Discussion

Discourse coherence is an independent notion: 
A reply to Deirdre Wilson☆

Rachel Giora*

Department of Linguistics, Tel Aviv University, Tel Aviv 69978, Israel

Received March 1977

Abstract

The purpose of this reply is to further substantiate the claim that discourse coherence is an independent notion (cf. Giora, 1997). To do that, I show that discourses may be relevant in Sperber and Wilson's (1986/1995) sense, but incoherent. Such discourses challenge the claim that discourse coherence may be derived from relevance (as argued by e.g., Sperber and Wilson, 1986/1995; Wilson, 1998). I further question the psychological reality of Sperber and Wilson's notion of relevance and the processing model their theory assumes.

1. Discourse coherence is an independent notion

From its inception, relevance theory (e.g., Sperber and Wilson 1986/1995, henceforth S&W) has been offered as an alternative account to the coherence-based view of discourse (e.g., Grice, 1975). "The main aim of relevance theory in the domain of verbal communication is to explain how utterances are understood. As a by-product, it claims to shed light on certain intuitions regarding discourse coherence" (Wilson, 1998: 58, emphasis added). Upon this view, we do not need an independent notion of discourse coherence, since "[t]opic-relevant utterances are only a subset of relevant utterances, and it is the notion of topic relevance which is derivative" (S&W: 217). In Giora (1997a), I argued against the view that coherence is a dependent notion by pointing out the dissociation between coherence and S&W's notion of relevance (henceforth S&W relevance). I showed that a discourse may be S&W relevant to an individual, but nevertheless judged incoherent by the very same individual, and vice versa: a discourse may be S&W irrelevant to an individual, but...
nevertheless judged coherent by that individual. Such evidence disputes the claim that coherence is a derivative notion.

Wilson (1998) has not disproved this dissociation. On the contrary, to judge from the kinds of contexts she constructs for her examples (13) and (14), she must accept the dissociation between S&W relevance and coherence: “[T]here are”, she says, “circumstances in which (13) or (14) would be both acceptable and consistent with the principle of relevance [i.e., S&W relevant, RG]. I will consider two types of case: in the first, their discourse segments will be intuitively unrelated [i.e., incoherent, RG]; in the second, they are intuitively related [i.e., coherent, RG] ... There are also several different types of case in which (13) and (14) would be unacceptable and inconsistent with the principle of relevance [i.e., S&W irrelevant, RG]. I will consider two. In the first, the two discourse segments are intuitively related [i.e., coherent, RG]; in the second, they are not [i.e., incoherent, RG]” (1998: 65–66). For instance, Case C is “unacceptable and inconsistent with the principle of relevance [i.e., S&W irrelevant, RG], despite the fact that its discourse segments are related [i.e., coherent RG]” (1998: 67). If a discourse may be S&W relevant but incoherent (and vice versa), then, it only follows that coherence is not a derivative notion.

Consider the following examples that might further substantiate the independence of the notion of coherence of S&W’s relevance notion. The first involves an utterance that may be equally coherent to two different audiences, yet more relevant to one of them. Wilson’s (1998) example (7), on an alternative analysis, may serve to illustrate this dissociation. Recall that (7), repeated here in (1), is said by the games teacher to Susan (who plays for the school cricket team but also brings her pet bats to school to tease her teacher) and her classmates:

(1) Please leave your bats at home tomorrow.

On hearing (1), each and every student of the class accesses assumption (a) to render (1) S&W relevant:

(a) I bring my (cricket) bats to class on certain days.

However, on hearing (1), Susan simultaneously accesses two assumptions – (a) and (b):

(b) I bring my (pet) bats to class on certain days.

For Susan, (a) and (b) are equally accessible, hence there is no reason to assume that (b) should be activated before (a) (for simultaneous activation of equally salient

---

1 Wilson (1998) rephrases this argument, substituting ‘optimal relevance’ and then ‘consistent with the principle of relevance’ for ‘S&W relevance’. This, however, does not alter the argument.

2 Note that two adjacent sentences are considered related (i.e., coherent) if they are related via a governing discourse-topic.

Moreover, (1) poses a further problem to a S&W account. There is no reason to dismiss the possibility that the ambiguity of (1) is actually consistent with the principle of relevance. On the contrary, the speaker could be highly S&W relevant to Susan at almost no cost, and could indeed intend her to process the ambiguity and derive the two resulting contextual implications. On this analysis, the speaker affects Susan’s cognitive environment to a larger extent than she affects those of the rest of her classmates. Since Susan does not have to make any extra effort accessing two assumptions simultaneously, the speaker is more relevant to her than to the rest, though to the audiences as a whole, the speaker is equally coherent.

But even on Wilson’s analysis (whereupon Susan first accesses the meaning of pet bats, but rejects it in favor of the meaning of cricket bats), the dissociation between coherence and S&W relevance is obvious. On Wilson’s analysis, while the utterance is equally coherent to all of the audience (including Susan), it is less relevant to Susan, since it is more effort consuming: Susan makes an extra effort, activating and suppressing an unnecessary assumption (upon which bats means her pets) in order to derive the speaker’s intended meaning (upon which she should not bring her cricket bats). Here, then, coherence and S&W relevance do not coincide.

As an example of a coherent but S&W irrelevant discourse, consider, again, the discourse cited in Giora (1997a: 24). Recall that although it was manifest to my students and to me that I would start my lecture by stating that I was leaving the university soon, and hence would propose a make-up-lecture timetable, I did not omit this mutually evident (i.e., S&W irrelevant) information as predicted by S&W. Rather, I started by stating the discourse-topic, marking it as old information (as required by coherence requirements): “As you well know, I am leaving soon, etc.” In her reply, Wilson (1998) suggests that, contrary to my view, the discourse in question is, after all, S&W relevant, because it confirms the addressees’ expectations. According to her, my students expected my leaving and expected me to announce it. On this view, at the start of my class, my students had two weak assumptions which I reinforced by confirming. But this was not the case. I was entirely uninformative in stating that I was about to leave; the students in question needed no confirmation for this ‘old’ piece of news. I was also entirely predictable in making this announcement. I was not even refreshing their memory – evoking the set of assumptions necessary for the comprehension of the discourse that followed. All I was doing was starting a new discourse segment by signaling that it was distinct from the informal conversations we had conducted during the break. Though the discourse topic was old information, and though my students were aware of the fact that the context has changed (which makes the need to alert them to that fact redundant), the messages added to it (e.g., the list of make-up lectures) conformed to the graded informativeness requirement (Giora, 1988). Here, again, the example testifies to the dissociation of coherence and S&W relevance.

For evidence that suppression is effortful, see e.g., Gernsbacher et al. (1990, 1995), Keysar (1994), Gernsbacher and Robertson (1995).
Along the same lines, a lecture in physics may be entirely S&W irrelevant to me, but I would be able to judge it as coherent; the filibusterer may be entirely S&W irrelevant, but that doesn’t mean she does not emit well-formed discourses (S&W: 159). And it is precisely the well-formedness of her discourse which grants her the right to S&W irrelevance. Had she been incoherent, in addition to being S&W irrelevant, she would not have been allowed to keep talking.

However, to falsify the claim that discourse coherence is a derivative notion, it is sufficient to demonstrate that a S&W relevant discourse may be incoherent. Recall, first, that Wilson herself (1998) provides us with such examples: On her analysis, her examples (13) and (14) are discourses which are acceptable and consistent with the principle of relevance, but in one case their discourse segments are intuitively unrelated (i.e., incoherent).

As an illustrative example consider Smith and Wilson’s (1992: 5), in which someone walks into a lecture announcing that the building is on fire. Inasmuch as this utterance is incoherent – being unrelated to the lecture topic under discussion – it is highly S&W relevant.

Or, consider, again, the illustrative instance (2b) provided by Blass (1990: 22), which lends itself to two interpretations, both consistent with the principle of S&W relevance, though one (the mention of Susan’s assertion) is more coherent than the other (the speaker’s assertion):

(2) a. What did Susan say?
   b. You have dropped your purse.

Evidence from such S&W relevant but incoherent discourses disputes the claim that discourse coherence is a derivative notion. Though coherence and S&W relevance often coincide (i.e., discourses that are S&W relevant tend to be coherent and vice versa), there is, nevertheless, a theoretical interest in teasing them apart.

1.1. On the psychological reality of S&W relevance

If coherence derives from S&W relevance and is not an independent notion, then we should either have intuitions about S&W relevance only, or our intuitions regarding coherence and S&W relevance should not be distinguishable. In Giora (1997a), I argued that we do not have intuitions about S&W relevance, but that we do about discourse coherence. As evidence, I cited S&W’s examples (3–5 below), which I will repeat here for convenience. Recall that S&W use these utterances to illustrate their notion of relevance in context. They claim that our intuitions that (4) is more S&W relevant than (3) and that (5) is less S&W relevant than (3) in the context of (a–c) are accounted for in terms of contextual effects measured against processing effort (S&W: 126–127):

(a) People who are getting married should consult a doctor about possible hereditary risks to their children.
(b) Two people both of whom have thalassemia should be warned against having children.
On S&W's analysis, (3) and (4) are equally difficult to process, because they are similar in length and require the same context (a–c). However, while (3) has only one contextual implication, i.e., (6), (4) has an additional contextual implication, (7):

(6) Susan and Bill should consult a doctor about possible hereditary risks to their children.
(7) Susan and Bill should be warned against having children.

In Giora (1997a: 19–22), I argued that, on the assumption that contexts are searched for, the context needed to render (3) S&W relevant is smaller than that needed to render (4) S&W relevant. To render (3) S&W relevant, only two assumptions (a, c) should be activated. However, this is not true of (4). On processing (4), (b) should be added to the context. (4) then, is not really more relevant than (3), as S&W tried to show. Though (4) has more contextual implications than (3), it is also more effort consuming. If we have intuitions that (3) and (4) are somewhat distinct, they are not accountable in terms of S&W's relevance notion.

However, even if, for the sake of the argument, we forgo for a while the idea that contexts are searched for, (3) is not more relevant than (5): while (3) has one contextual implication, (6), processed at the cost of accessing three assumptions, (a–c), (5) has two contextual implications, (6) and (7), processed at the cost of accessing four assumptions – (a–c), and an additional proposition (that 1976 was a great year for French wines). By simple calculation, (3) is 0.3 S&W relevant, while (5) is 0.5 S&W relevant. How is it possible that we have intuitions that (3) is more relevant than (5) in the above context? I suggest that it is not the case that we have intuitions about S&W relevance, but rather about coherence. (3) is indeed more coherent than (5) in the above context. Whereas (5) comprises a message that is unrelated to the discourse topic discussed, (3) does not (cf. Giora, 1985a,b).

Recall, once again, Blass' example (2 above). If coherence is a derivative notion, then both readings, the one upon which (2b) is both coherent and S&W relevant and the one upon which it is S&W relevant but less incoherent, should induce the same intuition. However, if coherence is independent of S&W relevance, then the two readings should induce different intuitions. It seems quite obvious that the latter holds. Thus, if our intuitions about the reading in which S&W relevance and coherence do not coincide are different from those regarding the reading on which they do, then coherence and S&W relevance are independent notions. Though, on the

---

4 I owe this remark to Mordechai Ben-Ari (personal communication).
basis of this example, it is not quite clear that we do not have intuitions about S&W relevance, it is quite clear that we do have intuitions about coherence.\textsuperscript{5}

2. The relevance requirement

Wilson (1998) also criticizes the requirements for discourse coherence that I have proposed. According to the relevance requirement (Giora, 1985a,b), a discourse segment is coherent iff its various propositions are either related to a discourse topic, preferably mentioned and placed in the beginning of the discourse, or marked as digressing from relevance (see also Giora, 1997a: 22–23). According to this view, a topic switch, as in the assertive reading of (2b), can be alleviated by an explicit (either verbal or nonverbal) digressive marker, such as \textit{oops}, which cues the addressee as to the change of topic.

How many marked digressions are acceptable? The amount of digression and length of digression affect discourse coherence. A discourse segment that abounds in (marked) digressions is bound to be incoherent, because it violates the first clause of the relevance requirement. As shown by Wilson’s (1998) example, it is not clear what discourse topic is discussed in the following:

(8) Today is Tuesday. Incidentally, I can dance the tango. By the way, there are tadpoles in my village pond, etc.

Recall that coherence is not achieved by local connectedness between adjacent sentences, but by a relation to a governing discourse-topic (Giora, 1985a). (8) obviously fails to have one. Wilson is right in pointing out that the number of marked digressions should be controlled by the first clause of the relevance requirement.

Indeed, in Giora (1991, 1993), I have shown that discourses comprising marked digressions (from either relevance or informativeness) are more difficult to understand than the same discourses without (the marked) digression. I have also shown that the longer the digression, the greater the difficulty. Though marking the digression alleviates the relative incoherence, a discourse containing a digression is relatively marked compared to a discourse that does not contain digressions. However, such a discourse is more coherent than discourses which do not mark their digressions. Recall that although these discourses are considerably less coherent, they may nonetheless be S&W relevant, e.g., (2b).\textsuperscript{6}

\textsuperscript{5} However, if my analysis of this example (Giora, 1997a: 31) is correct, and the less-coherent reading of (2b) is also less relevant, because the context switch is effortful, then my analysis here will not support the dissociation hypothesis.

\textsuperscript{6} Note, also, that the choice of digressive markers is not arbitrary, but constrained by the type of digression. The more drastic the digression, the more explicit, emphatic, informative, or apologetic the digressive marker.
3. Segmentation and the graded informativeness requirement

Wilson (1998) further questions my proposal that discourse segmentation is discourse-topic oriented. Upon the theory of coherence I propose, a formal segment marks the boundaries of a segment topic. Segmentation occurs either before or immediately after the introduction of a new segment topic (Giora, 1983a,b, 1988; Giora and Lee, 1996). In both cases, it occurs after the most informative but still relevant message in a given segment has been introduced. This relevant and most informative message can both mark the boundary of the given discourse segment and, at the same time, constitute the next segment topic.

Are readers sensitive to discourse segmentation? Does segmentation affect their coherence judgments? According to my findings, the answers to these questions are affirmative. For instance, in Giora and Lee (1996), we show that subjects’ paragraphing coincides with authors’. In Giora (1997a), the subjects I presented passages (9) and (10) below were also sensitive to paragraphing. They judged (10) as more coherent than (9). Indeed, (10) conforms to the graded informativeness requirement, ending with the most informative message relative to the discourse topic mentioned in its beginning. Wilson (1998), however, suggests that the subjects’ judgment might have been affected by paragraph marking rather than by coherence considerations. She cites Stark (1988) and Hofmann (1989), who have shown that a full NP often marks initial paragraph boundaries. According to Wilson, then, my subjects might have been guided by the paragraph marker rather than by the paragraph content, i.e., by its information structure.

Indeed, full NPs mark segment initial boundaries, but they also mark segment final boundaries (e.g., Longacre, 1979). Thus, if my subjects had been insensitive to the contents, i.e., to the graded informativeness structure of the text, as suggested by Wilson, they should have seen no difference between (9) and (10), because in both the paragraph ends with a mention of a full NP (penicillin).

(9) It has often occurred in the history of science that an important discovery was come upon by chance. A scientist looking into one matter, unexpectedly came upon another which was far more important than the one he was looking into. Penicillin is a result of such a discovery. Penicillin was accidentally discovered by Fleming in 1928 ...  

(10) It has often occurred in the history of science that an important discovery was come upon by chance. A scientist looking into one matter, unexpectedly came upon another which was far more important than the one he was looking into. Penicillin is a result of such a discovery.

Penicillin was accidentally discovered by Fleming in 1928 ...

Contrary to Wilson’s assumption, the use of a lower accessibility marker than required by the informativity consideration in paragraph initial position (e.g., the use of *penicillin* in the beginning of the second paragraph in (10)) is predicted by the theory of coherence in question. It evinces, among other things, that the new discourse topic is not an NP, but a full proposition, as argued by Giora (1985a) and Enq (1986).
Note that marking paragraph boundaries by highly informative, or low accessibility markers (cf. Ariel, 1990, 1991), where less informative markers could be used, goes beyond the occurrence of full NPs. For instance, in Giora and Lee (1996), we attest to the role of full pronouns as opposed to zero pronouns in marking boundaries in Mandarin Chinese. More relevant to our discussion here is our analysis of Li and Thompson’s (1979) findings. We show that their findings demonstrate that subjects are sensitive to discourse topic boundaries. In their study, Li and Thompson presented subjects with unparagraphed Mandarin texts. They removed all third person pronouns (a lower accessibility marker than a zero pronoun), and asked their subjects to insert pronouns where they felt they were needed. As predicted by Giora and Lee, full pronouns were realized in the beginnings, but also at the ends of discourse topic segments. This shows that comprehenders are sensitive to paragraphing even in the absence of overt signals. Moreover, it suggests that the repetition of the full NP (penicillin) could just as well function to signal the end of the paragraph. It is, therefore, more plausible to assume that the subjects who found (10) more coherent than (9) were not blind to the discourse structure.

Wilson (1998), however, argues that paragraphing cannot be explained in terms of a coherence-based theory. As evidence against a coherence-based theory, she cites Stark (1988), who showed that absence of paragraphing and arbitrary paragraphing did not affect readers’ coherence judgments. Unfortunately, however, Stark’s findings cannot support a S&W relevance-based theory either. According to Wilson, “the processes that go around paragraph breaks (anaphoric processes, intonation changes, pauses, slowed reading speeds, etc.) save the addressee wasted effort by alerting him to the fact that a switch in contexts is about to take place” (1998: 71). A S&W relevance-based theory thus predicts shorter reading times for a paragraphed than for unparagraphed or arbitrarily paragraphed text. Contrary to predictions, however, Stark’s (1988) subjects did not take less time to read the appropriately paragraphed text versions. Thus, Stark’s findings cannot support relevance theory any more than they do the coherence theory in question. Incidentally, Stark herself questions her own results.

Let us, however, examine the claim that the subjects’ preference of (10) over (9) does not pose any difficulty for relevance theory, since, according to that theory, paragraphing markers alert the reader to the oncoming change of context, thereby saving him wasted effort. Indeed, alerting a reader to a change of context seems to constitute a reasonable motivation for paragraphing, since a new discourse-topic, allegedly necessitating a switch of contexts, is introduced. However, a close examination of the structure of (10) reveals that (10) is a discourse sample in which formal segmentation occurs after the new discourse topic is introduced. The second paragraph of (10) discusses the discourse topic (i.e., the accidental discovery of penicillin) placed in the end of the first paragraph. Such paragraphing cannot alert the reader that a context switch is about to take place. Why then mark the beginning of

---

8 See also Schiffman (1984) for a preference for that (a relatively low accessibility marker) over it (a relatively high accessibility marker) in initial and final mentions.
the next paragraph? According to the coherence theory I propose (Giora, 1983a,b, 1986, 1988), the next paragraph is formally marked so as to warn comprehenders against processing this new paragraph in relation to the previous paragraph topic (i.e., important scientific discoveries happening by chance).

Note that, perhaps contrary to appearances, I do not agree with Chafe’s (1987) assumption (cited in Wilson, 1998: 71) that “[i]n a well-organized text, each utterance within a given paragraph should be interpretable in a context determined by the interpretation of its immediate predecessor”. Instead, I propose that in a well-organized text, each utterance within a given paragraph is interpretable in relation to the paragraph topic. The apparent coherence relations between adjacent propositions are only a by-product of the coherence relation to a governing discourse-topic (Giora, 1985a). Though in the discourse in (10), the first proposition of the second paragraph appears interpretable in relation to its immediately preceding proposition, it is to its paragraph topic proposition that it is related, which is placed at the end of the preceding paragraph. As noted, such paragraphing does not coincide with a context switch and, consequently, cannot alert comprehenders that a context switch is imminent. It does, however, warn them against relating the oncoming messages to the previous discourse topic.

4. What processing model does relevance theory assume?

If S&W really propose an alternative to Grice’s coherence-based theory, the processing model they assume must differ from that of Grice’s (1975). While according to Grice’s account, interlocutors must know the norms of communication, must be able to detect an overt violation of the norms, and derive an implicature as a result, according to S&W, interlocutors “need no more to know the principle of relevance to communicate than they need to know the principles of genetics to reproduce. Communicators do not ‘follow’ the principle of relevance; and they could not violate it even if they wanted to. The principle of relevance applies without exception” (p. 162). Thus, according to Grice, if language is used nonliterally, i.e., overtly violating norms of communication, comprehenders will be engaged in a multiple-stage process (dubbed ‘dual/sequential/special process’). They will process the literal meaning first, then reject it as the intended meaning and derive the nonliteral intended meaning.

Upon S&W’s account, nonliteral meanings should be accessed directly, without comprehenders having to process the unintended literal meaning first (see also Gibbs, 1986). The idea is that contexts are searched for, rather than given; hence comprehenders access (only) those assumptions that render an utterance S&W relevant. Since the literal meaning is S&W irrelevant, i.e., it may not enrich the comprehender’s set of assumptions with implications consistent with the principle of relevance, it need not be processed. As illustration, consider the process assumed by S&W to underlie the comprehension of the metaphor **His ink is pale**: “A strictly literal construal of this utterance is clearly ruled out: it is hard to see what relevance could attach to knowing the color of a poet’s ink. Nor is there any strong implica-
ture. The only way of establishing the relevance of this utterance is to look for a wide range of very weak implicatures” (p. 237).

Similarly, according to Vicente (1996), the literal meaning of utterances intended to be interpreted nonliterally is not accessed and rejected (as Grice would predict); “In relevance theory terms the propositions expressed by these utterances are not communicated ... [T]he same applies to other logical and contextual implications that would be derived on a strictly literal interpretation but whose derivation is inhibited when the utterance is being interpreted ‘loosely’. Interpretation is therefore not delayed in any way by the fact that the proposition expressed by the utterance is false” (1996: 202–203). Relevance theory, then, seems to dispense with the sequential process.

This indeed is an alternative to Grice. According to S&W, then, but contra Grice, figurative language “requires no special interpretative abilities or procedures: it is a natural outcome of some very general abilities and procedures used in verbal communication” (1986: 237). Indeed, for over 15 years, psycholinguists and cognitive psychologists have adduced evidence in favor of equivalent processing for literal and nonliteral language (e.g., Ortony et al., 1978; Gibbs, 1982, 1984, 1986, 1994; Hoffman and Kemper, 1987; Gibbs and Gerrig, 1989; Glucksberg, 1989; Keysar, 1989, 1994; Glucksberg and Keyser, 1990; Keysar and Glucksberg, 1992; Glucksberg et al., 1982; Gildea and Glucksberg 1983; but see Dascal, 1987, 1989; Giora, 1997b, and Giora and Fein, forthcoming(a,b) for a critique of the equivalence processing hypothesis).

However, it is not quite clear that this is the processing model relevance theory actually assumes. More often than not, it seems that what is assumed to be a direct process is, in fact, a ‘dual process’. For instance, according to S&W, irony is an instance of “a garden-path utterance, likely to cause the reader momentary processing difficulties later offset by appropriate rewards. One first reads it as an ordinary assertion, is led to the absurd conclusion that ... and only then reinterprets echoically” (p. 242, emphasis added. For a similar processing assumption, see Wilson, 1998, regarding the analysis of example (7); Papafragou, 1996: 184, regarding processing of metonymy). More specifically, S&W claim that “[C]ontrary to first appearances, the principle of relevance does make it possible to use an item by item testing strategy in comprehension” (p. 170). This is precisely the processing model assumed by Grice and his followers regarding e.g., irony comprehension (e.g., Giora, 1995; Giora and Fein, forthcoming(a); Giora et al., forthcoming). Whither the difference, then?

In sum, despite its merit as a comprehension model, relevance theory cannot totally dispense with coherence-based theories. The notion of relevance to the individual, as valuable as it is, cannot subsume the notion of discourse coherence.9

---

9 It is also unclear to me how the direct process approach here can be reconciled with S&W’s commitment to modularity. Recall that the linguistic module automatically produces the linguistic meaning of propositions, and, in most cases, this coincides with the literal meaning S&W claim the addressee does not process.

10 For an attempt to reconcile the direct vs. sequential processes see Giora (1997b).
References