THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Pronouns

Citation for published version:
Rohde, H 2019, Pronouns. in C Cummins & N Katsos (eds), *The Oxford Handbook of Experimental Semantics and Pragmatics.* Oxford Handbooks, Oxford University Press, Oxford, pp. 452-473. https://doi.org/10.1093/oxfordhb/9780198791768.013.21

Digital Object Identifier (DOI):
10.1093/oxfordhb/9780198791768.013.21

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Early version, also known as pre-print

Published In:
The Oxford Handbook of Experimental Semantics and Pragmatics

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 30. Jul. 2023

Pronoun interpretation and production

Hannah Rohde[a]

[a] *University of Edinburgh, Department of Linguistics & English Language, Edinburgh, UK*

Why do pronouns matter to the study of semantics and pragmatics? A long-standing answer to that question has been that pronouns are the fruit flies of discourse – that their interpretation reflects the deeper semantic meaning of a passage and can thus provide a window into the forces at play in establishing meaning across clauses. This view has motivated much of the experimental work on pronoun interpretation and production.

What that work has revealed, however, is that the story is in fact more complex: Pronouns are not transparent windows into underlying discourse structure and meaning; rather the pronominal form itself imposes unique and important pragmatic constraints on what speakers and comprehenders do with pronouns. These two viewpoints on the role of pronouns—as fruit flies to deeper meaning and as linguistic elements with independent effects—are illustrated below with two well-known examples from the pronoun literature. Through these examples and a discussion of a number of experimental studies, this chapter presents a history of the pronoun puzzle. It then introduces a recent attempt to reconcile competing approaches by situating pronouns within a more general model of pragmatic communication.

To start, consider examples (1a-b). These examples shows how ambiguous pronouns receive their interpretation via the general reasoning that comprehenders use in order to make a discourse make sense:

(1) *Effects of general reasoning in pronoun interpretation (from Winograd, 1972)*
   a. The city council denied the demonstrators a permit because they feared violence.
   b. The city council denied the demonstrators a permit because they advocated violence.

A comprehender who encounters (1a) or (1b) must figure out how the speaker intended the clauses to relate. For (1a), doing so requires reasoning about how the denial of a permit could be linked to fear and identifying which event participant's fear could most plausibly account for that denial. The establishment of a coherent discourse thus yields an interpretation of the pronoun *they* as referring to the city council. A similar process for (1b) yields a different interpretation: It is not clear how a preference for violence from the city council would result

in the demonstrators being denied a permit, but if the pronoun is taken to refer to the demonstrators, then it is possible to reason how demonstrators who advocate violence could be worrisome to the council and how that worry could lead to the denial of a permit.

In these cases, it seems that the pronoun itself dictates little; it is the reasoning in pursuit of coherence that determines the pronoun's referent. An approach which takes into account the role of coherence thus assigns pronouns an epiphenomenal status: Their meaning is cast simply as a side effect of the comprehender's process of establishing a coherent discourse (Hobbs, 1979; Kehler, 2002; Hobbs et al., 1993; Rohde et al., 2007). But that approach is not the only one, and in the end, it leaves unexplained certain interesting facts about pronoun use.

If pronouns are simply unbound variables whose meaning is assigned in the course of general discourse processing (Hobbs, 1979, *inter alia*), one might expect processing a name to be *simpler* than processing a pronoun. After all, a name can pick out a unique referent whereas a pronoun may require additional reasoning to resolve. On the other hand, if the pronominal form itself dictates the contexts in which it is more or less appropriate, using a name in such a context could induce a penalty. Indeed, models like Centering Theory (Grosz et al., 1995) encode a preference for speakers to use pronouns to refer back to the most prominent referent in the discourse, where prominence is linked to a referent's syntactic position in the previous utterance. The two passages in (2a-b) represent materials from a reading time study comparing the processing of pronouns and names when the antecedent is the previous syntactic subject (Gordon et al., 1993).

(2) *Comparison of pronouns and names: the Repeated Name Penalty (Gordon et al., 1993)*

a. *Pronoun-Name Condition*
Bruno was the bully of the neighborhood.
He chased Tommy all the way home from school one day.
He watched Tommy hide behind a big tree and start to cry.
He yelled at Tommy so loudly that all the neighbors came outside.

b. *Name-Name Condition*
Bruno was the bully of the neighborhood.
Bruno chased Tommy all the way home from school one day.
Bruno watched Tommy hide behind a big tree and start to cry.
Bruno yelled at Tommy so loudly that all the neighbors came outside.

The results of the study showed that pronouns do not necessarily give rise to comprehension difficulty; rather, the passage containing a number of ambiguous pronouns (2a) was read faster than the version with only unambiguous names (2b). This suggests that there may be contexts in which pronouns are expected over names and in which comprehension difficulty emerges not because of the

content of the passage's message, but the form that that message takes. An approach which takes into account a passage's surface form (its structure and word choices) thus is concerned with a speaker's decision of how to realize an intended message.

Reconciling the facts illustrated in (1-2) requires understanding when and why pronouns are used in discourse, questions that have interested researchers from across Linguistics, Philosophy, Psychology, and Computer Science. The facts are further complicated by a number of syntactic, semantic, and pragmatic biases that have been identified in experimental studies on pronoun use. This chapter reviews that work and the questions raised. It then steps back to ask whether it's really possible to understand pronouns by looking at pronouns themselves. A new approach is laid out that shifts perspective from pronouns specifically to a generative model of language more broadly. Such a model captures the choices speakers make in choosing messages to convey and choosing words with which to convey those messages. In the case of pronouns, a generative model makes an important distinction between the choice of which *referent* a speaker will mention and the *form* that that reference may take, thereby providing an account of the patterns observed in (1-2). In this way, pronouns are still fruit flies, but they serve a new purpose of providing a test case for how generative models can inform research in pragmatics.

1. One pronoun, many factors, different modeling approaches

The history of research on pronouns overwhelmingly targets one particular type—the 3rd person singular pronoun—almost to the exclusion of other pronominal forms.[1] Despite this focus, the resulting picture has not been simple. Many studies have grappled with the question of what a comprehender does when they encounter a pronoun, and a number of different factors have been identified, some of which appear to contradict each other. More recent work manages to resolve some of these discrepancies by showing how other mechanisms active in a discourse may determine the role a given factor plays in a particular context. These puzzles, and hints at directions for reconcilation, are reviewed below.

1.1. Does syntax matter?

To start, an oft-cited constraint on pronoun interpretation involves surface syntactic structure: This is a preference for antecedents that have appeared in SUBJECT position (Frederiksen, 1981; Gernsbacher & Hargreaves, 1988; Kameyama, 1996; Arnold, 1998; Arnold & Lao, 2015; Hartshorne et al.,

[1] The focus in this chapter will be further restricted to work conducted on English 3rd person singular pronouns. Work on other pronominal forms (null pronouns, demonstratives) has been pursued for other languages with different pronominal systems. Indeed pronouns can be used in varied and surprising ways (e.g., *they* in contexts without a readily available plural antecedent, abstract event-referring *this*, etc.), but the 3rd person pronouns *he/she* will be sufficient to get us started. In Section 5 we return briefly to these issues.

2015, see also the interpretation algorithm for Centering Theory, Brennan et al., 1987). For example, the ambiguous pronoun *him* in (3a) is said to resolve preferentially to John, the subject of the context sentence about the hitting event. However, if that same event is described in a passive construction with Bill promoted to subject position, as in (3b), the preference is said to switch to Bill.

(3) *Subjecthood preference (examples from Kameyama (1996))*
a. John hit Bill. Mary told him to go home. [him=John]
b. Bill was hit by John. Mary told him to go home. [him=Bill]

Intuitions about examples like (3a-b) suggest a role for syntax in antecedent selection, as does experimental work (Gernsbacher, 1989; Crawley et al., 1990; Speelman & Kirsner, 1990; Arnold, 2010) and corpus work (Arnold, 1998) that shows preferential resolution to the first-mentioned or subject referent over other referents. The subject preference has been assessed with a variety of procedures: e.g., reading time and timed/untimed antecedent selection tasks.

Another way that syntax has been implicated in pronoun interpretation is via a preference for antecedents in syntactically parallel positions. Antecedent selection tasks and reading time tasks (Smyth, 1994; Chambers & Smyth, 1998; Sheldon, 1974) show a preference for the subject referent in passages like (4a) and the object referent in passages like (4b).

(4) *Parallelism preference*
a. John kicked Bill in the leg. He punched Mary in the arm. [he = subject John]
b. John kicked Bill in the leg. Mary punched him in the arm. [him = object Bill]

Already we have two factors (subjecthood, parallelism) that can conflict: In passages like (4b), what dictates whether the pronoun *him* will be resolved to the subject antecedent (John) or the parallel antecedent (Bill)? One proposal is that discourses in which these two preferences manifest differ in the inference processes underlying coherence establishment (Kehler, 2002). What is notable about passages like (3a-b) and (4a-b) is the way the sentences within them relate to each other: The examples used to illustrate a subject preference describe narrative sequences of events (compatible with the connective *then*), whereas the examples in support of a parallelism preference both convey parallel events (compatible with connectives like *similarly* or *likewise*). In answer to the question *Does syntax matter?*, one can see that the appearance of syntax-based preferences may depend on the way that the sentences combine to form a coherent discourse.

Narration and Parallel relations represent just two out of a larger inventory of possible relations ("coherence relations") that can be inferred to hold between clauses (Asher & Lascarides, 2003; Hobbs, 1979; Kehler, 2002; Mann & Thompson, 1988; Prasad et al., 2008; Sanders et al., 1992). The pronoun literature that targets surface order and syntactic constraints (Smyth, 1994; Crawley

et al., 1990) concedes a role for real-world knowledge and plausibility. However, explicitly testing such a role requires a different tactic of direct assessments of semantic biases and coherence relations (see Stevenson et al., 1994; Crinean & Garnham, 2006; Stewart et al., 1998; Rohde & Kehler, 2014). To understand the role of general reasoning in pronoun interpretation, we turn next to a particular type of semantic bias. This will connect up with the generative model alluded to at the outset of this chapter.

1.2. Does thematic role matter?

In contrast to posited surface structural constraints, biases associated with the lexical semantics of particular verbs have been shown to cut across syntactic categories. Examples (5a-b) contain verbs belonging to the class of so-called implicit causality verbs (Garvey & Caramazza, 1974, *inter alia*).

(5) *Implicit causality bias to causally implicated referent (examples from Caramazza et al., 1977)*
   a. Mary annoyed Sue because she had stolen a tennis racket. [she=Mary]
   b. Mary scolded Sue because she had stolen a tennis racket. [she=Sue]

(5a-b) are minimal pairs, differing only in the first verb. The two referents Mary and Sue appear as subject and object, respectively, in both examples. However, the pronoun *she* in (5a) is interpreted most plausibly to refer to Mary, in (5b) to Sue. This difference is attributed to the way *annoy* and *scold* assign their thematic roles: Mary as the subject of *annoy* in (5a) is the causally implicated stimulus referent and Sue in object position is the experiencer, whereas in (5b), Mary is the agent and it is Sue who is the causally implicated referent. Verbs like these have been shown to induce a pronoun interpretation preference to the causally implicated referent, independent of the syntactic position of that referent. This has been demonstrated using a number of methodologies (for a review, see Koornneef & Sanders, 2013).

For example, in story continuation tasks in which participants are prompted with a passage up to and including the pronoun, the participants show a preference to use the pronoun to refer to the causally implicated referent (e.g., Garvey & Caramazza, 1974; Stevenson et al., 1994; Ferstl et al., 2011; Hartshorne & Snedeker, 2012; Stewart et al., 2000). Online measures show a corresponding pattern: Participants read more quickly when the causally implicated referent is mentioned than an alternate referent (e.g., via a gendered pronoun in a context with one male and one female referent; Koornneef & van Berkum, 2006); in visual-world eyetracking paradigms, participants look anticipatorily to a picture of the causally implicated referent even before hearing a pronoun (Pyykkönen & Järvikivi, 2009).

Thematic role effects can also be seen in transfer-of-possession contexts in which the source and goal thematic roles occupy the subject or non-subject position. Examples (6a-b) show how such sentences have been presented in story-continuation tasks (Stevenson et al., 1994; Arnold, 2001; Kehler et al., 2008). In continuations following both (6a-b), participants use the pronoun to

refer to the goal referent, Bob, as often or more often than the source referent, John.

(6) *Transfer-of-possession contexts with goal bias*
a. John handed a book to Bob. He _____
b. Bob received a book from John. He _____

However, the bias to the causally implicated referent for implicit causality contexts and the bias to the goal for transfer-of-possession contexts are not uniform. When pronoun interpretation patterns are broken down by coherence relation, either by explicitly indicating the relation with a connective (Stevenson et al., 2000; Koornneef & van Berkum, 2006) or by inferring the relation via annotation of participants' continuations (Kehler et al., 2008; Arnold, 2001), subpatterns emerge. The bias to the causally implicated referent is strongest in continuations that describe, unsurprisingly, a cause (Explanation relations); the bias to the goal referent is strongest in relations that describe what happened next (Narration or Result relations), presumably due to the goal's association with the end state of a transfer event. The interpretation of the pronoun is thus conditioned on the operative coherence relation. This can be seen in examples (7a-b) and (8a-b).

(7) *Effect of coherence relation in implicit causality contexts*
a. John amazed Bill. He could walk on his hands. [He=John$_{stimulus}$]
b. John amazed Bill. He clapped his hands in awe. [He=Bill$_{experiencer}$]

(8) *Effect of coherence relation in transfer-of-possession contexts*
a. John handed a book to Bill. He opened it to the first page.[He=Bill$_{goal}$]
b. John handed a book to Bill. He handed him *War and Peace*. [He=John$_{source}$]

The first example in each pair shows coreference in accordance with the oft-reported thematic role biases: the causally implicated referent for Explanations following implicit causality verbs in (7a), the goal for Narration relations following transfer-of-possession verbs in (8a). The second example illustrates alternative coreference biases associated with different coherence relations: the experiencer for a Result relation in (7b) (see "implicit consequentiality", Stewart et al., 1998, Crinean & Garnham, 2006, Pickering & Majid, 2007), the source for an Elaboration relation in (8b).

Early reports of thematic role biases were assessed by collapsing across coherence relations. The picture that emerges when coherence is taken into account is that the purported bias to the causally implicated referent arises from the bias to that referent in Explanations and, crucially, the *frequency* of Explanation relations in implicit causality contexts; similarly, the purported bias to the goal arises from the frequency of Narration/Result relations in transfer-of-possession contexts (Kehler et al., 2008).

What do these observations about thematic role biases and coherence relations mean for a model of pronoun interpretation? If we think of a comprehender's task as one of reverse engineering the speaker's intended message, then

verb class can provide a cue to the message the speaker may be trying to convey. For example, if a speaker describes a situation using an implicit causality verb, the comprehender may expect that the speaker will use the next utterance to provide an explanation and, furthermore, that the causally implicated referent will be mentioned as part of that explanation. In this way, a comprehender can attempt to reconstruct the speaker's intended meaning by considering how the discourse could most plausibly have been generated — which coherence relation is most likely to hold following an implicit causality verb and which referent is most likely to be mentioned given that relation? We return to generative models in section 4. See also Degen and Tanenhaus' chapter (this volume) on the role of prediction in constraint-based pragmatic processing, specifically the relevance of discourse structure, world knowledge, and estimates of the speaker's epistemic state to inform priors over possible messages.

Having identified a role for discourse-level coherence relations in pronoun interpretation, the next two sections consider other ways in which discourse structure and information structure guide pronoun use.

1.3. Does distance matter?

Beyond the short passages discussed thus far, there are claims regarding the interpretation of pronouns in larger discourses—for example, the claim that pronouns favor antecedents that were mentioned RECENTLY. Intuitively, a referent introduced in a previous chapter of this handbook is unlikely to be re-mentioned with a pronoun in the sentence you're currently reading. Indeed, pronouns have been found to yield more processing difficulty if the antecedent is mentioned two or three sentences back than if the antecedent appears in the immediately preceding sentence, as assessed via full-sentence reading times, response times for antecedent selection, and reaction times for judging the familiarity of a probe word related to a potential antecedent (Clark & Sengul, 1979; see also Gernsbacher, 1989; Chang, 1980; Ehrlich & Rayner, 1983; Arnold, 1998).

But how is recency determined? The distance to some referents may be large if one counts words or clauses, but small if the metric reflects the hierarchical combination of sentences in a discourse. Earlier work on discourse parsing suggested that the availability of particular pronoun∼antecedent relationships was constrained by the possible attachment positions of subsequent clauses in an unfolding discourse. The Right Frontier Constraint limits the addition of new clauses to open nodes on the right edge of the discourse structure (Polanyi, 1988; see also Asher, 1993; Asher & Lascarides, 2003; Malt, 1985). Figure 1 illustrates the right frontier of a discourse, representing the sentences of the discourse as nodes in a tree with semantic links shown as branches (e.g., the coherence relation expressed by *because* in (1) would link a parent and daughter node in a tree like Figure 1). When asked to select an antecedent for a pronoun, comprehenders appear sensitive to the role of structure, favoring antecedents in open positions over closed positions, even when the linear distance is greater (Holler & Irmen, 2007).

To illustrate the role of the right frontier in pronoun resolution, consider the pronoun *it* in the last utterance of the dialogue excerpt in (9). Despite

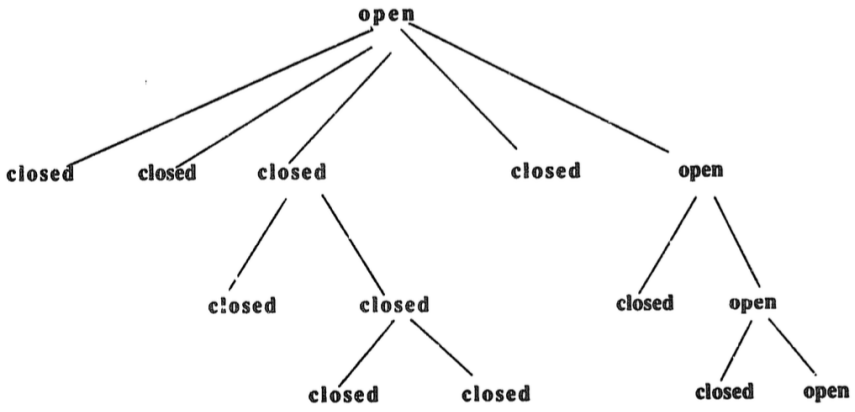

Figure 1: A sample discourse structure containing open positions to which subsequent clauses may attach (from Polanyi, 1988, p.613); pronouns favor antecedents in open positions.

a half dozen intervening utterances and the presence of several other available referents, *it* in (9i) manages to pick out an antecedent (the stuck bolt) mentioned only at the beginning of the dialogue.

(9) *Naturally occurring task-oriented dialogue showing effect of discourse structure on pronoun interpretation (Grosz, 1977). Discourse participants are an Expert (E) and an Apprentice (A) who are working together remotely to disassemble an engine. A camera feed shows the Expert the Apprentice's work space.*

a. A: One bolt$_v$ is stuck.
b. A: I'm trying to use both the pliers$_y$ and the wrench$_z$ to get it$_v$ unstuck, but I haven't had much luck.
c. E: Don't use the pliers$_y$.
d. E: Show me what you're doing.
e. A: I'm pointing at the bolts$_x$.
f. E: Show me the 12" combination wrench$_w$, please.
g. A: OK.
h. E: Good, now show me the 12" box wrench$_u$.
i. A: I already got it$_v$ loosened.

In an analysis of this dialogue and the broader role of intentional structure in discourse, Roberts (2016) identifies an overarching goal shared by the dialogue participants (to disassemble the engine) and a subgoal introduced at (9a) (to remove the stuck bolt). The subsequent utterances (9b-h) address the execution of plans and subplans for achieving the subgoal. In this way, the utterance at (9a) about the problem (*One bolt is stuck*) is an open node in the discourse structure, making it a possible attachment point for the utterance at (9i) about the problem's resolution (*I already got it loosened*). The pronoun in (9i) can thus be resolved to bolt$_v$ as a structurally accessible referent on the right frontier.

1.4. Does topichood matter?

Example (9) demonstrates how recency can depend on the model of the discourse. In a discourse structure, each sentence is said to stand in a coherence relation with prior material (Asher & Lascarides, 2003; Hobbs, 1979; Kehler, 2002; Mann & Thompson, 1988; Prasad et al., 2008; Sanders et al., 1992) or to address a question or goal relevant to the preceding discourse (Roberts, 1996). Inferring discourse goals involves inferring what is at issue, or what is topical. One way of understanding the use of the pronoun *it* in the dialogue in (9) is to consider the information status of its antecedent, bolt$_v$, in the dialogue. In information structural terms, bolt$_v$ is the topic—it is what the discourse is about.

In the context of a dialogue, topichood can be determined by the goals and intentions of the interlocutors. Information status can also be signaled more locally via sentence structure. The subject preference illustrated in examples (3a-b) and the Repeated Name Penalty in (2a-b) may reflect the information status of syntactic subjects. In English, subject position is the default position for topics (Chafe, 1976; Reinhart, 1981; Zubin, 1979). Moreover, being the subject of a passive (as in (3b) *Bill was hit by John*) is an even stronger signal

that that referent is likely to be the topic (Ward & Birner, 2004; Prince, 1985; Shibatani, 1985). Since speakers are known to use pronouns when they maintain the current topic, it seems reasonable that comprehenders ought to take topichood into account when they encounter a pronoun.

The field has long noted this link between topicality and the use of underspecified referential forms like pronouns, with researchers assessing a referent's status via discourse structure and sentence structure as well as measures of cognitive accessibility (Sanford & Garrod, 1981; Ariel, 1990; Gernsbacher, 1990; Givón, 1983; Gundel et al., 1993; Grosz et al., 1995; Prince, 1985; Walker et al., 1994; Cowles & Ferreira, 2012; Arnold & Lao, 2015). The importance of a referent's status in a passage has been confirmed in a range of studies on pronoun interpretation and production. In reading time studies, sentences with pronouns are found to be processed faster when the pronoun refers to a topical antecedent than a non-topical antecedent (as controlled by primacy of mention and scenario dependence; Anderson et al., 1983; see also Garrod et al., 1994) or when the antecedent is introduced with a name instead of a definite noun phrase (perhaps because names are used for principal protagonists; Sanford et al., 1988). In a study of children's storytelling, more sophisticated speakers were found to use pronouns to refer to the protagonist (Karmiloff-Smith, 1985). In adult productions, again the pronoun is used to refer to the main actor (Marslen-Wilson et al., 1982).

Lastly, to see the effects of passivization, consider the examples in (10a-b). These represent stimuli for a set of studies using familiarity with a probe word as a measure of antecedent accessibility (Speelman & Kirsner, 1990).

(10) *Effect of passivization on topicality (examples from Speelman & Kirsner, 1990)*
   a. A nurse threw the shoe over the balcony.
   b. The shoe was thrown by a nurse over the balcony.

Participants in the study read a sentence like (10a) or (10b) and then saw the first word of a subsequent sentence: either a pronoun *She* or another noun phrase *A neighbor*. The task required them to indicate at that point whether a probe word was familiar or not. The findings showed that the probe word *nurse* was recognized more quickly following the pronoun *She* than *A neighbor*, but that this pattern only held for active voice stimuli like that in (10a), not for the passive voice (10b) when the nurse is no longer in a topical position.

For the Repeated Name Penalty and the passivization effects, it is only the sentence's surface structure and form that changes, not the event-level semantics. But in many of the results discussed earlier, surface structure provided no cue to the intended interpretation, which depended instead on semantic biases or general reasoning. The next section summarizes this distinction between models built around meaning and models built around form.

1.5. Prior models of pronoun use

On one hand, a coherence-driven model of pronouns (Hobbs, 1979; Kehler, 2002) highlights the importance of comprehenders' reasoning about the way that

sentences relate. Such a model, however, provides no mechanism to account for behavior that varies with surface form—e.g., alternations that arise with different syntactic structures or different anaphoric forms. This modeling bias may reflect the artificial intelligence work of the time. Computational systems were being built that could reason about the logical form of an utterance and draw conclusions with respect to a knowledge base of axioms about the world (e.g., Hobbs et al., 1993). Sentence form was treated as a separate postscript link between the underlying meaning and the surface realization.

On the other hand, models that take into account the surface forms of the utterances in which a pronoun appears (Grosz et al., 1995) make no effort to incorporate deep semantic reasoning, yet they capture patterns that cannot be ignored—e.g., a link between subject antecedents and pronominal forms. The development of Centering Theory reflects the pressure in computational linguistics to build systems around directly observable features—the theory's definitions and rules make use of grammatical role and referential form, properties of a sentence that are easily discernable from surface input and don't require deeper reasoning. Centering Theory can make predictions across a wide variety of discourse contexts precisely because it doesn't depend on a sophisticated engine for deriving domain-specific meaning. Its effectiveness likely reflects the fact that grammatical role and referential form are symptoms of deeper information structural factors like topichood.

The results of the experimental work reviewed above can be linked to each of these two approaches. The coherence-driven approach accounts for evidence of thematic role biases and some of the interpretations of the syntactic and distance-based metrics. The form-driven approach accounts for topic-driven effects like the subjecthood preference and the Repeated Name Penalty. It is only recent work that has proposed a reconciliation of the two approaches (Kehler & Rohde, 2013). This new work, to be described in what follows, makes an important distinction between factors that influence which referent is the most likely candidate for re-mention and factors that influence how that referent will be mentioned.

A need for such a distinction is apparent in several previous studies. For example, Hudson-D'Zmura & Tanenhaus (1998) analyze implicit causality contexts, specifically those in which the causally implicated referent is mentioned in object position. In a pair of studies, one set of participants wrote story continuations (e.g., *Max despises Ross because he* _____), and another set of participants made speeded sensibility judgments about passages containing either a subject-referring or object-referring pronoun (e.g., *Max despises Ross. He always gives Max/Ross a hard time.*). The results showed an interesting dissociation between the referents favored for mention in story continuations (the causally implicated object) and those whose mention with a pronoun yielded the highest sensibility judgments and shortest decision latencies (the subject). At the level of meaning, participants' story continuations showed that they were attending to the lexical semantics of the verb and the coherence relation signaled by the connective *because*. At the level of surface form, participants' speeded ratings showed they were sensitive to the subjecthood of the antecedent.

An adequate model of pronoun use must therefore capture both the meaning-driven and form-driven patterns. We need to understand which referents are likely to be mentioned and which referents are likely to be mentioned with a pronoun since those probabilities need not be the same. Such a model requires recasting the research question from a pronoun-resolution-specific question ('Given a pronoun, who does it refer to?') to one that models coreference more generally ('Given estimates of likely messages and message forms, how does a comprehender infer what the speaker intended?').

2. Stop looking at pronouns to understand pronouns

A range of different methodologies have been used to understand pronoun interpretation, including reading time (e.g., Wolf et al., 2004), probe-word response (e.g., Greene et al. (1992)), speeded sensibility judgments (e.g., Foraker & McElree, 2007), brain response (e.g., Van Berkum et al., 2007), and eyetracking in the visual world paradigm (e.g., Arnold et al., 2000). This work has tried to understand pronoun interpretation largely by looking at how people interpret pronouns. But therein may lie the mistake. If a pronoun leads to reading time slowdowns, longer response latencies, neuronal activity associated with ambiguity, and delayed looks to a referent in a visual scene, how do we know if that response indicates difficulty with *who* has been mentioned or with *how*? We need to include in our analysis contexts in which the form of reference is not a pronoun (see Almor & Nair, 2007).

Consider the following excerpts from two different speakers' narrations of the same sequence in a video recording of a basketball game (Brennan, 1995).

(11) *Choices in referential form (Brennan, 1995)*
a. Wolverines movin' it slowly up the back court,
over to number...
back to forty-one,
he shoots from three points... no.

b. Number thirty passes it off to forty-one.
Forty-one goes up for the shot, and he misses.

The nature of the events being described (a pass from one player to another followed by an attempted shot) dictate which referent is most likely to be mentioned after the pass. In both speakers' versions, player forty-one is clearly in possession of the ball after the pass. However, despite the high probability that the speaker will generate a statement about player forty-one, the first speaker generates a pronominal reference and the second does not. This example distills the distinction that a generative model must make between the probability of who to mention and how, and that those two probabilities need not be the same. Across (11a-b), the observed probability of starting the last utterance with a mention of forty-one is 1 but the probability of pronominalizing that reference is only .5.

A comprehender encountering the pronoun in the last sentence of (11a) would likely have no trouble resolving it to player forty-one since the event semantics guarantee that the person shooting the ball must be the one who received it on the previous pass. So how do we reconcile the fact that a comprehender would easily interpret a pronoun in this context to player forty-one with the fact that a speaker wouldn't necessarily produce a pronoun for that referent? Why aren't production and interpretation mirror images of each other?

The literature often assumes a direct mapping between production and interpretation. This is apparent in claims that speakers use pronouns to denote referents they believe are associated with a high degree of activation in the cognitive state of the comprehender (e.g., Gundel et al., 1993, see discussions of salience in von Heusinger, 2002; Bach, 1994). The assumption is typically that a speaker is permitted to use a pronoun to refer to the single most prominent referent precisely because a comprehender will interpret that pronoun to refer to the single most prominent referent.[2] Under such an assumption, the remaining task for experimental research is to identify the factors that determine referent prominence. That research has yielded the set of factors reviewed in Section 1.

But examples like (11b), in which the speaker uses a non-pronominal form to mention a highly salient referent, are a challenge to the mirror image assumption. A further challenge comes from story continuation data like the following from Stevenson et al. (1994).

(12) *Story continuation results with and without pronoun prompt (Stevenson et al., 1994)*
   a. Mary annoyed Sue. _____ [63% Mary]
   b. Mary scolded Sue._____ [72% Sue]
   c. Mary annoyed Sue. She _____ [78% Mary]
   d. Mary scolded Sue. She _____ [66% Sue]

Recall that *annoy* and *scold* are classified as implicit causality verbs, with *annoy* being subject-biased and *scold* being object-biased in that participants' continuations favor the causally implicated subject referent Mary for (12a) and the causally implicated object referent Sue for (12b). What Stevenson et al. found was that these preferences shifted in a condition with a pronoun prompt. The pronoun is fully ambiguous, but its presence served to increase the number of continuations about the subject (meaning more about the causally implicated referent for (12c), fewer for (12d)). To understand this, we need a model that positions pronouns within the space of possible anaphoric forms a speaker might produce. In § 3 generative models are introduced as a framework that can provide the insight we need for a better understanding of pronouns (see § 4).

[2]Other terms for the most prominent referent include salient, accessible, activated, in focus, the center of attention, etc. All are associated with the same assumption that pronoun production and interpretation are mirror images of each other.

3. Likely messages and likely forms

This section is your statistical interlude to introduce generative models. Generative models provide a way to understand the output of a linguistic system by considering what forces generated that output. We'll start with an example. Imagine we're looking ahead to US election night in November 2016. The results are coming in and a speaker—whose political party you know—utters the following potentially ambiguous sentence:

(13) The awful one has won!

In order to understand the speaker's intended message—i.e., whether the winner is Democratic candidate Hillary Clinton or Republican candidate Donald Trump—one needs to consider both the probabilities of the possible messages and the probabilities with which this speaker would have generated different utterance forms to convey each message. What's of interest is how the prior probability of a given message combines with the appropriateness of a particular message form in this context (i.e., when produced by this speaker).

Let's say that the probabilities that the different messages would be generated are determined by real-world knowledge, independent of the speaker, as in (14).

(14) *Probabilities over possible speaker messages*[3]
p(`Clinton won`) = .8
p(`Trump won`) = .2

Next, let's say that the probability the message is generated using the form in (13) varies with the message and what is known about the discourse context of the utterance, namely the political party of the speaker, as in (15). Note that the values associated with these forms are invented and necessarily sidestep the issue of how many possible forms there are that a speaker might produce; here I treat the set of available forms as small and closed, with the expression "The awful one has won" being one of the only possible epithets. The conditional probability is written p(outcome | condition).

(15) *Probabilities over possible forms, conditioned on message, with the assumption that the speaker is a Democrat*
p("The awful one has won" | `Clinton won`)=.25
p("The awful one has won" | `Trump won`)=.99

Here, the speaker who is a Democrat is very likely (probability estimate 0.99) to generate a disparaging description of Trump, but since many Democrats are also wary of Clinton, knowing the speaker is a Democrat doesn't preclude the possibility that a negative description of Clinton will be generated (hence

[3]Election outcomes reflect the 30 June 2016 estimates from `http://projects.fivethirtyeight.com/2016-election-forecast/`.

probability estimate .25). Note that the probabilities in (15) do not sum to 1. This is expected. What must sum to 1 are the probabilities for all the forms that could be generated to convey a particular meaning. For example, utterances that described Trump with kinder words are assumed to be very unlikely for a Democrat (the remaining probability .01).

The comprehender's job is to recover a likely meaning given the form of the utterance. That job is encapsulated as the conditional probability of a message given a particular form: p(message | form). Bayes Rule, if you're not familiar with it, can be used to reframe that probability in terms of two probabilities about what the speaker might have done: the prior probability of the intended message and the probability that that message would be conveyed with that particular form. See (16). Bayes' Rule is a general theorem of probability theory that can be used to describe any belief updating (here, about a message) in light of new evidence (here, the utterance form).[4]

(16) *Bayes' Rule* p(message | form) $\sim$ p(message) $\times$ p(form | message)

Now let's apply Bayes' Rule to the interpretation of the speaker's utterance on election night, as in (17a). Bayes' Rule states that the probability that the speaker's intended message is `Clinton won`, given the words "The awful one has won", is proportional to the prior probability that this speaker would generate such a message (is `Clinton won` a likely message a priori?) combined with the likelihood that such a message would be conveyed in this context with those words (would this speaker, a known Democrat, produce a negative description of Clinton?). Formula (17b) shows a similar calculation for the alternative intended message that `Trump won`.

(17) a. *Democratic speaker intends to convey the message that Clinton won*
p(`Clinton won` | "The awful one has won")
$\sim$ p(`Clinton won`) $\times$ p("The awful one has won" | `Clinton won`)
$\sim .8 \times .25 = .200$

b. *Democratic speaker intends to convey the message that Trump won*
p(`Trump won` | "The awful one has won")
$\sim$ p(`Trump won`) $\times$ p("The awful one has won" | `Trump won`)
$\sim .2 \times .99 = .198$

It's a toss up: The probability that the sentence in (13) would be interpreted to mean `Clinton won` is roughly equal to the probability that it means `Trump won`. The sentence is confusing when uttered by a Democratic speaker precisely because neither message has a clear advantage given the priors over likely election outcomes and the production probabilities for a Democratic speaker.

[4]The symbol $\sim$ indicates that the probability on the left side of the formula is *proportional* to the value computed on the right side. For the exact value, the right side would include a denominator summing over the numerator for all messages.

Now consider what message would be inferred if that same utterance were generated by a Republican. The probabilities of the election outcomes remain the same—assuming that the Republican speaker is following the same polls and election predictions. But the production probabilities for that particular utterance *form* are different in this case: Here I estimate that the Republican speaker is highly likely (.99) to use negative words to describe Clinton, but there may be Republicans who also are wary of Trump (and perhaps more than there are Clinton-wary Democrats, hence an estimate of .3).

(18) a. *Republican speaker intends to convey that Clinton won*
p(`Clinton won` | "The awful one has won")
$\sim$ p(`Clinton won`) $\times$ p("The awful one has won" | `Clinton won`)
$\sim .8 \times .99 = .792$

b. *Democratic speaker intends to convey that Trump won*
p(`Trump won` | "The awful one has won")
$\sim$ p(`Trump won`) $\times$ p("The awful one has won" | `Trump won`)
$\sim .2 \times .3 = .06$

In this case, with a Republican speaker, the intended message is much more obvious. The scenario in which the speaker intended to convey `Clinton won` receives a much higher probability than the one in which they intended to convey `Trump won`.

The example here shows how real-world knowledge can guide the probabilities of different intended messages and how the appropriateness of particular forms varies by context (i.e., depending on the political affiliation of the speaker). In the next section we apply Bayes' Rule to model pronoun use, for which it can likewise be used to combine biases about intended messages with estimates of the appropriateness of particular forms in a particular discourse context.

4. Pronouns in a generative model: a Bayesian approach

Just as a speaker's ambiguous utterance about the winner of an election can be decoded by modeling the process by which the utterance was generated, an ambiguous pronoun can likewise be decoded by considering how it was generated. Bayes' Rule in (19) describes how to update one's belief that a particular referent has been mentioned, given that a pronoun has been encountered, (p(referent | pronoun)). It does so by combining the prior belief that that referent would be mentioned, p(referent), with the likelihood that that referent would be mentioned with a pronoun, p(pronoun | referent).

(19) *Bayes' Rule for pronouns (as proposed in Kehler et al., 2008)*
p(referent | pronoun) $\sim$ p(referent) $\times$ p(pronoun | referent)

The probability on the left-hand side of (19) represents the comprehender's interpretation problem of decoding an ambiguous pronoun. The probabilities

on the right are the estimates of the speaker's encoding of the intended referent. From (19), it is clear that interpretation and production of pronouns are linked. However, the two are not mirror images. Rather, Bayes' Rule shows how they are related via the prior.

Bayes' Rule is not a model of pronouns; it is simply a statement of mathematical truth. However, it is useful for understanding pronouns because it makes it impossible to conflate the probability of mention and the probability of pronominalization. The proposal in (20) explicitly links different features to the two probabilities.

(20) *Proposal for a division of labor in a Bayesian model of pronouns (Kehler & Rohde, 2013)*
Different factors influence the probability of mention (the prior) and the probability of pronominalization (the likelihood). The prior probability of mention is influenced primarily by semantic and coherence-driven cues. The likelihood of forms is influenced primarily by structural properties of the preceding discourse.

To see how the Bayesian approach and this division-of-labor proposal play out, this section describes two studies that manipulate coherence-driven and structural factors.

4.1. Coherence-driven factors influence next mention

Let's return to the class of implicit causality verbs discussed above. That earlier work by Stevenson et al. had noted different coreference patterns depending on the presence or absence of a pronoun. The Bayesian approach can help us make sense of that result. The story continuation paradigm allows us to explicitly measure the prior probability of next mention, the likelihood of pronominalization, and the pronoun interpretation probability for the available referents, as shown in (21). By using story continuation prompts with and without a pronoun ((21a-c) versus (21d-f)), we can test which factors influence pronoun interpretation versus production. By using verbs with opposing implicit causality biases (subject-bias vs. object-bias) and non-implicit-causality (Non-IC) verbs, we can test for coherence-driven effects and structural effects because the causally implicated referent appears either in subject position or object position (Rohde, 2008; see also Kehler & Rohde, 2013; Fukumura & van Gompel, 2010). The prompts in (21) served as materials for a study reported in Rohde & Kehler (2014).

(21) *Influence of coherence in interpretation and production of pronouns?*

a. [Subj-bias] Mary annoyed Sue. She ___	}	pronoun prompt
b. [Obj-bias] Mary scolded Sue. She ___		to estimate
c. [Non-IC] Mary babysat Sue. She ___		p(ref|pronoun)
d. [Subj-bias] Mary annoyed Sue. ___	}	full-stop prompt
e. [Obj-bias] Mary scolded Sue.___		to estimate
f. [Non-IC] Mary babysat Sue.___		p(ref), p(pronoun|ref)

The results of the study support the division-of-labor proposal. Figure 2 shows the percentage of continuations that started with a mention of the subject. The light bars confirm that participants' choice of who to mention varies by verb class, in keeping with the proposal that the prior is sensitive to coherence-driven factors: IC contexts favor Explanation coherence relations; subject-biased IC verbs favor Explanations about the subject; object-biased IC verbs favor Explanations about the object; Non-IC verbs favor other coherence relations with less consistent subject/object preferences. The dark bars show that participants' pronoun interpretation also varies by verb class, in keeping with the Bayesian approach: The coherence-driven effects on the prior contribute to the overall pattern of pronoun interpretation. These coherence-driven effects on choice of mention emerge as a main effect of verb class.

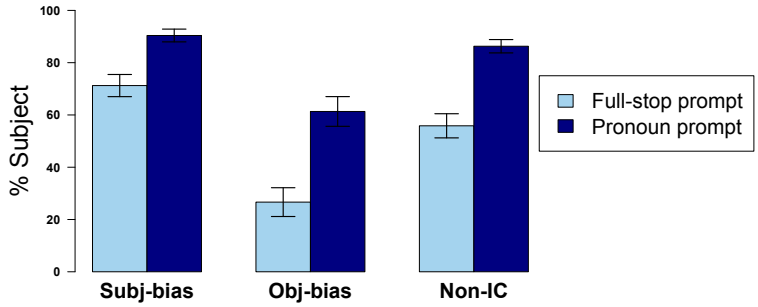

Figure 2: Implicit causality story continuation results: Verb class and prompt type influence next mention ($\underline{\text{p(referent|pronoun)}} \sim \underline{\text{p(referent)}} \times \text{p(pronoun|referent)}$), as reported in Rohde & Kehler (2014).

In addition, the dark bars in Figure 2 are higher than the light bars, a main effect of prompt type. This would make sense if the likelihood of a pronoun being produced were higher for subjects than non-subjects because, in the Bayesian approach, that likelihood term also contributes to the pronoun interpretation probability.

Indeed, Figure 3 shows that surface structure matters for pronominalization: More pronouns are produced when the antecedent is the subject of the preceding sentence, regardless of verb class. Even though object-biased IC verbs favor the object for re-mention, those mentions of the object are no more likely to be pronominalized than objects of subject-biased IC verbs (a main effect of antecedent grammatical role but no effect or interaction with verb class). Other contexts show a similar dissociation between prior probability of next mention and the probability of pronominalization (Kaiser, 2010; Kehler & Rohde, 2017; though cf. Rosa & Arnold, 2017 and Davies and Arnold's chapter (this volume) for a competing approach to the role of predictability in pronominalisation).

Having shown how coherence-driven factors influence pronoun interpreta-

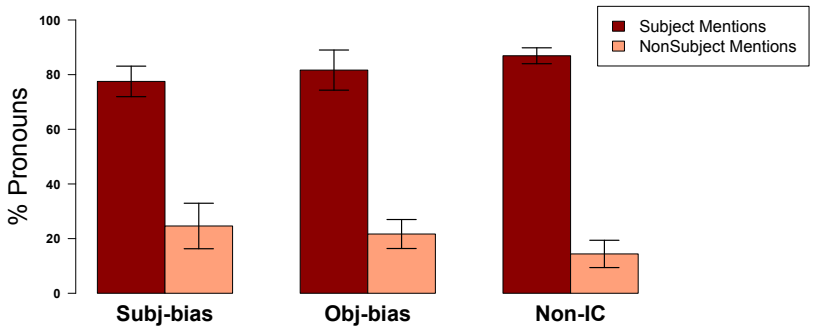

Figure 3: Implicit causality story continuation results: Antecedent grammatical role, but not verb class, influences pronoun production (p(referent|pronoun) $\sim$ p(referent) $\times$ p(pronoun|referent)), as reported in Rohde & Kehler (2014).

tion via the prior, the next study targets the structural effects that influence pronoun interpretation via the likelihood of pronominalization. The question is whether structural effects simply reflect grammatical structure (i.e., subjecthood) or whether such effects themselves are driven by information structure (i.e., topichood; see Bosch & Umbach, 2006 for a similar question regarding the interpretation of personal pronouns and demonstrative pronouns in German).

4.2. Topichood influences pronominalization

The finding that pronouns are produced at a higher rate for subject antecedents than non-subject antecedents makes sense if the central function of pronouns (in English) is to signal a continuation of the current topic (Ariel, 1990; Grosz et al., 1995; Gundel et al., 1993; Lambrecht, 1994) and if subject and topic (in English) are highly correlated. Declining rates of pronominalization for antecedents in more oblique structural positions may reflect the declining likelihood that an entity in that position is the topic.

A second study reported in Rohde & Kehler (2014) manipulates topichood while holding grammatical role constant in order to test whether the subject bias in pronoun production reflects subjecthood or topichood. The materials for the study used a voice manipulation, as in (22).

(22) *Do pronominalization rates reflect syntactic structure or information structure?*

a. [active] Amanda amazed Brittany. ___ } full-stop prompt shows
b. [passive] Brittany was amazed by Amanda. ___ } p(ref), p(pronoun|ref)

c. [active] Amanda amazed Brittany. She ___ } pronoun prompt shows
d. [passive] Brittany was amazed by Amanda. She ___ } p(ref|pronoun)

As noted above, the subject of an active clause in English is considered a default topic, but the passive's promotion of another constituent to subject position makes that subject even more likely to be the topic. If the choice to produce

a pronoun reflects topichood more generally, the prediction is that participants who re-mention the subject of the passive prompt (22b) will be even more likely to use a pronoun than those who re-mention the subject of the active prompt (22a). As such, a main effect of antecedent grammatical role is expected and would replicate the previous study's structure-driven effect. The critical new finding would be an antecedent grammatical role × voice interaction, whereby subjecthood increases pronominalization more in the passive.

The presence of the pronoun prompt is hence expected to have a greater impact on choice of mention in the passive than the active condition: The passive prompt (22d) is predicted to yield more subject continuations than the active prompt (22c), both of which are predicted to yield more subject continuations than the full-stop prompts (22a,b). Note that, because the active/passive manipulation changes the position of the causally implicated referent, the percentage of subject continuations is predicted to vary for coherence-driven reasons between the active and the passive as well. Main effects of voice and prompt type would replicate the previous study's coherence-driven and structure-driven effects. Of interest is the voice × prompt interaction, whereby a pronoun prompt has a greater impact on next mention in the passive condition.

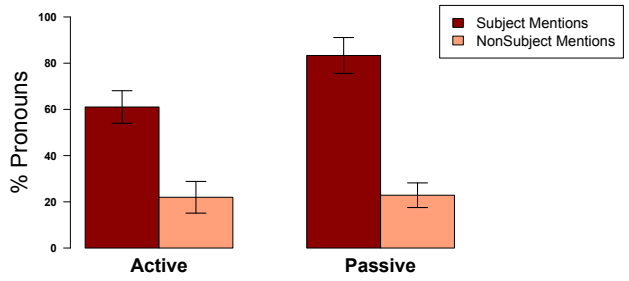

Figure 4: Voice manipulation story continuation results: Antecedent grammatical role and voice combine to influence pronoun production (p(referent|pronoun) ∼ p(referent) × p(pronoun|referent)), as reported in Rohde & Kehler (2014).

The results are shown in Figures 4 and 5. All predictions were borne out. The rates of pronominalization in Figure 4 replicate the main effect of antecedent grammatical role and show the predicted antecedent grammatical role × voice interaction: more pronouns for subject referents than object referents, and even more when the referent is the subject of a passive. The rates of subject re-mention in Figure 5 likewise replicate the previous study's effects and show the predicted voice × prompt interaction: more subject continuations for the active than the passive and more for the pronoun prompt than the full-stop prompt, and even more for the passive pronoun prompt condition.

These two studies point to the importance of pragmatics in pronoun interpretation, via general reasoning underlying the establishment of discourse coherence and information structure underlying the choices in surface realiza-

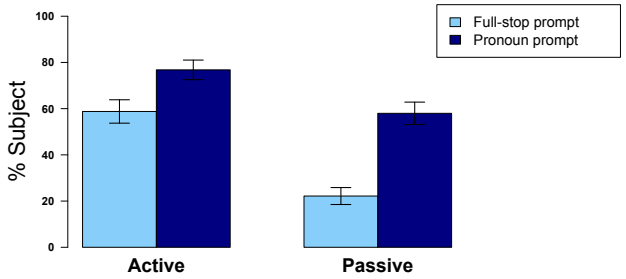

Figure 5: Voice manipulation story continuation results: Voice and prompt type combine to influence next mention (p(referent|pronoun) ∼ p(referent) × p(pronoun|referent)), as reported in Rohde & Kehler (2014).

tion. The first study highlighted the role of the verb in guiding coherence-driven biases. The second study highlighted the role of syntactic structure in signaling topichood. Other non-structural factors may also influence topichood. For example, the presence of more than one referent in a context may reduce each one's likelihood of being the topic, and there is experimental evidence showing that speakers indeed produce fewer pronouns when the context includes two referents compared to one (Arnold & Griffin, 2007). Kim's chapter (this volume) likewise discusses the role of information structure in domains such as coreference, including implicit causality contexts. See also Tonhauser's chapter on the complexity of establishing focus/topic categories and the repercussions for theories of meaning.

4.3. Model comparison

The findings from these two studies support the Bayesian approach and the proposal of a division of labor between factors that influence choice of mention and choice of form. But how well are these results captured by alternative models?

The coherence-driven model described in Section 1 focuses on meaning at the expense of form. In these two studies, coherence-driven effects on pronoun interpretation emerge as biases to the causally implicated referent. These biases in the data can be measured in participants' preferences in the full-stop condition: p(referent), the prior over referents. Focusing on the prior is the essence of Arnold's (2001) Expectancy Hypothesis—a model in which comprehenders are posited to interpret a pronoun to refer to the referent they most expect to hear mentioned next, with a variety of semantic and syntactic features being included as cues to this expectation. The problem is that a model based only on the prior overestimates the bias to the object referent for object-biased implicit causality verbs and underestimates the bias to the subject referent for subject-biased verbs. Why? Because it fails to take into account how the presence of a pronominal form increases the number of subject interpretations.

In contrast, a model that focuses only on surface structural factors ignores meaning-driven effects. In these two studies, structural effects emerge as a bias to produce more pronouns for referents which were previously mentioned in subject position, particularly as the subject of a passive construction. This bias can be measured in participants' behavior in the full-stop condition: p(pronoun|referent), the likelihood of pronominalization. Focusing on the likelihood of pronominalization as a model of pronoun interpretation is at the heart of the mirror-image assumption—i.e., the assumption that speakers produce pronouns to refer to a particular referent only if a comprehender who hears that pronoun will interpret it to refer to the correct referent. The problem is that a model based on the likelihood alone ignores the prior. There are cases in which the prior expectation for a mention of the non-subject referent is strong enough to shift comprehenders' pronoun interpretation to that referent, even though they may be aware that pronouns are not the preferred form for re-mentioning that referent.

The Bayesian approach incorporates the insights of both these types of models. It also shifts the analysis of pronouns from the question of how they are resolved to a question of how they are generated. Older models (e.g., Garnham & Oakhill, 1985) had focused on comprehenders' interpretation as the target phenomenon and drawn conclusions about pragmatic inference playing a role only in a later stage of the interpretation process. By switching to a generative model, the focus is on how a comprehender estimates the speaker's production process, and it suggests that the components of that process rely immediately and inextricably on pragmatic concepts like coherence and information structure.

5. How does a generative model clarify pronoun puzzles?

The Bayesian approach presented here reconciles two competing approaches to pronouns—one that targets coherence (Hobbs, 1979; Kehler, 2002) and one that targets surface structure and form (Grosz et al., 1995). Encouragingly, it also makes specific predictions that are borne out in experimental data. In addition, the Bayesian approach helps clarify the meaning of "prominence", a term whose sense within the coreference literature is often muddled in discussions of speakers' and comprehenders' respective biases to produce and interpret pronouns to refer to "prominent" entities. The Bayesian approach makes it clear that a referent can be prominent for coherence-driven reasons, such that a speaker is likely to mention that referent, or it can be prominent for information-structural reasons, such that, if mentioned, it is likely to be referenced with a pronoun. These two senses of prominence—prominence for mention and prominence for pronominalization—need to be understood on their own terms, particularly because they appear to reflect different properties of the discourse.

Moving to a generative model also invites a reconsideration of a number of existing results in the literature. It's possible that questions that were previously framed in terms of resolution can be better addressed from a generative

perspective. For example, researchers have asked what factors guide pronoun resolution and whether those factors differ from the factors that guide the processing of names and full noun phrases (Sanford et al., 1988; Garrod et al., 1994). Sanford et al. report that referents are more likely to be mentioned again if they have first been introduced with a name than with a description and that sentences with a pronoun are read faster if the antecedent was introduced with a name than with a description. But without knowing what *form* of reference is favored by speakers when re-mentioning the available referents, we can't tell whether the reading time results reflect facilitation from the form of reference (pronouns may be appropriate for antecedents introduced with names) or the choice of reference (the prior for re-mentioning an antecedent introduced with a name may be high even if the likelihood of using a pronoun is low) or both.

This chapter has focused on 3rd person singular pronouns and contexts in which those pronouns refer to a human referent. But this is only the tip of the iceberg for the use of pronouns. For example, the pronoun *it* and the demonstrative *this* can be used to refer back to events rather than entities, as in (23), or even to a sequence of utterances that comprise a rhetorical argument, as in (24).

(23) The dog tripped, fell into his food bowl, and splashed the cat. It was a disaster.

(24) The cat got annoyed because the food splashed its paws. The food splashed because the dog fell in the bowl. The dog fell because he tripped. This is what someone told me.

One can imagine using the Bayesian approach to model what kinds of messages a speaker might intend—messages about the entities, the events, or the discourse itself—and what kinds of referential forms would be used (see Kaiser et al., 2009; Brown-Schmidt et al., 2005 for contrasts between *him/himself* and *it/that*). Under a generative model, a comprehender who encounters the pronoun *it* in (23) must consider whether the speaker would have been more likely to reference one of the entities (the dog, the bowl, the cat), one of the events (the dog-falling event), or the sequence of events (the disastrous trip+fall+splash) and what forms would have been used in each case ("the dog", "it", "Fido", "the fall", "the scene", etc.). In addition, they may consider the possible ambiguity of the resulting expression. See Davies and Arnold's chapter (this volume) on trade-offs between informativity and ambiguity in choice of referring expression more broadly.

Not only is this an opportunity to reassess pronoun use in English, but the Bayesian approach may also provide a framework for thinking about the referential system in other languages. In some languages, a human referent may be mentioned again with, say, a pronoun or a demonstrative, or a null or overt pronoun. Existing work in this area has pointed to a division of labor between particular referential forms and the antecedents they are compatible with (e.g., Carminati's (2002) Position of Antecedent Hypothesis), but the kinds of tasks that are used often conflate preferences for *who* gets mentioned and *how*.

It may also be worth considering how second language learners master these two dimensions of reference in a non-native language, and whether success at estimating who will be mentioned is distinct from understanding what form is most appropriate (see Grüter et al., 2017).

Lastly, the approach advocated here situates pronouns within one of the leading frameworks for thinking about language production and interpretation. Generative models have been applied to a number of pragmatic phenomena (Frank & Goodman, 2012; Goodman & Lassiter, 2015; Franke & Jäger, 2016), and this chapter makes the case that generative models can also clarify our understanding of pronouns. If you're new to generative models, then pronouns may provide a convenient case study for seeing the explanatory power of such models and the role of experimentation in testing such models.

Acknowledgements

The perspectives taken in this chapter reflect my PhD training with Andrew Kehler and our ongoing collaborations. I also am grateful to Jeff Elman for asking me "Why pronouns?" nearly a decade ago. I thank the participants at the Anaphora & Coherence workshop at the 2016 North American Summer School on Logic, Language, and Information (NASSLLI) at Rutgers University for their input on this material and their invitation to consider some particularly tricky examples.

Almor, A., & Nair, V. A. (2007). The form of referential expressions in discourse. *Language and Linguistics Compass*, *1*, 84–99.

Anderson, A., Garrod, S., & Sanford, A. (1983). The accessibility of pronominal antecedents as a function of episode shifts in narrative text. *Quarterly Journal of Experimental Psychology*, *35*, 427–440.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

Arnold, J. E. (1998). *Reference Form and Discourse Patterns*. Ph.D. thesis Stanford University.

Arnold, J. E. (2001). The effects of thematic roles on pronoun use and frequency of reference. *Discourse Processes*, *31*, 137–162.

Arnold, J. E. (2010). How speakers refer: the role of accessibility. *Language and Linguistic Compass*, *4*, 187–203.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition*, *76*, B13–B26.

Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language*, *56*, 521–536.

Arnold, J. E., & Lao, S. C. (2015). Effects of psychological attention on pronoun comprehension. *Language, Cognition, and Neuroscience*, *30*, 832–852.

Asher, N. (1993). *Reference to abstract objects in discourse*. Boston: Kluwer Academic.

Asher, N., & Lascarides, A. (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.

Bach, K. (1994). *Thought and reference*. Oxford: Clarendon Press.

Bosch, P., & Umbach, C. (2006). Reference determination for demonstrative pronouns. In *Proceedings of the conference on intersentential pronominal reference in child and adult language* (pp. 39–51).

Brennan, S. E. (1995). Centering attention in discourse. *Language and Cognitive Processes*, *10*, 137–167.

Brennan, S. E., Friedman, M. W., & Pollard, C. J. (1987). A centering approach to pronouns. In *Proceedings of the 25th Meeting of the Association for Computational Linguistics*.

Brown-Schmidt, S., Byron, D. K., & Tanenhaus, M. K. (2005). Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language*, *53*, 292–313.

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

Carminati, M. N. (2002). *The processing of Italian subject pronouns*. Ph.D. thesis University of Massachusetts at Amherst.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25–56). New York: Academic Press Inc.

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

Chang, F. R. (1980). Active memory processes in visual sentence comprehension: Clause effects and pronominal reference. *Memory and Cognition*, *1*, 58–64.

Clark, H. H., & Sengul, C. J. . (1979). In search of referents for nouns and pronouns. *Memory and Cognition*, *7*, 35–41.

Cowles, H. W., & Ferreira, V. S. (2012). The influence of topic status on written and spoken sentence production. *Discourse Processes*, *49*, 1–28.

Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, *19*, 245.

Crinean, M., & Garnham, A. (2006). Implicit causality, implicit consequentiality and semantic roles. *Language and Cognitive Processes*, *21*, 636–648.

Ehrlich, K., & Rayner, K. (1983). Pronouns assignment and semantic integration during reading: Eye-movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior*, *22*, 75–87.

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

Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language*, *56*, 357–383.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998.

Franke, M., & Jäger, G. (2016). Probabilistic pragmatics, or why bayes' rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft*, *35*, 3–44.

Frederiksen, J. R. (1981). Understanding anaphora: Rules used by readers in assigning pronominal referents. *Discourse Processes*, *4*, 323–347.

Fukumura, K., & van Gompel, P. G. (2010). Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language*, *62*, 52–66.

Garnham, A., & Oakhill, J. (1985). On-line resolution of anaphoric pronouns: effects of inference making and verb semantics. *British Journal of Psychology*, *76*, 385–393.

Garrod, S., Freudenthal, D., & Boyle, E. (1994). The role of different types of anaphor in the online resolution of sentences in a discourse. *Journal of Memory and Language*, *33*, 39–68.

Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, *5*, 459–464.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, *32*, 99–156.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, *27*, 699–717.

Givón, T. (1983). Topic continuity in discourse: an introduction. In T. Givón (Ed.), *Topic continuity in discourse: a quantitative cross-language study* (pp. 1–42). Amsterdam: John Benjamins Publishing.

Goodman, N. D., & Lassiter, D. (2015). Probabilistic semantics and pragmatics: Uncertainty in language and thought. In S. Lappin, & C. Fox (Eds.), *The Handbook of Contemporary Semantic Theory, 2nd Edition*. Wiley-Blackwell.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, *17*, 311–347.

Greene, S. B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 266–283.

Grosz, B. J. (1977). *The representation and use of focus in dialogue understanding*. Technical Report No. 151, Artificial Intelligence Center, SRI International, Menlo Park, CA.

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

Grüter, T., Rohde, H., & Schafer, A. (2017). Coreference and discourse coherence in l2: The roles of grammatical aspect and referential form. *Linguistic Approaches to Bilingualism*, *7*, 199–229.

Gundel, J., Hedberg, H., & Zacharski, R. (1993). Referring expressions in discourse. *Language*, *69*, 274–307.

Hartshorne, J. K., Nappa, R., & Snedeker, J. (2015). Development of the first-mention bias. *Journal of Child Language*, *42(2)*, 2–25.

Hartshorne, J. K., & Snedeker, J. (2012). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes*, *28*, 1–35.

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

Hobbs, J. R., Stickel, M. E., Appelt, D. E., & Martin, P. (1993). Interpretation as abduction. *Artificial Intelligence*, *63*, 69–142.

Holler, A., & Irmen, L. (2007). Empirically assessing effects of the right frontier constraint. In A. Branco (Ed.), *Anaphora: Analysis, algorithms and applications*.

Hudson-D'Zmura, S., & Tanenhaus, M. K. (1998). Assigning antecedents to ambiguous pronouns: The role of the center of attention as the default assignment. In M. Walker, E. Prince, & A. Joshi (Eds.), *Centering Theory in Discourse* (pp. 199–226). Oxford: Oxford University Press.

Kaiser, E. (2010). Investigating the consequences of focus on the production and comprehension of referring expressions. *International Review of Pragmatics*, *2*, 266–297.

Kaiser, E., Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition*, *112*, 55–80.

Kameyama, M. (1996). Indefeasible semantics and defeasible pragmatics. In M. Kanazawa, C. Pinon, & H. de Swart (Eds.), *Quantifiers, Deduction, and Context*.

Karmiloff-Smith, A. (1985). Language and cognitive processes from a developmental perspective. *Language and Cognitive Processes*, *1*, 61–85.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (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*, *3*, 1–37.

Kehler, A., & Rohde, H. (2017). Evaluating an expectation-driven QUD model of discourse interpretation. *Discourse Processes*, *54*, 219–238.

Koornneef, A. W., & Sanders, T. J. M. (2013). Establishing coherence relations in discourse: The influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes*, *28*, 1169–1206.

Koornneef, A. W., & van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, *54*, 445–465.

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Malt, B. C. (1985). The role of discourse structure in understanding anaphora. *Journal of Memory and language*, *24*, 271–289.

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, *8*, 243–281.

Marslen-Wilson, W., Levy, E., & Tyler, L. K. (1982). Speech, place, and action. In R. J. Jarvella, & W. Klein (Eds.), *Producing interpretable discourse: The establishment and maintenance of reference.* (pp. 339–378). Chichester: Wiley.

Pickering, M. J., & Majid, A. (2007). What are implicit causality and consequentiality? *Language and Cognitive Processes*, *22*, 780–788.

Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics*, *12*, 601–638.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A. K., & Webber, B. L. (2008). The Penn Discourse TreeBank 2.0. In *LREC*. Citeseer.

Prince, E. (1985). Fancy syntax and 'shared knowledge'. *Journal of Pragmatics*, *9*, 65–81.

Pyykkönen, P., & Järvikivi, J. (2009). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*, (pp. 1–12).

Reinhart, T. (1981). Pragmatics and linguistics: Analysis of sentence topics. *Philosophica*, *27*, 53–94.

Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. *OSU Working Papers in Linguistics*, *49: Papers in Semantics*.

Roberts, C. (2016). Coherence and anaphora. In *Proceedings of the NASSLLI Workshop on Anaphora and Coherence*.

Rohde, H. (2008). *Coherence-Driven Effects in Sentence and Discourse Processing*. Ph.D. thesis University of California, San Diego.

Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition, and Neuroscience*, (pp. 912–927).

Rohde, H., Kehler, A., & Elman, J. E. (2007). Pronoun interpretation as a side effect of discourse coherence. In *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 617–622).

Rosa, E. C., & Arnold, J. E. (2017). Predictability affects production: Thematic roles affect reference form selection. *Journal of Memory and Language*, *94*, 43–60.

Sanders, T. J., Spooren, W. P., & Noordman, L. G. (1992). Toward a taxonomy of coherence relations. *Discourse processes*, *15*, 1–35.

Sanford, A. J., & Garrod, S. C. (1981). *Understanding written language*. New York: Wiley.

Sanford, A. J., Moar, K., & Garrod, S. C. (1988). Proper names as controllers of discourse focus. *Language and Speech*, *31*, 43–56.

Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in english. *Journal of Verbal Learning and Verbal Behavior*, *13*, 272.

Shibatani, M. (1985). Passives and related constructions: A prototype analysis. *Language*, *61*, 821–848.

Smyth, R. (1994). Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, *23*, 197.

Speelman, C. P., & Kirsner, K. (1990). The representation of text-based and situation-based information in discourse comprehension. *Journal of Memory and Language*, *29*, 119–132.

Stevenson, R., Crawley, R., & Kleinman, D. (1994). Thematic roles, focusing and the representation of events. *Language and Cognitive Processes*, *9*, 519–548.

Stevenson, R., Knott, A., Oberlander, J., & McDonald, S. (2000). Interpreting pronouns and connectives: Interactions among focusing, thematic roles, and coherence relations. *Language and Cognitive Processes*, *15*, 225–262.

Stewart, A. J., Pickering, M. J., & Sanford, A. J. (1998). Implicit consequentiality. In *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 1031–1036).

Stewart, A. J., Pickering, M. J., & Sanford, A. J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory and Language*, *42*, 423–443.

Van Berkum, J. J. A., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*, *1146*, 158 171.

von Heusinger, K. (2002). Reference and representation of pronouns. In H. Wiese, & H. Simon (Eds.), *Pronouns: Grammar and Representation* (pp. 109–135). Amsterdam: Benjamins.

Walker, M., Iida, M., & Cote, S. (1994). Japanese discourse and the process of centering. *Computational Linguistics*, *20*, 193–232.

Ward, G., & Birner, B. (2004). Information structure and non-canonical syntax. In L. R. Horn, & G. Ward (Eds.), *The handbook of pragmatics* (pp. 153–174). Oxford: Basil Blackwell.

Winograd, T. (1972). *Natural Language Understanding*. New York: Academic Press.

Wolf, F., Gibson, E., & Desmet, T. (2004). Discourse coherence and pronoun resolution. *Language and Cognitive Processes*, *19*, 665–675.

Zubin, D. (1979). Discourse function of morphology: The focus system in German. In T. Givón (Ed.), *Syntax and semantics 12: Discourse and syntax*. New York: Academic Press.