position pronoun is more likely to be interpreted as referring to a subject antecedent when the subject-position pronoun co-occurs with an object-position pronoun in the same clause (*he daxed him*) than when the pronoun co-occurs with a name (*he daxed Tom*) or in an intransitive (*he daxed*).

Our research is not intended to be a direct or exhaustive test of Centering Theory (e.g., Grosz et al., 1995; Walker et al., 1998), which is a rich framework that goes far beyond the topics we touch upon here. However, our results are compatible with a foundational intuition that underlies Centering Theory – namely, that the referential properties of the entire clause play a role in discourse coherence and reference resolution. Let's recap this in more detail:

Recall that in Two-Pronoun conditions (e.g. *he daxed him*), both the subject and object of the preceding clause are mentioned. Here, interpreting the subject-position pronoun as referring to the preceding object results in an interpretation that *demotes* the preceding subject to the less privileged object position in the current clause while *promoting* the preceding object to the subject position of the current clause – a RETAIN transition which is less coherent than a CONTINUE transition, where the subject pronoun refers to the preceding subject and the object pronoun to the object.

In contrast, if the pronoun-containing clause only has one pronoun (intransitive: *he daxed* or transitive with a (non-)anaphoric name: *he daxed Tom*), then according to Centering Theory, the subject pronoun can be interpreted as referring to *either* the preceding subject *or* object with no effect on the coherence of the transition. Promoting the preceding object to the subject position (pronoun refers to object) or maintaining the preceding subject in the privileged subject position (pronoun refers to subject) both yield a highly coherent CONTINUE transition in Centering terms.

Our finding that, with IC2 verbs in both experiments and IC1 verbs in Experiment 2, presence of an object pronoun (*he daxed him*) boosts the likelihood of *he* referring to the subject – or, conversely, that absence of an object pronoun boosts the likelihood of *he* referring to the object – shows that comprehenders actively utilize referential information that occurs later in the clause, after the subject pronoun, and prefer to make discourse transitions as coherent as possible. These results are compatible with Kaiser's (2009) observation – based on production data, with a different class of verbs, and no explicit referential structure manipulation – that 'promoting' a preceding *object* by realizing it in subject position with a pronoun makes subsequent mention of the preceding *subject* in the same clause dispreferred.

However, our results also suggest that effects of IC verb cues in the *preceding* clause are stronger than effects of referential structure in the *pronoun-containing* clause. There

---

[1] In principle, the pronouns could refer to some other, previously unmentioned person. However, participants' responses make it clear that the pronouns are interpreted as referring to one of the two mentioned and depicted characters.

are at least two possible reasons for this. *First*, the asymmetry may stem from the inherent robustness of implicit causality effects (e.g. IC biases can overcome the default subject bias, as found by Hartshorne & Snedeker, 2013 and others). It could simply be that IC verb biases have an inherently stronger effect than referential structure. In this case, post-pronominal referential structure could turn out to have a stronger effect when compared to other (weaker) pre-pronominal cues. *Second*, it could be that pre-pronominal information, present in the clause that precedes the pronoun-containing clause, is privileged in guiding pronoun resolution due to the incremental nature of language processing: Other things being equal, pre-pronominal information could consistently have a stronger effect than post-pronominal information. If a pronoun is first interpreted based on pre-pronominal information, then when post-pronominal information becomes available, the previous co-referential link may need to be revised. However, it is reasonable to assume that such revision is cognitively costly/effortful. Under this view, pre-pronominal factors (as long as they are reliable enough to be considered by the processor) are inherently more influential than post-pronominal factors. Further work is needed on this.

Finally, let us consider the difference between IC1 and IC2 verbs. Contrary to our expectations, we only found effects of referential structure with IC1 verbs in Experiment 2, whereas IC2 verb conditions show effects of referential structure in both experiments. Furthermore, even when we *do* see an effect of referential structure with IC1 verbs (Experiment 2), the effect is weaker than with IC2 verbs. Thus, overall, it appears that pronoun resolutions in configurations where the first clause contains an IC1 verb is less susceptible to effects of referential structure than with IC2 verbs.

This may be due to recency effects combined with the slightly weaker bias of our IC1 verbs relative to our IC2 verbs. If it's the case that more recently-mentioned referents (e.g. objects) are more 'in focus' than less recently-mentioned referents, this kind of recency effect could weaken the subject bias stemming from IC1 verbs and boost the object bias of IC2 verbs – this could be a task-based effect. This may have obscured potential referential structure effects with IC1 verbs in Experiment 1 and weakened them in Experiment 2. However, the absence of *any* effects of referential structure with IC1 verbs in Experiment 1 brings up the possibility that this may partly also stem from the semantics of intransitive subjects and subjects with the thematic role of stimulus (IC1 verbs). In future work, we aim to further investigate the verb effect asymmetries. We also plan to use real-time methods to directly assess the incremental use of post-pronominal information.

Our results provide novel evidence that pronoun resolution is guided by both pre-pronominal and post-pronominal factors. Empirically and methodologically, our results highlight the importance of considering both what came before the pronoun and what comes after. These findings are compatible with approaches such as Centering Theory and coherence-based accounts.

References

Arnold, J. (1998). *Reference form and discourse patterns*. Doctoral dissertation, Stanford University.

Caramazza, A., Grober, E., Garvey, C., & Yates, J. (1977). Comprehension of anaphoric pronouns. *Journal of Verbal Learning and Verbal Behaviour*, 16, 601-609.

Chambers, C. G., & Smyth, R. (1998). Structural parallelism and discourse coherence: A test of centering theory. *Journal of Memory and Language*, 39(4), 593-608.

Crawley, A., Stevenson, J., & Kleinman, D. (1990). The use of Heuristic Strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, 19, 245-264.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21, 203–225.

Ferstl, E. C., Garnham, A., & Manouilidou, C. (2011). Implicit causality bias in English: A corpus of 300 verbs. *Behavior Research Methods,* 43(1), 124-135.

Hartshorne, J. K., & Snedeker, J. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes*, *28*(10), 1474-1508.

Hobbs, J. (1979). Coherence and coreference. *Cognitive Science*, 3, 67-90.

Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *language*, 251-299.

Kaiser, E. (2009). Investigating effects of structural and information-structural factors on pronoun resolution. In M. Zimmermann and C. Féry (eds.), *Information Structure from Different Perspectives*. NY: Oxford University Press, 332-353.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. (2008). Coherence and Coreference Revisited. *Journal of Semantics,* 25, 1-44.

Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics,* 39(1-2), 1-37.

McKoon, G., Greene, S. B., & Ratcliff, R. (1993). Discourse models, pronoun resolution, and the implicit causality of verbs. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 19(5), 1040.

R Core Team. (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rohde, H., & Kehler, A. (2008). The bidirectional influence between coherence establishment and pronoun interpretation. *Poster presented at the 21st Annual CUNY Conference on Human Sentence Processing.*

Stevenson, R., Nelson, A., and Stenning, K. (1995). The role of parallelism in strategies of pronoun comprehension. *Language and Speech*, 38, 393-418.

Walker, M. A., Joshi, A. K., & Prince, E. F. (1998). Centering in naturally occurring discourse: An overview. *Centering theory in discourse*, 128.

Winograd, T. (1972). Understanding natural language. *Cognitive psychology*, *3*(1), 1-191.