Croatian Journal of Philosophy
Vol. XXIII, No. 69, 2023
https://doi.org/10.52685/cjp.23.69.3
Received: September 20, 2023
Accepted: November 13, 2023

From Coherence Relations to the Grammar of Pronouns and Tense

MAGDALENA KAUFMANN*
University of Connecticut, Storrs, USA

Stojnić (2021) argues that the content of linguistic utterances is determined by the rules of natural language grammar more stringently than what is generally assumed. She proposes specifically that coherence relations are encoded by the linguistic structures and determine what individuals count as most prominent, thereby serving as the referents of free ("demonstrative") pronouns. In this paper, I take a close look at the empirical evidence from English and Serbian that she offers in support of this position. Considering these data points in connection with additional linguistic data (also from German and Japanese), I argue that there is no compelling evidence for the assumption that coherence relations directly determine the resolution of pronouns. Instead, grammatical restrictions imposed by different types of pronouns and tenses have a larger impact on the meaning conventionally expressed by complex utterances than what is generally assumed in the literature on coherence relations.

Keywords: Pronouns; tense; coherence relations; information structure; discourse relations.

1. *Introduction*

In her 2021 book *Context and Coherence*, Una Stojnić develops and defends the claim that grammar determines the content expressed by

* This research was supported in part by the author's NSF grant #2116972 "Research on conditional and modal language". For discussion of data and theories, I am indebted to Ivana Jovović, Stefan Kaufmann, Teruyuki Mizuno, Rick Nouwen, Tyler Poisson, Philippe Schlenker, Julian Schlöder, Walter Shaw, and the participants of *Philosophy of Linguistics and Language*, Interuniversity Center Dubrovnik, September 5–9, 2022.

complex natural language expressions to a significantly higher degree than what is standardly assumed:

> [L]anguage–grammar–itself is far more expressive and pervasive than has been assumed; the resolution of context-sensitivity is entirely a matter of linguistic convention. (Stojnić 2021: 5)

In developing an account along these lines, she assigns a uniform, non-ambiguous interpretation to pronouns and to modals, which offers strikingly elegant accounts of quandaries for even a dynamic truth-conditional theory or apparent failures of otherwise intuitive inference patterns. Stojnić (2021) argues, moreover, that crucial linguistic conventions that determine content for the items in question...

> have gone unnoticed, because their principal domains are entire discourses, not just their constituent words and sentences. While it is not controversial that the way sentences are constructed depends on conventions of syntax and semantics which specify the rules by which individual expressions combine, I similarly argue for rules–*discourse conventions*–that specify how individual sentences combine to form a discourse. (Stojnić 2021: 5)

It is this latter point that I take issue with: I do not think that Stojnić (2021) establishes convincingly that coherence relations directly address the resolution of pronouns, and I would like to defend the more traditional picture of the interplay between pronominal resolution and coherence relation resolution in parallel, a position Stojnić explicitly rejects.

Reconciling the points with Stojnić's (2021) main claim as reproduced initially, I would like to argue that it is rather the grammar of pronouns and tenses that is richer than what has been assumed in some parts of the literature. Restrictions imposed by the grammar of these elements will thus reduce the choice of otherwise possible (that is, salient enough) referents for a given pronoun; concerning the remaining options, the choice has to be made between pairings of compatible coherence relations with the content resulting from the alternative pronominal resolutions.

2. *A unified interpretation for pronouns*

At first glance, Stojnić's take on pronouns may sound like a standard credo in linguistic semantics:

> The meaning of a pronoun is simple, uniform, and unambiguous; as a first pass, a pronoun selects the most prominent candidate interpretation—what is "at the center of the attention" at the point in discourse at which it occurs. (2021: 40)

Upon closer inspection, this is probably better considered a standard desideratum that most accounts tacitly stop short of delivering on.

With the advent of dynamic semantics (specifically, *file change semantics*, Heim 1988; *Discourse Representation Theory* (DRT, Kamp and Reyle 1993; *Dynamic Predicate Logic*, Groenendijk and Stokhof 1991), it has become standard to assume that specific linguistic phe-

nomena require us to take into account that context evolves between utterances and also over the course of a single utterance. These developments affect the interpretation of subsequent linguistic material, a phenomenon that is captured by reconceptualizing linguistic meanings as *context change potentials*: functions from input contexts to output contexts, and Stojnić is firmly rooted in that tradition.

Dynamic notions of meaning have been put to use fruitfully when explaining coreference across sentences (e.g. (1b)) and covariation effects within (e.g. (1a)), standardly indicated by coindexation:

(1) a. If a boss$_i$ has employees$_j$ she$_i$ should treat them$_j$ well.
b. A woman$_i$ met a girl$_j$. She$_{i/j}$ greeted her$_{j/i}$.

In capturing such phenomena, dynamic theories also seek to explain that the availability of an antecedent or binder is constrained by the linguistic context; negation, for instance, blocks phenomena as illustrated in (1):[1]

(2) a. Mary does not have a violin. #She will (not) bring *it* to the party.
b. #If Mary does not have a violin, she will not bring *it* to the party.

Despite their impressive traction for explaining these phenomena, when it comes to a treatment of the pronouns themselves, mainstream versions of dynamic semantics tacitly accept infinite ambiguity. Pronouns are translated as variables to be assigned referents by assignments. Treating contexts as sets of assignments or world-assignment pairs[2] provides a way of modeling limited information as associated with an intended referent (Heim 1982, Kamp and Reyle 1993, Groenendijk et al. 1996). For instance, the discourse in (1b) can be integrated successfully even when the interlocutors are unable to identify more precisely the two individuals the speaker has in mind.

(3) a. A phonetic form [hi:] realizes the lexical element he_i for some $i \in$ N.
b. For any $i \in$ N, he_i is translated as x_i.
c. For any $i \in$ N, the interpretation of x_i with respect to a variable assignment g is $g(i)$ (an element of the domain of individuals provided by the model of interpretation).

If we assume moreover that assignments are undefined unless a suitable referent has been stored at an index, then the treatment sketched in (3) provides significant restrictions on what indexations could be available at any given point. Pronouns themselves, however, are still associated with infinitely many possible translations.

A uniform interpretation as 'the most prominent male/female' would be preferable but does not intuitively match the behavior of the pronominal elements. For instance, in the individual domain, Stojnić (2021) shows that in (4a) (in the absence of contrastive intonation),

[1] Dynamic theories also capture the similarity between possible pronominal resolution patterns and presupposition projection (van der Sandt 1992).

[2] Contexts can also be treated as sets of sets of assignments to model plurality, (see van den Berg 1996; Brasoveanu 2006).

he has to refer to John even at the expense of infelicity. In contrast, the definite description *the contextually salient male* is understood straightforwardly as referring to a different individual in (4b).

(4) a. #John came to the party, and he did not come.
b. John came to the party, and the contextually salient male did not come.

An attempt at analyzing the pronoun as picking out the contextually most salient male thus seems doomed. This reasoning, however, is fallacious. By developing a formal account, Stojnić shows that an implementation of this idea need not behave like the object language definite descriptions we can use to describe it, i.e., English phrases like *the contextually salient female*/*the contextually salient male*. Technically, for her, interpretation proceeds with respect to a sequence of possible referents (the *attentional state*), which Stojnić calls a stack, emphasizing the special status of the top-most element.[3] Pronouns are translated as strings beginning with '@', and these expressions are interpreted as picking out the top-most element of the stack that meets the requirement imposed by their lexical content (e.g. being male for *he*, or being female for *she*: these pronouns are translated as *@he* and *@she*, respectively). Specific assumptions about how the stack is affected when updated with a simple predication as expressed by the first conjunct in (4a) ensure that at the point where the pronoun *he* is encountered, the top-most element of the sequence that meets the requirement of being male is John. Specifically, the subject referent gets stored in the top-most position by default, followed by the object referent for a two-place predicate. Pronouns with suitable features are therefore expected to refer to the subject by default. And indeed, subject orientation seems a reasonable default for English pronouns.[4]

[3] Other positions are accessible directly via indexation, making the formal object behave more like a sequence proper.

[4] It is worth noting that the use of stacks to provide an index-free (and thereby semantically unambiguous) treatment of pronouns is not unprecedented. Motivations and data coverage differ across the approaches that come to mind. For instance, van Eijck (2001) develops a stack-based dynamic semantics for pronouns to address the issue of *destructive assignments*: in dynamic accounts, indefinites are taken to introduce new referents. If coreference and binding is handled through indices (variables), a separate requirement has to ensure that the process uses a fresh variable (and does not therefore overwrite the information associated with a variable that is already in use). Nouwen (2003) observes that placing indices in the structure struggles to account for certain patterns of plural coreference in discourse as exemplified in (i) (his (5.8)): despite the difference in interpretation, the plural pronoun *them* has the same antecedent *exactly two papers*.

(i) Three students each wrote exactly two papers. They each sent them to L&P.
a. 'each student sends all the written paper'
b. 'each student sends just their own papers'

Schlenker (2005) develops a stack-based semantic version of classical binding theory. Modal Centering Theory as mentioned in Stojnić (2021) develops related accounts specifically to capture phenomena in the modal and temporal domain (Bittner 2011; Murray 2014). In contrast, sign languages might provide evidence for the existence

Stojnić (2021) does not work out an account for definite descriptions like *the contextually salient male*. As a first pass, these could be analyzed similarly: they would then pick out the top ranked referent in the attentional state that meets the descriptive content (of being a contextually salient male).[5] At first glance, this seems to bring back the issue we set out to avoid—the definite description should behave the same as the pronoun, and the contrast in (4) would remain unaccounted for. This, however, fails to take into account that the technical notion of being listed at the top of the stack thanks to grammatical mechanisms is crucially different from the notion of being the individual that counts as contextually most salient. For instance, there might be a particular individual nobody dares to name but which dominates a conversation about other individuals. Or else, the discourse goal (the question under discussion) may consist in characterizing a particular individual (maybe regarding their chances in an upcoming election), making that person conversationally most salient in this respect. But in doing so, the individual sentences may ascribe properties to their aides, thus associating these individuals with the top position of the stack temporarily. In short, the predicate *conversationally (most) salient* need not be interpreted as true only of the individual at the top of the stack with respect to which interpretation proceeds. The difference between the English phrase *conversationally most prominent* and the technical notion of being at the top of the stack would probably become most obvious when extending Stojnić's (2021) account to quantifier bound pronouns, as in *Every student thinks that he is smart*. A natural way of doing this while retaining the uniform semantics of the pronoun would be to 'borrow' the top-position to run over all the values in the domain, which would of course not lead to each individual becoming conversationally most prominent at least for a split second (see also Stojnić 2021: 42, for a suggestion along these lines). Beyond this, Stojnić's interpretation with respect to an attentional state offers the possibility to directly compare the behavior of pronouns and definite descriptions. As *R(eferential)-expressions* (Binding Theory, Chomsky 1981), definite descriptions are generally expected not to appear in the scope of a co-indexed (binding or coreferential) expression. It is therefore not implausible to assume that they come with a restriction or at least a bias against picking out referents that are salient enough to be accessible for pronominal reference.[6] Systematic exceptions, on which definite descriptions behave

of indices in natural languages (Lillo-Martin and Klima 1990; Schlenker 2018). A comparison of these different stack-theoretic accounts (as well as variable-free approaches more in general, Jacobson 2012) and an evaluation of the arguments from sign linguistics have to be left to future research.

[5] An analysis along these lines is provided by von Heusinger (2004), who models the changing referential prominence of various referents of a NP as choice functions and identifies *he* with *the male* for this purpose.

[6] A treatment along these lines is given in Schlenker's (2005)–also stack-based–semantic implementation of the Binding Principles. Definite descriptions and proper

like anaphoric pronouns, result, however, when the definite description is deaccented as in (5b). In this case, *the shed* has to be interpreted as anaphoric to the cottage, which is thus additionally classified as being merely a shed. In contrast, the accented occurrence in (5a) cannot be interpreted as anaphoric to John's old cottage: with pitch accent on the noun phrase, a new referent is introduced (Umbach 2002, her (1); *bridge anaphora*):

(5) (John has an old cottage.)
a. Last summer he reconstructed the SHED.
b. Last summer he RECONSTRUCTED the shed.

Related effects with definite descriptions in the place of anaphoric pronouns occur in newspaper jargon, where, as Riester (2009) points out, they serve to activate contextually available information that has not yet been discussed (or which is introduced by accommodation):

(6) Gerhard lives in Munich. The father or triplets is 42 years old.

The interaction with prosody is reminiscent of an effect Stojnić observes for pronouns: (4a) can become felicitous when the pronoun is stressed (see discussion in Sect. 4.1).

Ultimately, a full assessment of Stojnić's predictions for the contrast in (4) awaits an application of definite descriptions into the framework. The system seems flexible enough, however, to capture both the similarities and the differences between pronouns and definite descriptions like *the contextually salient male*.

Building on the idea that modals come with contextually supplied domain arguments, Stojnić can reduce the contrast in (7) to a similar problem (Stojnić 2021: 100, her (62) vs. (63)):

(7) a. #If it is not raining and it might be raining, then I am uninformed about the weather.
b. If it is not raining and the body of information *i* does not rule raining out, then I am uninformed about the weather.

Crucially, (7a) 'is not incoherent because there is no body of information that the context can select that yields a plausible interpretation', Stojnić (2021: 121) attests; 'rather, it is incoherent because the context determines the body of information that delivers an inconsistent interpretation' (Stojnić 2021: 121). Findings like (7a) have been adduced as evidence for non-propositionalism about epistemic modals. Stojnić's response is decisive: 'the problem is not, [...], in the idea that these expressions express truth-conditional content; the problem is in the underlying assumption of how a context operates to determine these truth-conditions'. (2021: 7)

As shown for the individual domain, Stojnić's specific implementation of 'most prominent/contextually salient' as the top-ranked one

names are subject to *Avoid Redundancy*, which blocks their use for referents already stored in the sequence of individuals that have been introduced in the linguistic discourse.

of all referents tracked in the on-going conversation avoids the issue elegantly. As a sort of corollary to her main thesis about the power of linguistic convention in the determination of content, we should thus also embrace a word of caution: "Beware of metalinguistic naivete!". Linguistic expressions come with various layers of static and dynamic meaning; faithfully paraphrasing formal accounts in English, our informal metalanguage, is tricky and cannot be taken to refute these accounts.

3. *Coherence relations and pronominal resolution*

Standard dynamic theories focus on what referents are available to be picked up by pronouns at any given point of the conversation. To allow for this, different syntactic objects are standardly differentiated by coindexation (or suitable choices of variables when translating to a formal representation language, Nouwen 2003 for discussion). For (1b) the two salient options that are felicitous out of the blue are given in (8), where indefinites are taken to modify assignments by storing individuals that have the noun phrase property (being a woman and being a girl, respectively) under the indices they carry.

(8) a. A_1 woman met a_2 girl. She$_1$ greeted her$_2$.
b. A_1 woman met a_2 girl. She$_1$ greeted her$_2$.

Stojnić (2021) argues that it is a genuinely grammatical phenomenon that in such cases only one resolution is available, and that content is therefore determined by grammar beyond what is usually assumed. Specifically, the burden is placed on coherence relations. Operative at the level of entire texts, these determine what referent pronouns like *that* or *she* pick out on any given occasion of use. The effect is illustrated with the classical example in (9) (modified from Smyth 1994: see also Hobbs 1979; Kehler 2002; Asher and Lascarides 2003 for discussion of this and similar examples):

(9) Phil tickled Stanley. Liz poked him.

The second sentence in the sequence can be understood as specifying a result of the event in the first (perhaps a show of disapproval), or as an event in parallel to what is described by the first. The resolution of the pronoun *him* stands and falls with the resolution of the coherence relation to Result or Parallel: we understand *him* as referring to Phil when we take the poking to be a result of the tickling, and to Stanley when we conceive of the tickling and the poking as two parallel events.

> [M]ost extant theories treat it as a pragmatic default. Standard coherence theoretic accounts interpret this correlation as evidence of an inferential relationship between a speaker's intention in organizing the discourse and her referential intentions. I argue that this is a a mistake: there is a tighter connection between discourse coherence and pronoun resolution, one underscored by linguistic convention. [...] I argue that discourse relations that connect and organize utterances are a part of the grammar of a language,

> and that they govern the resolution of context-sensitivity as a matter of grammar, too. (Stojnić 2021: 6)

Once a discourse relation is inferred, it determines the resolution of the pronouns by linguistic convention. Emphasizing the difference to the standard picture, Stojnić writes:

> There is reason to think that it's a mistake to treat (3) [our (9)] as harboring *two separate* ambiguities, or *two separate* underspecified elements that must be resolved in turn—one involving discourse coherence, another concerning pronoun resolution. The examples suggest that, once a coherence relation is established, a certain pronoun resolution is automatically set up. [...] there's good reason to conclude that pronoun resolution is *settled* by whichever coherence relations organize a discourse. (Stojnić 2021: 65)

As much as I agree with Stojnić that grammar has a significantly more decisive say in pronominal resolution than what is standardly assumed, I do think that this is a move in the wrong direction: there is good reason to think that the standard account has it right. While discourse relations are known to interact with, and to constrain, pronominal resolution, Stojnić's arguments that coherence relations asymmetrically determine pronominal resolution do not strike me as convincing. I will ultimately argue that the grammar of pronouns and the grammar of tense have a larger role to play than much of the standard literature on pronominal resolution acknowledges. Once the standard picture is updated to reflect this,[7] the data discussed in Stojnić (2021) can be integrated into the standard picture very naturally.

To explore this, like Stojnić, I will assume that, when interpreting utterances in context, language users integrate them into the given linguistic discourse that can be represented in Stojnić's translation language or also in the one of *Segmented Discourse Representation Theory* (SDRT, Asher and Lascarides 2003). Predicates representing coherence relations are a crucial building block in this (according to the SDRT convention, they appear typeset in small caps in the following).

Stojnić emphasizes that the construction of a coherent discourse is subject to constraints similar to what we observe at the sentence level; consider for instance well-known Principles A and B of binding theory (Chomsky 1981). While the reflexive pronoun *himself* in (10a) has to be coreferential with a c-commanding expression in the relevant binding domain (roughly, the smallest domain containing also a subject), the personal pronoun *him* in (10b) cannot be coreferent with another expression in this domain. At the level of interpretation, we obtain the effect of obligatory coreference and disjoint reference as illustrated in (10).[8]

[7] Note that especially DRT already includes very fine-grained studies of temporality.

[8] See Heim (1993) for the need to make the constraint sensitive to presupposed coreference. Consider her example in (i):

(i) Zelda must be the author. She praises her to the sky.

When coreference is not taken for granted but is established in the very sentence, the reflexive pronoun is felicitously replaced by the regular personal pronoun.

(10) a. Phil is shaving himself.
b. Phil is shaving him.

Similarly to this sentence internal constraint on the interpretation of object pronouns as depending on the referent of the subject, specific coherence relations can indeed go hand in hand with constraints on the resolution of personal pronouns. It is, however, far from clear that this requires the coherence relations to directly modify the ranking that underlies pronominal resolution. For instance, the coherence relation PARALLEL enforces a particular interpretation of the pronoun. On the SDRT account this follows, because two sentential structures can be related by PARALLEL only if their arguments match (Asher 1993).[9] Resolutions to individuals other than Stanley fail to instantiate a structure that could count as parallel to the one assigned to the first sentence *Phil poked Stanley*, and the two utterance units can thus not be connected with the relation PARALLEL. Therefore, PARALLEL imposes a constraint on the resolution of the pronoun also on the standard picture, but this constraint is indirect: there is no need to associate PARALLEL with an effect on the stack representing the referents that have been introduced in the discourse; the requirement it imposes (that the predication expressed by two units be parallel in the part of the predicate and each of the arguments) can be met only under one specific resolution of the pronoun and it can therefore not co-occur with the other option.

Other discourse relations, like RESULT or EXPLANATION lack even an indirect formal effect of that sort: they can be inferred as long as the contents of the two sentences under whatever resolution of the pronouns can plausibly be conceived of as standing in the relevant relation; restrictions are imposed, however, on the temporal order of the eventualities involved. For RESULT, the first discourse unit has to describe an eventuality that causes and hence precedes an eventuality described by the second (Asher and Lascarides 2003: 155). For EXPLANATION, the eventuality described by the first discourse unit may not precede the eventuality described by the second. If the second describes an event, this has to strictly precede the eventuality described by the first discourse unit (Asher and Lascarides 2003: 159).

In contrast, Stojnić (2021) argues for a direct connection between coherence relations and pronominal resolutions. She maintains that, in the following discourse, they stand and fall together: if the second sentence is understood as describing a result, its subject pronoun *he* is resolved as referring to John (the matrix subject); if it is understood as an explanation, the subject pronoun *he* is resolved as referring to Tim (the matrix object). To derive this, a representation of the discourse that relates the translations of the two individual sentences by EXPLANATION comes with an operator that reorders the attentional state by

[9] For instance, building on Kehler (2002), Altshuler and Truswell (2022) require that for two discourse units to be related by parallel, they have to share a common theme and all elements that differ have to be similar.

promoting the object over the subject. The intuition seems to be that 'something about the object' explains John's disappointment. In contrast, RESULT maintains the default order with subject prominence.

(11) John was disappointed with Tim.
a. He fired him. RESULT
b. He disobeyed him. EXPLANATION

In contrast, the standard picture (as for instance SDRT) allows for all possible combinations that are compatible with the requirements of the individual coherence relations (at least four in this case), the most coherent one of which will be selected by language users (following the principle of *Maximize Discourse Coherence* (MDC) as laid out in Asher and Lascarides 2003, their section 5.10). Crucially, this selection operates on the different options of pairing coherence relations with pronominal resolutions, rather than letting a coherence relation directly determine the pronominal resolution.[10] In short, according to SDRT (which represents the standard theory in this respect), interpreting a discourse thus involves choosing between all possible combinations of coherence relations and pronominal resolutions. According to Stojnić (2021), interpreting a discourse relation involves choosing between various possible coherence relations (characterized in parts by their impact on pronominal resolutions).

In the following, I will turn to two pieces of empirical evidence that Stojnić offers in support of the idea that coherence relations themselves determine pronominal resolution. I will argue that they rather support the standard theory as reflected in the SDRT framework. In section 4.3 I discuss a third empirical phenomenon which, despite at first glance supporting Stojnić's conceptualization, ultimately also supports independence of coherence relations and pronominal resolution.

4. *Disentangling coherence relations and pronominal resolution*

4.1 *Crosslinguistic variation in coherence relations?*

While processes relying on general purpose reasoning are standardly expected to be invariant across languages, effects tied to linguistic conventions are known to vary across languages, making data along these lines important arguments in favor of the one or the other. Aiming to support the assumption that coherence relations are part of grammar rather than of general all-purpose reasoning, Stojnić (2021) argues that

[10] The discussion in this section simplifies the SDRT picture by ignoring another structural restriction imposed on the resolution of pronouns by the distinction of coordinating and subordinating coherence relations. This difference impacts what sites are, at any point in the discourse, available for the attachment of a subsequent sentence (*Right Frontier Constraint*) that is, with what previous sentence it can be connected. The attachment site then constrains further what referents are available to resolve pronouns to at any given point in the discourse.

the impact of coherence relations on pronominal resolution varies even across Indo-European languages. While English shows the determination by coherence relation discussed above, Serbian does not. Serbian can realize subject pronouns overtly or covertly. Independently of the coherence relation inferred between the two sentences, in subject position, Serbian overt pronouns pick up the object of a previous sentence, covert pronouns pick up the subject of the previous sentence (Stojnić 2021: 69, (48)).[11]

(12) Džon je bio razočaran Timom.
John-NOM is-PRS-3MS be-PPA-3MS disappointed-ADJ-3MS Tim-INS
'John was disappointed with Tim.'
a. Otpustio ga je.
Fired-PPA-3MS him is-PRS-3MS
'He fired him.' (= 'John fired Tim.')
b. On ga je otpustio.
He him is-PRS-3MS fired-PPA-3MS
'He fired him.' (= 'Tim fired John.')

Obviously, Serbian coherence relations cannot determine prominence ranking (contrary to what Stojnić has argued for English); if they did, they would overwrite the grammatical information the pronouns need to access. Stojnić (2021) considers this difference between English and Serbian evidence for the linguistic/conventional nature of CRs.

Upon closer inspection, however, it seems more natural to locate the variation in the pronouns. Across languages, different pronominal elements are well-known to differ in interpretative possibilities: *weak pronouns* (including covert pronouns and clitic pronouns) appear, for instance, as bound variables or anaphoric to linguistically salient elements, whereas *strong pronouns* pick up other referents and can typically not be bound (see Patel-Grosz and Grosz 2017 for overview and references to earlier work). For English, Stojnić herself shows a difference in behavior depending on the prosodic contour realized on the pronoun: with unmarked intonation, (13) cannot be read as referring to Mary, no matter how much Mary's jumping up and down and yelling loudly may make her the most salient female in the conversational context and the center of our attention. As pointed out in Stojnić (2021) fn. 14, stress on the pronoun in combination with a pointing gesture will make Mary the referent (note that pointing at Mary is still infelicitous if the pronoun is not stressed). Stress on the pronoun in the absence of a pointing gesture may still retain Betty as a referent if (13) is used to correct a previous utterance about some other female sitting down.

(13) Betty came in, and {she/ SHE} sat down.

[11] Serbian clitics have to appear in second position, in (12), they therefore either follow the verbal participle or the overt subject pronoun (which is not a clitic). The resulting change in word order is orthogonal to her point.

While these interpretative changes go hand in hand with the phonological shape of the pronoun as well as changes in the overall discourse setting (like previous commitments with respect to which individual sat down), they are perfectly compatible with keeping the discourse relation constant (the two conjuncts in (13) are naturally related by narration, for instance).

For German, relevant contrasts are discussed as obtaining between regular personal pronouns and *d(emonstrative)*-pronouns[12] (Patel-Grosz and Grosz 2017, their (55)):

(14) a. Hans$_1$ wollte mit Paul$_2$ joggen, aber er$_{1,2}$ war krank.
Hans wanted with Paul jog but he was sick
b. Hans$_1$ wollte mit Paul$_2$ joggen, aber der$_2$ war krank.
Hans wanted with Paul jog but DEM was sick
'Hans wanted to go jogging with Paul, but he was sick.'
(adapted from Bosch et al. 2003)

Here, too, without a change in coherence relations (in both cases, the second coordinate stands in contrast to the first, as cued by *but*, and specifies an obstacle to the realization of the desire), we find a difference in pronominal resolution: the personal pronoun can pick up either Hans or Peter, while the d-pronoun can only be resolved to the non-subject participant (Peter). Hinterwimmer (2015) states this as an information-structural constraint: d-pronouns cannot refer to aboutness topics, which is the discourse status typically assigned to referential, unstressed preverbal subjects. The findings from Serbian are best understood along these lines, as well. Jovović (2022, 2023) explores the occurrence restrictions of overt and covert subject pronouns in Serbian. In non-subject positions, where covert pronouns are unavailable, the contrast is replicated by strong pronouns in contrast to clitics. Like the German d-pronouns, Serbian overt pronouns (or more generally, strong pronouns) cannot be anaphoric to a topic.[13]

The relevant linguistic conventions and crosslinguistic differences are thus better located in the grammar of the pronouns than in the grammar of coherence relations. A move along these lines reflects independent findings that different types of pronominals (covert pronouns, reflexives, phonologically reduced pronouns, full pronouns, stressed pronouns) differ in terms of resolution possibilities (Montalbetti 1984;

[12] Stojnić (2021) labels all free personal pronouns demonstrative pronouns; in the linguistic literature, the term is sometimes restricted to specific types of pronouns that are more closely related to pronouns like *this* or *that* (e.g. in German the pronoun series that is homophonous to the definite determiner).

[13] In Jovović's rendering, strong pronouns are acceptable in three environments: (i) when the pronoun's antecedent bears new information focus, (ii) when the pronoun itself is constrastively focused, or (iii) in the presence of a co-sentential focalized adverb. Jovović mostly aims to show that acceptability contrasts that were previously considered to involve syntactic violations (Despić 2011) are better analyzed in terms of the information structural requirements of different pronominal types. She does not develop an analysis of the patterns described.

Patel-Grosz and Grosz 2017; see also the binding principles as exemplified partly in (10)).

4.2 *Different explanations in English*

In (11), repeated here for convenience, we see a clear effect towards a preference of pronoun resolution depending on result, which keeps the default prominence of the subject, and explanation, which promotes the object.

(11) John was disappointed with Tim.
  a. He fired him. RESULT
  b. He disobeyed him. EXPLANATION

However, Stojnić (2021) points out that explanation need not always promote the object. It is therefore proposed that explanation comes in two variants, depending on what argument is promoted to be the most salient one.

(15) The city denied the demonstrators a permit.
  a. They feared violence. $\text{EXPLANATION}_{\text{Subject-based}}$
  b. They advocated violence. $\text{EXPLANATION}_{\text{Object-based}}$

Stojnić adduces contrasts with overtly encoded causal relations as independent evidence for this distinction: similarly to the two different explanation-relations, lexical verbs induce an implicit bias towards the causally implicated agent.[14]

(16) a. Sue frightened Mary because she was boisterous.
  'because Sue is boisterous' (*she* = Sue/subject)
  b. Sue feared Mary because she was boisterous.
  'because Mary is boisterous' (*she* = Mary/object)

Upon closer inspection, however, we find that the preference for pronominal resolution is determined not so much by the higher verb and the agent it causally implicates, but rather by the content of the propositions expressed by the embedded clause (with the pronoun resolved to either participant):

(17) a. Sue frightened Mary because she found her boisterous.
  'because Mary finds Sue boisterous' (*she* = Mary/object)
  b. Sue feared Mary because she found her was boisterous.
  'because Sue finds Mary boisterous' (*she* = Sue/subject)

The *because*-clauses with the evaluative predicate *find* are naturally understood as an explanation for the emotional state that is ascribed to a participant in the matrix clause. The pronoun *she* is thus resolved to the emotionally impacted participant, not the causally implicated one. We thus find the oppositive preferences for the subject pronoun *she* of the embedded clause even though the matrix clause as well as the

[14] In (16), I replace Stojnić's (2021) original choice *scary* with *boisterous* to at least reduce the bias from the lexical predicate which is applied to the pronoun in question.

coherence relation (overtly cued by *because*) have remained the same. This suggests that contrary to our first intuitions about (11), explanation does not determine whether the subject or the object is the most salient referent.

More in general, when playing with the contexts also for these sequences, it is possible to overwrite the initial preferences also for these specific examples while keeping the coherence relations constant. For (11a), imagine that Tim is a very uncertain person and, as a boss, does not take it well if he notices that someone is disappointed with him. John lets it transpire that he is disappointed with his boss, therefore Tim fires him. In this case, result is paired with a resolution of the second sentence's subject to the first sentence's object (contrary to the original intuition for the example). (11b) is harder to understand in a way such that an eventuality of John disobeying Tim is the reason for John's disappointment with Tim. Even if we imagine that John is disappointed with bosses that are weak enough to tolerate disobedience, this situation would be expressed preferably with an overt indication that the disobedience preceded the disappointment as in (18):

(18) John was disappointed with Tim. He had disobeyed him.
EXPLANATION, ok: *he* = John; ok: *he* = Tim

If the pronoun is anchored to the object, speakers apparently do not require this indication of temporal precedence. One speaker, however pointed out that they strongly prefer (18) to (11b) even on the reading discussed in Stojnić (2021), where the subject pronoun of the second sentence is resolved to Tim, the object of the first sentence. These last observations suggest that the initially dispreferred resolution of the pronoun cannot be paired up with a dispreferred resolution of the temporal ordering. While this phenomenon requires further investigation, it puts on the map tenses as yet another player in the game of how to resolve coherence relations and pronominal reference. All in all, the data discussed in this section strongly suggest that, even in English, coherence relations do not asymmetrically determine pronominal resolution.

4.3 *Might Japanese be coherence dominant?*

Stojnić (2021) aims to show that languages differ in the grammatical rules associated with particular coherence relations (specifically their impact on pronoun resolution). She uses this as an argument for the position that coherence relations form part of natural language grammar. In Sect. 4.1, we have seen that the evidence from Serbian is not entirely compelling in this regard. It might be interesting to investigate Japanese as a language which appears to display strong connections between coherence relations and grammatical markers, for instance in the inventory of conditional connectives (Arita 2004; Takubo 2020).

In English, sequences of eventive sentences in simple past tense can be related by EXPLANATION (as in (19b)), which requires the second event

(the pushing) to precede the first (the falling).[15] This contrasts with a discourse related by NARRATION or RESULT, in which the events are presented in the order of occurrence.

(19) Max fell.
a. John helped him up. NARRATION/RESULT
b. John pushed him. EXPLANATION

In Japanese, however, Explanation appears to require overt marking by the copula construction with *no da*, roughly 'it is...' (Kaufmann 2020: 416).

(20) Makkusu ga taore-ta.
Max NOM fall-PST
'Max fell.'
a. Zyon ga osi-ta.
John NOM push-PST
'[Then] John pushed him.' not: EXPLANATION
b. Zyon ga osi-ta no da.
John NOM push-PST NMLZ COP.NPST
'[That's because] John pushed him.' ok: EXPLANATION

Upon closer inspection, however, the issue seems to be with the temporal order of events, and not about a requirement to overtly mark a coherence relation: inserting a perfect marker makes the explanation reading available even without the nominalizer, cf. (21).

(21) Zyon ga osi-tei-ta (no da).
John NOM push-PERF-PST (NMLZ COP.NPST)
'[That's because] John pushed him.' ok: EXPLANATION

In this case, the crosslinguistic difference is probably better placed on the behavior of the tenses than on the behavior of the coherence relations. As evidenced above, a sequence of sentences in the simple past with eventive predicates is naturally interpreted in terms of temporal progression, (Kamp 1979; Kamp and Reyle 1993) but can be interpreted as going back in time to the cause of the event described by the first sentence. The temporal order presumed stands and falls with the choice of coherence relation: RESULT (and NARRATION) require one ordering, EXPLANATION requires the other. Things are different, however, in Japanese. We can make sense of the Japanese data if we assume that event progression is strict, that is, sequences of sentences in the simple past tense shift forward the time at which the events are located (the *reference time*, Reichenbach 1947). This is schematized in (22):

(22) For Japanese:
$Clause_1$-Past. $Clause_2$-Past.
entails: event-time($Clause_1$) < event-time($Clause_2$)

[15] As pointed out at the end of Sect. 4.2, not all speakers of English seem to be equally permissible in this respect.

As explanations cannot temporally follow their explanandum (Asher and Lascarides 2003 for restrictions on the parameters associated with discourse relations), relating these clauses by EXPLANATION is inconsistent with the very interpretation of EXPLANATION.

In contrast, the construction with nominalizer and copula construction (*no da* 'it is (the case) that') but also with the perfect *tei* describe the $Clause_2$-event as concluded at utterance time. The utterance time follows the time at which the event in the first clause is anchored, which renders the sequences in (20b) and (21) felicitous. The apparent effect of coherence relations is thus better understood as deriving from the grammar of tense.[16]

5. *Conclusions*

Stojnić (2021) maintains that the conventionally encoded meaning of complex natural language expressions determines the content conveyed to a much higher degree than what is standardly assumed. Over the course of a conversation, individuals and bodies of information talked about, are tracked in the attentional state, a sequence that records them ordered by prominence as determined (up to a presumedly small remnant of ambiguity) by the linguistic context. Coherence relations as inferred between subsequent sentences are part of the grammar and can directly impact how referents are ordered in the attentional state. Pronouns (and the parameters for the interpretation of modal verbs and adverbs) receive a unform treatment as picking out the highest ranked referents meeting their lexical requirements (as imposed for instance by gender features).

I have argued that Stojnić makes a valid point: data shown in the

[16] While Japanese main clauses thus fail to display a clear effect of coherence relations, the effect might still be real in conditionals. Japanese has a large inventory of conditional connectives, each subject to their own constraints on tenses, temporal order,... (Takubo 2020)

(i) *Mary-ga {ku-ru nara / ki-tara / ku-reba / kuru to / ki-tewa}, John-mo ku-ru.*
Mary-NOM come-NPST NARA / come-TARA / come-BA / come-NPST TO / come-TEWA John-also come-NPST.
'If Mary comes, John also comes.'

Only *nara* is fully felicitous for *indicative backtrackers* as in (ii), that is, conditionals where the antecedent specifies currently existing evidence that the event described in the consequent has taken place.

(ii) Ima sinku-ni koohii magu-ga { ar-u nara / ?ar-eba / ?at-tara}, John-wa
now sink-at coffee mug-nom { be-NPAST nara / be-COND / be-TARA} John-TOP
kesa kokoni ki-ta (hazu-da)
this.morning here come-PAST (must)
'If there is a coffee mug in the sink (right now), John was here this morning.'

In this case, inserting the perfect marker *-tei-* would not resolve the conflict, which suggests that it is not merely an issue of temporal order. Further research is required to fully understand the interaction between Japanese conditional markers and coherence relations.

book and novel data I have added support the finding that pronominal resolution is severely constrained by grammatical phenomena. However, contrary to what is argued in Stojnić (2021), there is no convincing evidence for the assumption that coherence relations directly determine pronominal resolution. The overlooked rules of grammar that Stojnić locates in the coherence relations are better located in the complex pronominal systems and temporal forms that we find in natural languages. Where the grammar of pronouns and tense allows for multiple resolutions, general all-purposes reasoning resolves combinations of coherence relations and pronominal resolutions to determine the most plausible contender if one exists. If several combinations are equally coherent according to independent criteria (as developed for instance in Asher and Lascarides 2003), we are facing an instance of ambiguity. Where, however, one combination is favored by criteria for discourse coherence, it remains to be determined if it is then also conventionally encoded. This is, for instance, the position advocated almost in passing by Asher and Lascarides (2003: 236):

> After all, degree of discourse coherence is a partial rather than a total order, and there could be more than one maximally coherent. update. In such a case, MDC won't pick a unique updated logical form from among the candidates, and this amounts to semantic ambiguity.

The quoted passage suggests that SDRT is committed to the assumption that, up to ambiguity as reflected in equally coherent segmented discourse representations, the propositional content of a declarative utterance[17] is a fact of the matter independent of the speaker's intentions and determined solely by the linguistic material uttered in a given conversational setting. Thanks to the non-monotonic nature of the logic underlying the construction of segmented discourse representations (*glue logic*), contextual knowledge plays a systematic role absent from Stojnić's framework, leaving room for extra-linguistic factors. The two approaches thus differ on how much role is granted to extralinguistic factors (a difference to be explored in more detail), but they agree that the content conveyed is settled objectively (and independently of speaker intentions). A main difference between Stojnić (2021) and SDRT remains in any case that the grammatical principle determining content for SDRT is not the grammar of coherence relations, but an overarching principle of Maximize Coherence.

To what extent we do take content to be determined to this high degree to begin with has to await further investigation. Potentially relevant data might come from considerations of how speakers react to moves in which their referential intentions were misrepresented: as far as clearly semantically encoded content is concerned, they can easily resort to 'but I said that...'. For combinations of pronominal resolu-

[17] Same for its equivalent in interrogatives, e.g. a set of propositions (Hamblin 1973; Karttunen 1977).

tions and coherence relations, it might be more natural to stick to 'but I meant that...'. To the best of my knowledge, data along these lines still remain to be investigated systematically.

References

Altshuler, D. and R. Truswell. 2022. *Coordination and the Syntax – Discourse Interface*. Oxford University Press.

Arita, S. 2004. *(Fu)kanzenjiseisetsu to nihongojoukenbun. ((In)complete tensedness and Japanese conditionals.)*. PhD thesis. Kyoto University.

Asher, N. 1993. *Reference to Abstract Objects in Discourse*. Dordrecht, Boston, and London: Kluwer.

Asher, N. and A. Lascarides. 2003. *Logics of Conversation*. Cambridge: Cambridge University Press.

van den Berg, M. 1996. *Some aspects of the Internal Structure of Discourse. The Dynamics of Nominal Anaphora*. PhD thesis. University of Amsterdam.

Bittner, M. 2011. "Time and modality without tenses or modals." In R. Musan and M. Rathert (eds.). *Tense Across Languages*. Tübingen: Niemeyer.

Bosch, P., Rozario, T., and Zhao, Y. 2003. "Demonstrative pronouns and personal pronouns: German *der* vs. *er*." In *Proceedings of the EACL 2003 Workshop on the Computational Treatment of Anaphora*. Association for Computational Linguistics, East Stroudsburg, 61–68.

Brasoveanu, A. 2006. *Structured Nominal and Modal Reference*. PhD thesis. Rutgers University.

Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris Publications.

Despić, M. 2011. *Syntax in the Absence of Determiner Phrase*. PhD thesis. University of Connecticut.

van Eijck, J. 2001. "Incremental dynamics." *Journal of Logic, Language, and Information* 10 (3): 319–351.

Groenendijk, J. and M. Stokhof. 1991. "Dynamic predicate logic." *Linguistics and Philosophy* 14: 39–100.

Groenendijk, J., Stokhof, M., and Veltman, F. 1996. "Coreference and modality." In S. Lappin (ed.). *Handbook of Contemporary Semantic Theory*. Oxford: Blackwell.

Hamblin, C. L. 1973. "Questions in Montague English." *Foundations of Language* 10: 41–53.

Heim, I. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis. University of Massachusetts, Amherst.

Heim, I. 1988. "On the projection problem for presuppositions." *Proceedings of the West Coast Conference on Formal Linguistics* 2: 114–125. Reprinted in Davis (Ed., 1991).

Heim, I. 1993. "Anaphora and semantic interpretation: A reinterpretation of Reinhart's approach." In O. Percus and U. Sauerland (eds.). *The Interpretive Tract*. MIT Working Papers in Linguistics 25, 205–246.

von Heusinger, K. 2004. "Choice functions and the anaphoric semantics of definite NPs." *Research on Language and Computation* 2: 309–329.

Hinterwimmer, S. 2015. "A unified account of the properties of German demonstrative pronouns." In P. Grosz, P. Patel-Grosz, and I. Yanovich (eds.). *NELS 40: Semantics Workshop on Pronouns*. University of Massachusetts, Graduate Linguistic Student Association, Amherst, 61–107.

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

Jacobson, P. 2012. "Direct compositionality." In W. Hinzen, E. Machery and M. Werning (eds.). *Oxford Handbook of Compositionality*. Oxford University Press.

Jovović, I. 2022. "Condition B and other conditions on pronominal licensing in Serbo-Croatian." *Linguistic Inquiry:* 1–20.

Jovović, I. 2023. *Pronouns at the interface of syntax and discourse structure*. PhD thesis. University of Connecticut.

Kamp, H. 1979. "Events, instants, and temporal reference." In R. Bäuerle, U. Egli, and A. von Stechow (eds.). *Semantics from Different Points of View*. Berlin: Springer.

Kamp, H. and U. Reyle. 1993. *From Discourse to Logic*. Dordrecht: Kluwer.

Karttunen, L. 1977. "Syntax and semantics of questions." *Linguistics and Philosophy* 1 (1): 3–44.

Kaufmann, S. 2020. "Formal treatments of tense and aspect." In W. Jacobsen and Y. Takubo (eds.). *The Handbook of Japanese Semantics and Pragmatics*. Berlin, Boston: Moulton de Gruyter, 371–422.

Kehler, A. 2002. *Coherence, Reference, and the Theory of Grammar*. Stanford: CLSI Publications.

Lillo-Martin, D. and Klima, E. S. 1990. "Pointing out differences: ASL pronouns in syntactic theory." In S. D. Fischer and P. Siple (eds.). *Theoretical Issues in Sign Language Research, Volume 1: Linguistics*. Chicago: The University of Chicago Press, 191–210.

Montalbetti, M. 1984. *After binding: On the interpretation of pronouns*. PhD thesis. MIT.

Murray, S. E. 2014. "Varieties of update." *Semantics and Pragmatics* 7 (2): 1–53.

Nouwen, R. 2003. *Plural Pronominal Anaphora in Context: Dynamic Aspects of Quantification*. PhD thesis. University of Utrecht, The Netherlands.

Patel-Grosz, P. and Grosz, P. G. 2017. "Revisiting pronominal typology." *Linguistic Inquiry* 48 (2): 259–297.

Reichenbach, H. 1947. *Elements of Symbolic Logic*. New York: Free Press.

Riester, A. 2009. "Partial accommodation and activation in definites." In *Proceedings of the 18th International Congress of Linguists*. Seoul: Linguistic Society of Korea, 134 –152.

van der Sandt, R. 1992. "Presupposition projection as anaphora resolution." *Journal of Semantics* 9: 333–377.

Schlenker, P. 2005. "The lazy Frenchman's approach to the subjunctive." In T. Geerts, I. van Ginneken and H. Jacobs (eds.). *Romance languages and linguistic theory 2003: Selected papers from 'Going Romance' 2003, Nijmegen, 20–22 November*. Amsterdam: John Benjamins, 269–310.

Schlenker, P. 2018. "Visible meaning: Sign language and the foundations of semantics." *Theoretical Linguistics* 44 (3–4): 123–208.

Smyth, R. 1994. "Grammatical determinants of ambiguous pronoun resolution." *Journal of Psycholinguistic Research* 23: 197–229.

Takubo, Y. 2020. "Conditionals in Japanese." In W. Jacobsen and Y. Takubo (eds.). *The Handbook of Japanese Semantics and Pragmatics*. Moulton de Gruyter.

Umbach, C. 2002. "(de)accenting definite descriptions." *Theoretical Linguistics* 27: 251–280.