View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Edinburgh Research Explorer

THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Discourse cues: Further evidence for the core contributor distinction

Citation for published version:
Oberlander, J & Moore, J 2006, Discourse cues: Further evidence for the core contributor distinction. vol. 12, 3 edn, Walter de Gruyter., 10.1515/cogl.2002.006

Digital Object Identifier (DOI):
10.1515/cogl.2002.006

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher final version (usually the publisher pdf)

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.

OPEN ACCESS

Download date: 20. Feb. 2015

Discourse cues: Further evidence for the core–contributor distinction*

JON OBERLANDER and JOHANNA D. MOORE

Abstract

Moser and Moore (1995, to appear) carried out a corpus study of discourse cues in tutorial dialogue. Their annotation uses Relational Discourse Analysis (RDA), which distinguishes core *elements (nuclei-like) from* contributors *(satellite-like). In their discussion of these results, Moser and Moore propose that clauses in the contributor–core order are harder to understand than clauses in core–contributor order, but do not attempt to explain why the "hard" order is ever used. Here, we recruit evidence from work by Stevenson and her collaborators, which substantiates the empirical claim. We then suggest that by distinguishing information structure (given–new) from intentional structure (core–contributor), we can explain why hard orders are surprisingly frequent. We note, however, that this cannot be the whole story, and show how the hierarchical RDA structure helps account for differences between discourse cues such as* since, so, this means, *and* therefore.

Keywords: *discourse cues; intentional structure; information structure; tutorial dialogue.*

1. Introduction

Discourse cues are connectives, words, or phrases that link together units in extended, coherent discourse. They include expressions such as *because*, *since*, *so*, *so that*, *this means*, and *therefore*. These discourse cues obviously occur in many different contexts. But why do people choose to use one rather than another, in a given context?

Moser and Moore (1995, to appear) carried out a corpus study of tutorial exchanges, in the domain of electronics problem solving. Their annotation uses a framework known as Relational Discourse Analysis (RDA), which distinguishes *core* elements (nuclei-like) from *contributors* (satellite-like). In their discussion of these results, Moser and Moore

Cognitive Linguistics 12–3 (2001), 317–332 0936–5907/01/0012–0317
© Walter de Gruyter

Brought to you by | University of Edinburgh
Authenticated
Download Date | 2/9/15 12:37 PM

propose that clauses in the contributor–core order are harder to understand than clauses in core–contributor order, but do not attempt to explain why the "hard" order is ever used.

Here, we recruit evidence from psycholinguistic work by Stevenson and her collaborators, which substantiates the empirical claim. We then suggest that by distinguishing information structure (given–new) from intentional structure (core–contributor), we can explain why hard orders are surprisingly frequent. We note, however, that this cannot be the whole story, and returning to corpus analysis, show how the hierarchical RDA structure helps account for differences between the discourse cues of interest.

2. Corpus observations

Moser and Moore (1995, to appear) gathered, annotated, and analyzed a corpus of tutorial dialogues, to help investigate the distribution of discourse cues, such as *because*, *also*, *although*, and *first*. The corpus consists of a series of written question-and-answer exchanges between students and tutors, discussing fault location and repair in the domain of electronic circuitry. Each student was solving an electronics problem using an intelligent tutoring system (SHERLOCK II, Lesgold et al. 1992). At the point in the task when the student requested an explanation, a human tutor produced a written answer, in the place of the usual computer-generated explanation. Thus, the corpus consists of a set of written human–human interactions, and the results reported are based on the analysis of 144 question–answer exchanges, which together contain around 850 clauses.

Example (1) is a relatively brief exchange, but gives a flavor of the type of exchange, the technical domain under discussion, and the types of discourse cue at issue (here, *because*). The student is testing components to locate a fault, and requests advice on strategy:

(1) Student: If the A2A3A5 is OK then, shouldn't I be testing the A2A3A6? I'm just moving back towards the source on the data path.

Tutor: Testing the A2A3A6 is fine, but you are not concentrating on the signal you discovered to be bad. The bad input to pin 21 on the A2A3A5 is the "80's" output of the A2A3A6. You need to find out if the A2A3A6 is receiving an "80's" input. Maybe the A2A3A8 is sending out this bad signal because it is being told to.

Moser and Moore approached this kind of discourse predicting that both the occurrence and placement of discourse cues were correlated with the

Brought to you by | University of Edinburgh
Authenticated
Download Date | 2/9/15 12:37 PM

functional relationship of intentional subordination between discourse units. For them,

> intentional subordination occurs when one discourse unit, the *core*, manifests a communicative purpose and another discourse unit, a contributor, helps to realize that purpose. Cues signal a structural or semantic aspect of how a core and its contributors fit together to form the segment. (to appear: 2)

The idea that speaker intentions dictate the hierarchical structure of discourse, and that the defining feature of a segment is that there be a recognizable segment purpose, is due to Grosz and Sidner (1986). The idea that discourse is hierarchically structured by pairwise relations in which one relatum (called the *nucleus*) is more central to the speaker's purpose than the others (called the *satellites*) is due to Mann and Thompson (1988). Moser and Moore (1996) point out the correspondence between the relation of dominance among intentions in Grosz and Sidner and the nucleus–satellite distinction in Mann and Thompson. In addition, several other models of discourse exploit a similar device (for instance, those due to Hobbs 1985, Polanyi 1988, Redeker 1990). Thus, Moser and Moore's annotation involved coding up all the core–contributor relations in the corpus, both within and between sentences. Note that until section 4.2, all of our examples are cases of sentence-internal relations.

Turning to the analysis, we find that, given a pair of elements, one containing a discourse cue, we can distinguish the linear (first, second) and functional (core, contributor) role of each element. Consider the two-clause examples here, where the recommendation to test is taken to be the core:

(2) a. *Since* Part2 is more likely to be damaged, you should test it first. [Contributor-1]
    b. Part2 is more likely to be damaged, *so* you should test it first. [Core-2]
    c. You should test Part2 first, *because* it is more likely to be damaged. [Contributor-2]
    d. #*So* you should test Part2 first, it is more likely to be damaged. [Core-1]

In (2a) and (2b), the clauses occur in contributor–core order. In (2a), the cue occurs on the contributor; combining order and function, we can call this a *contributor-1* placement. In (2b), the cue occurs on the core, and so we have a *core-2* placement for the cue. In (2c) and (2d), the clauses occur in core–contributor order. In (2c), the cue occurs on the contributor, and thus we have *contributor-2* placement. In (2d), we have placed a cue on the

Brought to you by | University of Edinburgh
Authenticated
Download Date | 2/9/15 12:37 PM

core, attempting a *core-1* placement; however, this particular utterance is ill formed.

Note in passing that in Relational Discourse Analysis what makes one element the core, rather than the contributor is a functional matter, not a semantic one. Operationally, the discourse analyst takes a segment and determines which part of it is the core by asking "what information the hearer is expected to believe or to understand as a result of the segment"; the contributing information is that which is provided by the speaker so that the hearer is more likely to believe or understand the core. Thus, there is no necessary relation between causes and cores or consequences and contributors. The fact that a single causal connective can have both a semantic and a pragmatic reading is related to this; in current terms, the match or mismatch between causal and intentional structure will depend on the communicative intentions as determined in the specific context.

Returning to (2d), it is notable that one of Moser and Moore's main findings was that cues are *never* placed on a core which occurs first (core-1). It is not just that *so* cannot occur as core-1; there is simply no cue that occurs there. When the core comes first, if a cue is used at all, it is placed on the contributor (contributor-2). Table 1 summarizes the distributions.

Note that it is true that (in this and other corpora) cue phrases can occur sentence-initially; for instance, *In order to test a relay you must provide a difference between the relay coil inputs*. However, it is clear that some cues function at a level below the intentional structure defined by core–contributor relations. Arguably, this example falls into that category, since it appears to fail the deletion and replacement diagnostics for detecting intentional substructure (cf. Mann and Thompson 1988 on nuclearity). Although these cues and this level of structure are included in the full RDA analysis, they are not the subject of this paper. It is therefore possible that Moser and Moore's finding generalizes: there is a set of segments in which cues occur sentence-initially, and a set of segments which can be analyzed into a core and contributor(s), with the core first; but these two sets are disjoint.

Table 1. *Distribution of cues with respect to both functional and linear location (*n = *153)*[a]

Functional order of relation	Linear placement of cue		
	First	Second	Both
core–contributor	0	53	0
contributor–core	38	57	5

[a]Reproduced from Moser and Moore (to appear: 51).

Brought to you by | University of Edinburgh
Authenticated
Download Date | 2/9/15 12:37 PM

Moser and Moore also note that discourse cues are in general substantially more likely to occur when the core *follows* the contributor: in these cases, the cue may be placed on either the contributor (contributor-1) or on the core (core-2). Table 2 summarizes the relative frequencies of occurrence versus nonoccurrence. One other point worth mentioning is that they found that particular cues (such as *since*, *so*, *because* or *this means*) have preferred positions, and rarely stray from them. Thus, *since*, *so*, and *because* almost always occur in the positions they occupy in example (2); and *this means* always occurs in core-2 position.

In their discussion of these results, Moser and Moore raise two questions; first:

The origin of order: What explains the relative order of core and contributor?

Their answer to this origin question is that

> In core:contributor order, the context of interpretation for the contributor includes the core. We would expect that the core would help a hearer to understand the contributor by motivating its utterance. In contributor:core order, the context of interpretation for the contributor does not include the core. It may be more difficult to understand a relation in this order. (to appear: 42)

It is the relative difficulty of the latter order that leads speakers to use discourse cues more frequently. It is thus marked in not one, but two senses. Conversely, when the core comes first, a cue indicating forthcoming support for it is superfluous, since whether it's supported or not does not affect its interpretation.

It might be accepted that core-1 cues do not occur, but argued that this is just a corollary of simple syntactic constraints on cues. Thus, ease of processing is immaterial: it is a matter of the resources available in the language. There can be little doubt that there is a significant correlation between syntactic and intentional subordination. However, it is also true that intentional relations can hold between discourse elements that have

Table 2. *Cued relations tended to be in contributor–core order*[a]

Order of relation	Number of relations	
	with cues	overall
core–contributor	53	129
contributor–core	100	157
Total	153	286

[a]Reproduced from Moser and Moore (to appear).

Brought to you by | University of Edinburgh
Authenticated
Download Date | 2/9/15 12:37 PM

no formal syntactic relationship. More importantly, some would argue that form follows function: the syntax reflects the functionality (and hence the psycholinguistic facts), rather than vice versa.

The second question, following from this, is

The existence of marked order: If contributor–core order is harder to understand, why is it ever used?

To this existence question, Moser and Moore attempt no answer. As a first step towards answering it, in the next section, we consider whether there is any empirical support for their answer to the origin question, and in particular, for the supposition that contributor–core order is relatively hard to understand.

3. The origin of order

Reading-time and comprehension studies provide important sources of evidence regarding relative ease of linguistic processing. To address the contributor–core issue, we need to look at data that carries out the right comparisons; in particular, we require experimental materials that manipulate core–contributor order (and thus the connective) while maintaining the materials. The alternation between *so* and *because* we saw earlier provides one kind of test:

(2) b. Part2 is more likely to be damaged, *so* you should test it first. [Core-2]
c. You should test Part2 first, *because* it is more likely to be damaged. [Contributor-2]

In (2b), we encountered a sentence in the (supposedly marked) contributor–core order; in (2c), we encountered the same material re-cast in core–contributor order. If the latter is "easier" to understand, we would predict that it would be read more quickly, and that people would be more likely to understand it correctly. The general question then is: are contributor-2 sentences read faster and understood better than core-2 sentences?

Stevenson and Urbanowicz (1995) carried out a series of reading-time studies which bear directly on this question. They were investigating differing influences on the interpretation of pronouns and proper names in discourse, and comparing effects due to centering (related to order of mention; cf. Grosz, Joshi, and Weinstein 1995) with those due to the thematic roles played by entities mentioned in the discourse, and those due to connectives. The thematic roles under consideration included: goal, source, agent, patient, experiencer, and stimulus. For instance, a sentence's

Brought to you by | University of Edinburgh
Authenticated
Download Date | 2/9/15 12:37 PM

goal is the entity towards which something moves, and the source is the entity from which it must move; a sentence's experiencer is an entity having a given experience, and the stimulus is the entity which gives rise to that experience. Their materials exemplify a range of manipulations, some of which can be seen in (3). Note that subscripts have been added here to help the reader determine referential identity; they were not in the original materials.

(3) a. Malcolm$_1$ won some money from Stuart$_2$ *because* he$_1$ was very good at poker.
b. Malcolm$_1$ won some money from Stuart$_2$ *because* he$_2$ was very bad at poker.
c. Malcolm$_1$ won some money from Stuart$_2$ *so* he$_1$ ended up feeling rich.
d. Malcolm$_1$ won some money from Stuart$_2$ *so* he$_2$ ended up feeling poor.

Sentences like those in (3) place the goal first (and the source second) in the first clause; further materials use verbs (such as *lost to*) which place the goal second (and the source first). The (a) and (b) cases use *because*, and the (c) and (d) cases use *so*; the (a) and (c) cases involve pronominal reference to the goal, and the (b) and (d) cases involve pronominal reference to the source. Further materials make the subsequent references by using repeated names instead of pronouns (e.g., *Malcolm* instead of *he* in [3a]). With all these materials, subjects carried out a self-paced reading task. Each sentence was presented one clause at a time, and after the second clause had been read, a *yes*/*no*-question was presented, the answer to which indicated how the pronoun had been assigned (in the pronoun condition, exemplied here). Time to read the second clause in milliseconds was recorded.

The materials were constructed this way because Stevenson's own hypothesis is that preferences due to centering constraints interact with those due to the thematic roles of the entities referred to. On this view, centering primarily influences *how* an entity introduced in one sentence will be referred to in the next (by pronoun, or by name, for instance); thematic roles influence *which* entities will be subsequently referred to (the goal, or the source, from the first sentence, for instance). Let us now focus on one specific aspect of their results, which bears directly on our current concern.

Analysis of variance showed a significant main effect of connective on reading time: "clauses were read more quickly in *because* than in *so* sentences" (1995: 330): mean times were 1676 milliseconds versus 1926 milliseconds. There was a marginally significant main effect on

Brought to you by | University of Edinburgh
Authenticated
Download Date | 2/9/15 12:37 PM

comprehension accuracy: "There were more correct responses when the questions were connected by *because* rather than *so*' (1995: 331): mean accuracy was 91 percent correct versus 83 percent. The significant reading-time effect was replicated when repeated names were used in place of pronouns: clauses were read faster in the *because* condition: mean times were 1555 versus 1754 milliseconds.

Recall now that the *because* sentences present information in core–contributor order, and *so* presents it in contributor–core order. Stevenson and Urbanowicz's results thus provide initial support for the view that core–contributor order is easier to process, both in terms of speed and accuracy.

4. The existence of marked order

But if contributor–core order is harder to understand, why is it ever used? If we can say (2c), why would we ever say (2b) or (2a)?

(2) a. *Since* part2 is more likely to be damaged, you should test it first. [Contributor-1]
b. Part2 is more likely to be damaged, *so* you should test it first. [Core-2]
c. You should test Part2 first, *because* it is more likely to be damaged. [Contributor-2]

This question can fruitfully be broken down into two parts. First, why use preposed *since* when postponed *because* is available? Secondly, why use *so* when *since* is available?

4.1. *The larger context*

The answer to the first part lies in the information structure of the utterance seen against its wider discourse context. Within a discourse context, the "harder" order may be less coherent locally, but more coherent globally. Thus, although dispreferred on local grounds, it may be required on global grounds.

Elhadad and McKeown (1990) point out that a given–new distinction applies within sentences like (2b) and (2c). In particular, following Halliday, the unmarked position for new information is seen to be towards the end of the sentence. So, sentences usually present information in given–new order. Consider (2b) embedded in a larger discourse context, as shown in (4), and compare it with embedding (2c) in the same context, as in (5):

(4) a. Part2 is moved frequently and liable to damage, but Part1 is not.
b. Part2 is more likely to be damaged, *so* you should test it first.

Brought to you by | University of Edinburgh
Authenticated
Download Date | 2/9/15 12:37 PM

(5) a. Part2 is moved frequently and liable to damage, but Part1 is not.
b. You should test Part2 first, *because* it is more likely to be damaged.

Given this (constructed) context, it seems highly appropriate to say that *Part2 is more likely to be damaged* is given in (4b), and that *you should test it first* is new. But if this is correct, then (5b) presents this same information in new–given order. If information is presented in this order without any fancy syntactic construction (like an *it*-cleft), then arguably the reader or listener will have to carry out extra work. In the absence of information to the contrary, we use the early part of the sentence to provide a link to the prior discourse context, and we then attach the later part of the sentence to this link. If new information is presented first, then extra inferential effort will be required to find a place to link it into the prior discourse—or the listener will just have to wait until the linking context arrives, later in the sentence.

To avoid this extra work, the remedy is that when a proposition is considered new to a discourse, then it should occur later than the given information which links it to the discourse model. If that proposition happens to be a core, then it will be realized as a core in second place. And because (as we have seen) cores are harder to understand when they occur in second place, a range of connectives is available to help readers build an appropriate interpretation of the contributor–core structure in which they appear.

The key point is that intentional subordination varies independently of information structure—information structure in the sense of given versus new information, that is. Moser and Moore (1996) have already elaborated the argument for intentional subordination to be seen as independent of what they term "informational structure"—in the sense of content relations, such as cause versus consequence. What the current discussion underlines is that one cannot account for the distributional facts about discourse cues purely at the intentional level. But what is needed here is an information-structural level—in the sense of one distinguishing given from new information.

4.2. *The remaining question*

Now we can see why *since* or *so* might be better choices than *because* in some circumstances. However, an issue remains. It might be accepted that cores should be placed in second position when the information structure demands it. But there are two ways of signaling this marked structure: by placing the discourse cue at the contributor-1 location (as with *since*), or at the core-2 location (as with *so*). Comparing these options, the former must

Brought to you by | University of Edinburgh
Authenticated
Download Date | 2/9/15 12:37 PM

be considered more co-operative: if the speaker or writer is departing from unmarked order for core and contributor, then the sooner he or she marks the departure, the better. The trouble with *so* is that the listener or reader may misinterpret the initial contributor as a core, until he or she processes the core-2 connective.

Thus, inspection of the tutorial dialogue corpus reveals that *so* is much less frequent than *since*, with just nine occurrences to *since*'s 29. But there are other connectives which should pattern with *so*, including *therefore* and *this means*; taking these into account, there are 32 core-2 cue tokens.[1]

So, the question now is: why do core-2 connectives occur with such frequency, when a contributor-1 alternative is available? The answer to this question does not lie so straight-forwardly in the given–new dichotomy. The corpus reveals interesting differences in the distribution of *so* and *since*. Notably, they tend to occur at different levels in the hierarchy, as defined by core–contributor relationships.

Depth can be measured in terms of the number of relations in the hierarchy directly above the cue; the greater the depth, the more heavily embedded the discourse segment containing the cue. The mean depth for *so* is 1.11, compared with 2.17 for *since*. Examples (6) and (7) are representative of the most common patterns. Recall that the students are solving electronic circuitry problems using an intelligent tutoring system, and then requesting explanations and guidance from a human tutor.

(6) Student: But, what if there was a broken run on high side, but not low side? Or pushed pin?
Tutor: If you would have gotten a short by measuring 60 to 12, this could have told you that your problem was caused by something wrong with the high side and the signal controlling the relay was not a fault. But, you had taken a measurement (A2A3A7 pin 54) that told you the signal controlling the relay was good, *so* there really was not need to measure the low side because you already know there was a problem with the measurement path.

(7) Student: Why?
Tutor: At this point, you know the input to the A1A3A15 is good, yet the output at pins 11 and 12 is bad. The next thing you should check is the inputs to the relays. *Since* you have the A1A3A15 extended from the previous test, you should check the inputs from the A2A3A10 and A2A3A5 before you start checking another card.

However, as Table 3 reminds us, there are just nine tokens of *so* in the corpus. Considering other core-2 cues, we find that *this means*, with eight

Brought to you by | University of Edinburgh
Authenticated
Download Date | 2/9/15 12:37 PM

Table 3.[a]

Type	Cue	Total tokens	Depth		Location		Percentage in final 20%
			mean	s.d.	mean	s.d.	
core-1	*because*	13	1.46	1.13	0.64	0.27	23
contributor-1	*since*	29	2.17	1.04	0.67	0.17	14
core-2	*so*	9	1.11	1.05	0.73	0.25	67
	this means	8	1.50	0.76	0.76	0.26	50
	therefore	15	1.80	1.08	0.68	0.25	27

[a]Some, but not all, core-2 cues tend to occur relatively high up in the hierarchy of the discourse, and relatively late in the discourse. Depth is measured in terms of the number of relations in the hierarchy directly above the cue. Location is measured in terms of the segment in which the cue occurs, expressed as a decimal fraction of the total number of segments in the discourse containing it. The percentage of cues within the final fifth of the discourse is computed from the distribution of locations.

tokens, patterns similarly, with a mean discourse depth of 1.5 (which is similar to that for *because*). However, *therefore*, with 15 tokens, seems closer to *since*, with a mean depth of 1.8.

To explore further these differences, we can track the location of the occurrence of a cue within its discourse context, to determine whether these depth features are related to discourse order. In particular, the segment in which the cue occurs can be located with respect to the overall sequence by expressing its location as a decimal fraction (to one significant figure) of the total number of segments in the discourse containing it. For example, if there were ten segments, and the cue occurred in the fourth segment, its location would be 0.4; if it occurred in the final segment, its location would be 1.0. The percentage of cues within the final fifth of the discourse can then be computed from the distribution of cue locations. Again, if there were ten segments in a single discourse, for this score, we would count those cues which occurred in segments nine or ten, and present this as a percentage of the total number of cues in that discourse. As Table 3 reveals, *since* and *because* have a roughly similar pattern of locations: their mean location is two-thirds of the way through the discourse, and around a fifth of them occur within the final fifth of any discourse. In fact, the slightly greater spread of occurrences of *because* fits with the fact that it is a little more likely than *since* to occur late in the discourse. But in both cases, no more tokens occur late than would be expected by chance.

By contrast, the distribution of both *so* and *this means* is skewed towards the end of the discourse. Their mean position is three-quarters of the way through the discourse, with two-thirds and one half, respectively, of

Brought to you by | University of Edinburgh
Authenticated
Download Date | 2/9/15 12:37 PM

all their tokens occurring within the final fifth. Again, *therefore*, which might be expected to pattern with the other core-2 connectives, ends up with around a third of its tokens within the final fifth: more than *since*, but a much smaller proportion than *so*.

Taken together, this suggests one of two hypotheses about core-2 connectives: the occurrence of such connectives late in, and close to the surface of, a stretch of discourse means that their utterance carries a conventional meaning which outweighs any local processing disadvantage. Indeed, we might generally expect that the beginning and end of a piece of discourse will tend to feature higher level contributions (cf. for instance Moser and Moore [1995] on embedding relations within relations). What we have found here certainly suggests that *so* might be associated with the summing up of a discourse—with spelling out a key point supported by the prior information. Particularly because there will be little further information to absorb after it, the apparent inconsiderateness of a speaker using *so* will be outweighed by the useful information that we are close to the end of the discourse, or at least the end of a substantial segment of it. Another cue phrase, *this means*, seems to have a similar pattern of occurrence, and may therefore fulfill the same function. However, *therefore* seems to occur more evenly throughout discourse, and at a greater hierarchical depth than its core-2 cohorts. Indeed, it seems closer to *since* in its pattern.

Another point of difference between *since* and *so* suggests a second hypothesis about core-2 connectives in general: processing limitations lead to their uncooperative use. The difference relates to syntactic issues touched on earlier. Because *so* and its cohorts are not subordinators, two clauses linked by *so* are essentially independent of one another, and it is therefore possible for them to occur in separate sentences. This freedom is not possible with *since*. This distinction makes a difference in terms of discourse production. Use of *since* to link clauses must be planned ahead, whereas use of *so* can opportunistically link the current clause to a previous one. A subordinator like *since* links cause and effect within a single sentence, but a coordinator like *so* can in principle link to larger segments of discourse. Its occurrence higher up in the discourse hierarchy is mirrored by its syntactic freedom. On this account, the clause preceding a *so* might even have been produced by another speaker.

Therefore, if a speaker is proceeding incrementally, and is not devoting substantial resources to looking ahead (or to revising the sequence of utterances before speaking), or has no control at all over the previous utterance, then it is possible that *so* and its relatives will occur more frequently than might otherwise be expected.

Brought to you by | University of Edinburgh
Authenticated
Download Date | 2/9/15 12:37 PM

It therefore seems that *so* might have a conventional summative function—and thus be deployed by considerate speakers. Or it might be evidence of lack of planning—and thus be deployed only by inconsiderate speakers, or by interrupters (cf. Oberlander 1998). With the relatively small collection of instances of *so* and *since* in the corpus, we cannot properly distinguish these possibilities. One obvious suggestion, however, is that the discourse depth and location of *so* and *this means* seem to fall within a common pattern, consistent with the summative function hypothesis. By contrast, the other main core-2 discourse cue, *therefore*, does not seem to signal the ending of a discourse segment: perhaps its behavior is consistent with the limited processing hypothesis.

One way or another, it seems that the hierarchical structure induced by core–contributor structure can help us explore and account for the occurrence of connectives in core-2 position, in the face of forces favoring the other positions for connectives.

5. Related work

It is interesting to note connections between this account, and work by Spooren (1989) and Noordman (2001), relating to *but* and *although*, respectively. Noordman considers how speakers are likely to continue a discourse after a complex sentence containing one of these connectives. The general account proposes that in the case of *although*, speakers prefer to continue from whichever clause was the main one; however, this preference is moderated by another preference, to continue from the rightmost clause in a complex sentence. This has connections both to the core–contributor distinction and to the information-structural issues already discussed.

One point of connection is that the Moser and Moore corpus study clearly indicates that *though* patterns as Noordman would predict. With six occurrences as a contributor-1 connective like *since*, it introduces an initial subordinate clause. If people prefer to continue from a main clause, particularly a right-hand one, then this helps explain why the *though*-clause is seen as a mere contributor. In parallel, it is notable that *but* behaves as Spooren would predict. With eight occurrences as a core-2 conjunctive connective like *so*, there is no main-clause effect, but there is an ordering effect.

However, there are several differences between the general accounts. First, ours is not tied to sentential syntax. Thus, on the one hand, whether *so* or *but* occur intra- or intersententially is not relevant to our account. On the other hand, it might be fruitful to consider in more detail the relation

Brought to you by | University of Edinburgh
Authenticated
Download Date | 2/9/15 12:37 PM

between syntactic constraints and intentional subordination. Secondly, extending Noordman's account, but still considering only complex sentences, we would expect that the preference to continue from the second clause is stronger in the case of *since* than in the case of *so*, because the main-clause effect strengthens the preference for continuing from the second clause. Recall example (2), repeated here in part, and imagine that subjects are asked to provide continations to discourse fragments like these:

(2) a. *Since* Part2 is more likely to be damaged, you should test it first.
    b. Part2 is more likely to be damaged, *so* you should test it first.

On the extended Noordman account, (2a) will evoke a higher proportion of continuations from *you should test it first* than will (2b). On our own account, there would be no particular difference in the preferences for continuations from (2a) as opposed to (2b).

Unfortunately, our corpus evidence cannot help distinguish these accounts, for a number of reasons that we have already touched on. Most crucially, there are very few continuations of any kind beyond sentences containing *so*. This makes it impossible to compare cases which continue the topic from the *so*-clause with cases which continue the topic from some other main clause or sentence.

One way ahead would be to set out to test some of the core–contributor claims using reading-time experiments. For instance, we contend that the contributor–core order is relatively hard, but that this might be ameliorated by given–new structure in the discourse context. Thus, we could test (*a*) how quickly people read matched constructions such as *A so B* and *B because A* without a prior discourse context; and compare this with (*b*) how quickly they read the same materials in a discourse context C in which A is given, and B is new; and with (*c*) how quickly they read them in another context C′, in which A is new and B is given.

We would predict that in (*a*) *so* is slower than *because*; in (*b*) the difference in speed is reduced or reversed; and in (*c*) the difference is the same as in (*a*) or increased.

6. Conclusion

Constructions like *A so B* might well be less internally coherent than constructions like *B because A*. There is empirical evidence which suggests they are harder to understand; and this supports the view that the difficulty arises because *A so B* is a contributor–core construction, while *B because A* is a core–contributor construction. However, given the right discourse

Brought to you by | University of Edinburgh
Authenticated
Download Date | 2/9/15 12:37 PM

context, *A so B* can have an information structure which coheres with that context where *B because A*'s information structure proves less coherent (and vice versa for a different context). This suggestion is empirically testable. If it is correct, then it helps explain Moser and Moore's puzzle; it would explain why contributor–core constructions—as in *A so B*—are surprisingly frequent, and demonstrably effective.

More generally, the core–contributor framework helps us explore differences in the patterns of behavior of closely related connectives, like *since*, *so*, *this means*, and *therefore*. In every case, attention to the broader discourse context makes a difference. So we conclude with a general moral. As Altmann and Steedman (1988) have said: there is no such thing as a null context.

Received 10 November 2000
Revision received 12 April 2001

University of Edinburgh

Notes

* We are grateful to the audience at the workshop on "Levels of Representation in Discourse" held in Edinburgh in July 1999 for interesting discussions, and to our two anonymous referees for their helpful comments on an earlier draft of the article. Authors' e-mail addresses: J.Oberlander@ed.ac.uk; J.Moore@ed.ac.uk. Both authors work at the Division of Informatics, University of Edinburgh.

1. In this connection, it is worth noting that the corpus does contain five tokens of *so* which are not classifiable as core-2. However, every one of these tokens in fact occurs as part of the cue phrase *so that*. This unambiguously signals a purpose clause, and according to the RDA analysis, neither the matrix nor the subordinate clause is core. Our subsequent discussion therefore excludes these instances. It is worth noting in passing that Stevenson, Knott, Oberlander, and McDonald (2000) analyze data on subjects' continuations of fragments ending with *so* and show that, with transfer verbs only, the *so* is taken to introduce a purpose clause. Nevertheless, in the current corpus, *so* on its own is never analyzable in this way.

References

Altmann, Gerry T. M. and Mark J. Steedman
1988 Interaction with context during human sentence processing. *Cognition* 30, 191–238.

Elhadad, Michael and Kathleen R. McKeown
1990 Generating connectives. In Karlgren, H. (ed.), *Proceedings of COLING90* vol. 3. Helsinki, 97–101.

Grosz, Barbara J. and Candace L. Sidner
1986 Attention, intentions, and the structure of discourse. *Computational Linguistics* 12, 175–203.

Grosz, Barbara J., Aravind K. Joshi, and Scott Weinstein.
1995 Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21(2), 203–225.

Brought to you by | University of Edinburgh
Authenticated
Download Date | 2/9/15 12:37 PM

Hobbs, Jerry R.
1985 On the coherence and structure of discourse. Report No. CSLI-85-37, Center for the Study of Language and Information, October 1985.

Lesgold, Alan M., Susanne Lajoie, Marilyn Bunzo, and Gary Eggan
1992 Sherlock: A coached practice environment for an electronics troubleshooting job. In Larkin, J. H. and R. W. Chabay (eds.), *Intelligent tutoring systems: Shared goals and complementary approaches*. Hillsdale, NJ: Erlbaum, 201–238.

Mann, William C. and Sandra A. Thompson
1988 Rhetorical structure theory: A theory of text organization. *Text* 8 (3), 243–281.

Moser, Megan and Johanna D. Moore
1995 Investigating cue placement and selection in tutorial discourse. In Uszkoreit, H. (ed.), *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, 130–135.
1996 Toward a synthesis of two accounts of discourse structure. *Computational Linguistics* 22 (3), 409–420.
to appear On the correlation of cues with discourse structure: Results from a corpus study. Submitted to *Discourse Processes*.

Noordman, Leo G. M.
2001 On the production of causal-contrastive *although*-sentences in context. In Sanders, T., J. Schilperoord, and W. Spooren (eds.), *Text Representation: Linguistic and Psycholinguistic Aspects*. Amsterdam: Benjamins, 153–181.

Oberlander, Jon
1998 Do the right thing … but expect the unexpected. *Computational Linguistics* 24, 501–507.

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

Redeker, Gisela
1990 Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics* 14, 367–381.

Spooren, Wilbert
1989 Some aspects of the form and interpretation of global contrastive coherence relations. Unpublished Ph.D. dissertation, Katholieke Universiteit Nijmegen.

Stevenson, Rosemary J. and Agnieszka J. Urbanowicz
1995 Structural focusing, thematic role focusing and the comprehension of pronouns. In Moore, J. D. and J. F. Lehman (eds.), *Proceedings of the 17th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum. 328–332.

Stevenson, Rosemary J., Alistair Knott, Jon Oberlander, and Sharon McDonald
2000 Interpreting pronouns and connectives: Interactions among focusing, thematic roles and coherence relations. *Language and Cognitive Processes* 15, 225–262.

Brought to you by | University of Edinburgh
Authenticated
Download Date | 2/9/15 12:37 PM