

The Multifunctionality of Discourse Markers

Eduard H. Hovy

Information Sciences Institute
of the University of Southern California
4676 Admiralty Way
Marina del Rey, CA 90292-6695
U.S.A.
tel: +1-310-822-1511
fax: +1-310-823-6714
email: hovy@isi.edu

November 1994

Abstract

This paper discusses the need for and nature of multifunctionality of discourse markers, signalling in parallel several simultaneous structures that underlie coherent discourse. Arguing that any adequate description of discourse requires at least four distinct structural analyses — semantic, interpersonal/goal-oriented, attentional/thematic, and rhetorical — one has to address the questions: what are the individual functions of these different structures, how do they interact, and how are they expressed in the text? With respect to the last question, it is clear that when constructing the discourse, the speaker has to select, from the available structuring cues or markers, the one(s) that minimise the overall structural ambiguity for the hearer. When included in the rhetorical structure, and hence in the text, these cues or markers assist the hearer in decoding the speaker's message into the various parallel structures, resulting in effective communication. Three collections of cues (semantic, interpersonal, and rhetorical) are given, and the overloading of meaning for rhetorical cues is discussed.

1 Introduction: Why is there Syntax?

I would like to start this paper setting up an analogy between discourse and syntax, which necessitates an assertion about syntax and then a question. In the following section, the analogy is extended to the subject of discourse structure and markers.

First, the assertion. I take it as obvious that, with language, we communicate more than semantics; in a single utterance, we can convey interpersonal, situational, and other information at the same time. Of the major types of information speakers convey, the semantic (propositional) is usually most apparent, since it is usually contained in the surface

content of the message. The interpersonal and situational meanings are mostly expressed indirectly, through text style, with occasional appropriate lexical insertions. This idea has been mentioned in many places, notably in [Halliday 85].

Now the question: Why is there syntax? That is, *why do languages all exhibit syntactic structure as a separate, autonomous construct?* The typical linguistic model of single sentences includes two principal structures: a structure that houses the semantic information (usually called the semantic structure, the f-structure in LFG [Bresnan 78], or possibly the deep structure [Chomsky 65]; the distinction is not relevant here) and a syntactic structure that expresses the surface form of the sentence. But why should there be two structures? If you have a well-formed semantic structure, possibly something like a case frame or a set of connected knowledge base assertions, and you define a regular traversal algorithm, possibly a depth-first left-to-right strategy, then for the surface sentence you can simply produce a string of pairs: semantic function and filler. In fact, most languages have almost this form: English and Dutch (with pairs of preposition and filler) somewhat so; Japanese (with pairs of case marker and filler) more so. Under this view there is no need for a distinct surface structure — the whole semantic structure is straightforwardly recoverable from the sentence itself. Why then have all languages evolved a syntax?

One may surmise: the syntax of the language is nothing other than the trace of the traversal of the semantic structure.

Unfortunately, however, this cannot hold: in general, the syntactic structure and the semantic structure are not isomorphic. Regardless of what theories of syntax and semantics one follows, non-isomorphism would be admitted, I believe, by all students of language.

If the structures are not isomorphic, then they contain different information. In particular, either one structure holds more information than the other, or else the structures are simply mutually unrelated in some places; that is:

- *either* semantic structure holds more information than syntactic structure;
- *or* syntactic structure holds more information than semantic structure;
- *or* each structure holds some information not in the other.

Ignoring questions of anaphora (in which a part of the syntactic structure holds less information than the corresponding part of the semantic structure), it seems a reasonable assumption that all semantic information required by the speaker is reflected in the syntactic structure. (If not, the speaker would have to extend the syntactic structure somehow to incorporate the remainder of what he or she wants to say, ending with a syntactic structure that *does* include all the semantics.) Therefore, the non-isomorphism between syntactic and semantic structure implies that the syntactic structure contains information that is not present in the semantics. So what then is this information?

A moment's thought provides the answer: the additional non-semantic information must support interpersonal and other communication. The difference between active and passive voice, for example, is thematic, not semantic. The difference between “the door is closed” spoken normally (as a statement) and spoken with a final rising intonational contour (as a question) is not semantic, it is Speech Act-related, hence interpersonal. The difference between completed and continuative aspect (as in “he sang” vs. “he is singing”) reflects the speaker's point of view of the action-event and is therefore also interpersonal.

One has to conclude that syntactic form is a structure that merges information about the sentence from several sources: semantics (for the primary content), discourse (for theme and focus), interpersonal goals and plans (for Speech Acts), and so on. The syntactic structure cannot be isomorphic to any one of these component source structures alone, since it houses more information than any of them do individually. Rather, the syntactic structure merges these disparate types of information into one structure. Necessarily, therefore, the syntactic structure is much closer to the surface form of the sentence than any of the other structures are.

2 A Model of Discourse Structure

We now turn to discourse. By analogy to the syntactic structure, I argue that the content of a discourse structure derives from several sources, and that it is a common, surface-level structure designed to house them all. The major sources for the content of a discourse are: the semantics of the message, the interpersonal Speech Acts, knowledge about stylistic preferability, and the guidance information included by the speaker to assist the hearer's understanding processes (namely information that signals theme, focus, and topic). We consider these four in turn.

1. Semantic information: Semantic information is information about the world and our perceptions of it, our non-emotionalised thoughts and dreams, etc. In logic, semantic information is that which can take a truth value; see for example the discourse representations of [Kamp 81, Asher 93]. In computational approaches, semantic information consists of propositions in a knowledge base, often represented as case frames using terms defined in a taxonomic ontology. Each clause in a discourse may contain semantic information, and if it does, this information is usually related to semantic information in nearby clauses by one or more semantic interpropositional relations (sometimes also called ideational or informational relations), such as CAUSE, TEMPORAL-SEQUENCE, PART-OF. A collection of approx. 40 semantic relations can be found in [Hovy et al. 92, Hovy & Maier 92] (see the Appendix to this paper).

2. Interpersonal information: Discourses are made for communication. From the highest levels and most elaborate intentions (to instruct or amuse someone), through mid-level goals (to explain a solution or justify an action), to the most basic Speech Acts (to inform of a fact, to ask a question), each clause and each block of clauses serves some communicative purpose of the speaker. Typically, discourse goals refer to the speaker, the hearer, and the desired effect on the hearer's state of knowledge, state of happiness, state of belief, etc. Often, these goals include portions of semantic information. If one partitions the discourse blocks and subblocks according to communicative purpose, one arrives at the kinds of discourse segments, each with a purpose, described in [Grosz & Sidner 86]. Adjacent blocks usually act as substeps in some larger more encompassing plan, which itself is a block with a discourse purpose. It is possible also to collect and taxonomise the discourse segment goals (although in previous work they have often been called "interpersonal discourse relations"); see for example [Hovy et al. 92, Hovy & Maier 92, Moore 89, Maybury 90].

Work in Computational Linguistics has paid some attention to these aspects of discourse in the past decade, notably in an area called Text Planning concerned with computer

programs that automatically plan coherent paragraphs of text given communicative goals and a knowledge base of information to convey. Recently, [Moore & Pollack 92] have made a good argument for the need for two distinct discourse structures, one based on semantic interrelatedness and the other on communicative goals. A multifunctional model of discourse similar to the one here is presented in [Redeker 91].

3. Stylistic considerations: We turn to a different type of information source: stylistic control. In general, a set of semantic propositions can be generated as numerous surface forms, even under a single primary interpersonal communicative goal. As argued in [Hovy 88], surface form variation is controlled by stylistic factors in order to achieve additional interpersonal goals. For example, a formal style can help achieve the goal to create some interpersonal distance; an opinionated style can help achieve the goal to be noticed, and so on. Since the links between interpersonal goals and surface form are often extremely complex, it is easier to view stylistic aspects as a separate factor in the forces constraining the form of the discourse. Their effect however is very clear and manifests itself on many levels, especially during the phase of sentence planning performed after text content selection and before surface realisation [Rambow & Korelsky 92, Hovy 92]. These tasks include:

- **Clause aggregation:** the operation of merging very similar representations into conjunctive clauses so as to remove redundancy: “Bush is sure to veto Bill 1711 and Bill 2104” (more fluent style) instead of “Bush is sure to veto Bill 1711. He is also sure to veto Bill 2104.” (more precise, detailed style). Typical cue words that signal aggregation are: *also, respectively, all except*. An example of aggregation is given later in the paper.
- **Certain types of lexical choice:** the determination of verbs can significantly affect the local intersentential structure of the discourse: “He sold her the book for \$15” (fluent, unmarked style) vs. “He gave her the book. Simultaneously, she gave him \$15” (more precise, detailed style).
- **Clause-internal structuring:** whether, for example, an attribute is realised as a relative clause: “the book, which is blue, . . .” (more verbose style) or as an adjective: “the blue book. . .” (less verbose style).
- **Pronominalisation:** a mechanism to remove redundancy and achieve greater internal cohesion, pronominalisation respects discourse boundaries and can thus be used to signal such boundaries.

4. Interpretation guidance: Pity the poor hearer! During communication, especially in a monologue situation, the hearer must build up several extremely complex structures — the meaning, the argument development, etc. — clause by clause, block by block. He or she must not only correctly interpret the meaning of each clause, but must then decide whether it is semantic, interpersonal, or both, attach it to the evolving structures in the right places, decide after attachment what new pieces of structure (if any) to build, keep track of assumptions against later verification, and then, if there’s any time left, think of appropriate responses! A sympathetic speaker will naturally try to help. One way to help is to signal the role of each new piece of information — is it primarily semantic, or interpersonal, or both — and to signal where and how to attach the new information. One can think of the

assistance information provided by the speaker in this regard as *interpretation guidance*. As argued in [Lavid & Hovy 93], the speaker signals theme to indicate to the hearer where in the evolving discourse the new sentence or group of sentences attaches; the speaker signals focus in the sentence to indicate to the hearer where in the clause the hearer should expend most inferential attention, and so on. Mechanisms to perform this signalling include voice (active and passive), clause constituent ordering, pronoun use (pronouns respect discourse boundaries), pitch range and stress (for spoken discourse), etc. Each type of signal involves well-understood cues. More about guidance later in the paper.

3 The Rhetorical Structure

Now, pity the poor speaker! How much information must be packed into each clause! It is easy (especially for a computer scientist) to imagine the potential for disaster in integrating semantic information, interpersonal intentions, stylistic considerations, and guidance information, all at the same time in a dynamic process. Small wonder that a distinct structure, one much closer to the surface form of the discourse, seems useful. Just as (in the case of single sentences) syntactic structures integrate information from various sources using a set of syntactic labels (noun, verb, preposition, subject, object, etc.), just so the discourse-level rhetorical structures require their own, multipurpose, type of interclause relation — rhetorical relations. Rhetorical relations are the presentational analogue of both semantic relations and interpersonal goals. They have been studied for a long time; see for example [Hobbs 79, Mann & Thompson 88, Sanders et al. 92], and see [Hovy & Maier 92] for a general survey and collection. In this paper I will use a fairly generic set of relations; I am not concerned here with any particular collection.

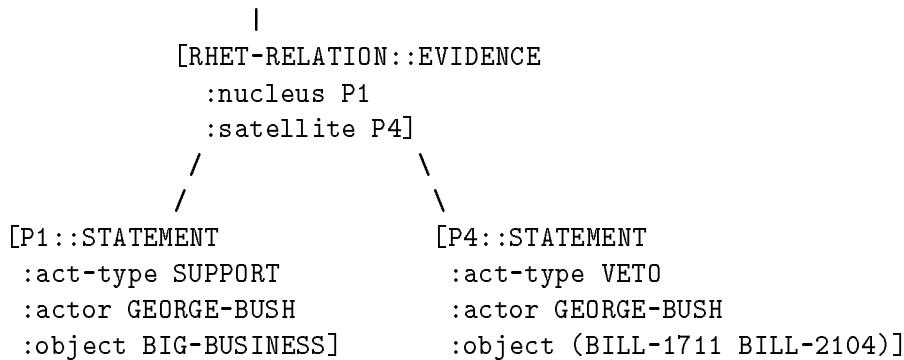
What is of more interest is the function and the use of relations in discourse. While I do not think it is useful to identify a unique rhetorical partner for each semantic relation and each interpersonal goal, it doesn't seem surprising that certain strong correlates exist. Just as semantic *agent* and *patient* pattern closely in English sentences with syntactic *subject* and *direct object* respectively, just so semantic TEMPORAL-SEQUENCE patterns closely with rhetorical PRESENTATIONAL-SEQUENCE. In fact, it also patterns closely with SPATIAL-SEQUENCE; such simplifications of semantic diversity are found in several areas, as where the semantic relations PART-WHOLE, PLAN-STEP, ABSTRACT-INSTANCE can be grouped to pattern with rhetorical ELABORATION. This type of “overloading” of information under a single relation appears also at the sentence level: consider how the preposition *for* signals both BENEFACTIVE and CAUSE semantic relations as in “Oliver brought the spoon for his sister” and “Oliver sang for his supper” respectively. Exactly which rhetorical relations are most useful to define as separate entities, and how they co-pattern with semantic relations, interpersonal goals, and control information, remains a matter of investigation.

The rhetorical discourse structure differs from the semantic and the interpersonal structure. Incorporating as it does the effects of both, as well as of other constraints on the discourse, it is much closer to the surface form of the text¹.

¹Rhetorical Structure Theory (RST) [Mann & Thompson 88], a fairly popular approach in this area, attempts to describe rhetorical structure using about 25 rhetorical relations. What has often been called a liability of RST, namely that its analyses mirror the text too closely, is in fact a virtue — it represents

domain-oriented knowledge base and the representations needed for stylistically adequate generation. The problem has been studied in the text generation community (see for example [Mann & Moore 80, Kempen 91, Dalianis & Hovy 93]) but is still a long way from being solved, involving as it does questions of conversational implicature [Horacek 92] and of style [Hovy 88]. As described in [Hovy 90], the presence of a discourse structure greatly reduces the problem of finding candidates for aggregation (from polynomial to sub-linear in the total number of clause-sized representation clusters).

The operation of aggregation (as well as many other sentence planning steps) can only be performed on a structure that is fairly close to the surface level, since semantic and interpersonal structures do not contain the appropriate surface-level information. In particular, interpersonal structures do not contain case frame details and semantic structures do not express the groupings of case frames in discourse blocks. Sentence planning is thus one way of motivating the independent existence of a separate rhetorical structure. Just to complete the example, the rhetorical structure for sentences (1) and (2) would be:



Notice the similarities and differences between this rhetorical structure and the semantic structures and interpersonal structure from which it derives.

4 The Role of Discourse Markers

During the process of interpreting a discourse, the hearer must perform several distinct operations. In an extremely simplified model, and without implying process orderings, these operations include the following. For each clause, the hearer has to decode²:

- the semantic information (possibly inverting some sentence planning operations),
- the interpersonal purpose (at the lowest, Speech Act, level),

and then has to follow the guidance information contained in the clause to determine:

- where in the rhetorical structure to attach the current clause, using the theme guidance information in the clause.

Once the new clause has been attached to the rhetorical structure, the hearer can begin to “unpack” the new attachment:

²For ease of visualisation, one may think of the semantic structure as a network of nodes, the interpersonal structure as a tree, and the rhetorical structure as a lattice (a tree in which some nodes may have multiple parents).

- attaching the new semantic information to the semantic structure, checking for knowledge-level inconsistencies, and possibly inferring new semantic information, concentrating inference especially on the portion(s) of the clause signalled to be in focus.
- growing the interpersonal argument structure with the new Speech Act and possibly inferring additional, higher-level, intentions.

The speaker has only three cue mechanisms with which to signal information about network, tree, or lattice attachment and focus location: stress (intonational or orthographic), word order, and cue words. Typically, the presence of a relation is signalled by a cue word; the presence of a sentence planning operation either by a cue word or a word order inversion, and any guidance information is signalled either by stress or by word order patterning.

Rhetorical structure is signalled by cue words indicating rhetorical relations. Examples of these relations and their cue words appear in [Martin 92].

Interpersonal/intentional structure is signalled by cue words indicating interpersonal communicative goals such as EXPLAIN. Examples can be found in [Moore 89, Maybury 90].

Semantic structure is signalled by cue words indicating semantic interproposition relations such as CAUSE. Examples can be found in [Hobbs 79, Dahlgren 88].

Stylistic variants are signalled by cue words indicating sentence planning operations such as *also* and *respectively* for aggregation, pronouns for pronominalisation, etc.

Guidance information is signalled by syntactic inversion of word placement, such as sentence-initial position for theme, post-verb position for focus, etc.; see [Lavid & Hovy 93, Hajičová 87],

4.1 Discourse Markers with Cue Phrases

In the model outlined here, discourse structure is signalled by several mechanisms, the most common being cue words or phrases for discourse relations and intentions. In English, for example, “in order to” and “because” signal CAUSE and “then” signals SEQUENCE. An effort to collect and taxonomize discourse structure relations over a period of several years is reported in [Hovy & Maier 92] and extended in [Hovy et al. 92]. The resulting three taxonomies—one for semantic relations, one for interpersonal intentions, and one for rhetorical relations—from the latter reference is reproduced in the Appendix.

As [Knott & Dale 94] have shown, one can infer a taxonomic structure over these relations by testing for substitutability of their cue phrases. For example, “then” can substitute for either “after” or “beside” in appropriate circumstances, showing the specialisation of SEQUENCE into TEMPORAL-SEQUENCE and SPATIAL-SEQUENCE respectively.

Unfortunately, it is a strange fact of English (and many other languages) that the number of discourse relation words/phrases seems to be at least one (and probably two) orders of magnitude smaller than the number of verbs or nouns. One might expect that, given 50.000 verbs (i.e., distinguishably different processes or actions), the number of semantically plausible interconnections among them is on the order of at least half of the total number of possible interconnections, thus half of 25×10^8 , a large number. But the number of semantic cue words appears to be fewer than 1.000. This implies that the semantic cue

words/phrases are highly vague and possibly even ambiguous (where I take vague to mean not distinguishing among similar possible candidates and ambiguous to mean indicating several different candidates simultaneously). The same sort of argument may hold for the interpersonal intentions, if one could quantify communicative intentions somehow.

With regard to the rhetorical relations, they have a multipurpose function essentially by design. Primarily, of course, they signal the articulation of the rhetorical structure. But since this structure mirrors (in many places) the underlying semantic and interpersonal/intentional structures so closely, it is small wonder that they also signal semantic interpropositional relations and interpersonal intentions. Not surprisingly, one can find several regions of very close correspondence among the three taxonomies of the Appendix. As a result, cue phrases such as “then” can sometimes signal pure rhetorical/presentational structure, as in the itemization of a list of unordered facts, and sometimes semantic structure, as in a sequence of events ordered by time. With such an impoverished lexicon of cue phrases in English, it is the hearer’s task to decide when an overloaded word such as “then” is rhetorical and when it is both rhetorical and semantic.

5 Conclusion

Given the differences between semantic and interpersonal information, the stylistic necessity of performing such sentence planning tasks as aggregation, pronominalisation, etc., not to mention the reorganisations of material caused by lexical choice, and given the complexity of managing the disparate effects of these operations, a fairly surface-level structure that records the realisation of the text becomes a necessity to ensure effective communication. The interclausal relations employed in such a rhetorical structure have to be fairly neutral in character, carrying as they do semantic, interpersonal, and guidance information simultaneously. The precise format of the rhetorical structure and its rhetorical relations remains a topic of ongoing study.

References

- [Asher 93] Asher, N. 1993. *Abstract Objects, Semantics and Anaphora*.
- [Bresnan 78] Bresnan, J. 1978. A Realistic Transformational Grammar. In *Linguistic Theory and Psychological Reality*, M. Halle, J. Bresnan, and G. Miller (eds), Cambridge: MIT Press (39–49).
- [Chomsky 65] Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
- [Dahlgren 88] Dahlgren, K. 1988. *Naive Semantics for Natural Language Understanding*. Boston: Kluwer Academic Press.
- [Dalianis & Hovy 93] Dalianis, H. and Hovy, E.H. 1993. Aggregation in Natural Language Generation. In *Proceedings of the 5th European Workshop on Natural Language Generation*, Pisa, Italy, 1993.
- [Grosz & Sidner 86] Grosz, B.J. and Sidner, C.L. 1986. Attention, Intentions, and the Structure of Discourse. *Journal of Computational Linguistics* 12(3) (175–204).
- [Hajičová 87] Hajičová, E. 1987. Focussing — A Meeting Point of Linguistics and Artificial Intelligence. *Artificial Intelligence II: Methodology, Systems, Applications*, Ph. Jorrand and V. Sgurev (eds), Amsterdam: Elsevier North-Holland Publishers (311–325).

- [Halliday 85] Halliday, M.A.K. 1985. *An Introduction to Functional Grammar*. Baltimore: Edward Arnold Press.
- [Hobbs 79] Hobbs, J.R. 1979. Coherence and Coreference. *Cognitive Science* 3(1) (67–90).
- [Horacek 92] Horacek, H. 1992. An Integrated View of Text Planning. In *Aspects of Automated Natural Language Generation*, R. Dale, E. Hovy, D. Rösner, O. Stock (eds). Heidelberg: Springer Verlag Lecture Notes in AI number 587 (57–72).
- [Hovy 88] Hovy, E.H. 1988. *Generating Natural Language Under Pragmatic Constraints*. Hillsdale: Lawrence Erlbaum Publishers.
- [Hovy 90] Hovy, E.H. 1990. Unresolved Issues in Paragraph Planning. In *Current Research in Natural Language Generation*, R. Dale, C. Mellish, M. Zock (eds), Academic Press.
- [Hovy 92] Hovy, E.H. 1992. Sentence Planning Requirements for Automated Explanation Generation. DIAMOD-Bericht no. 23, GMD, St. Augustin, Germany.
- [Hovy et al. 92] Hovy, E.H., Lavid, J., Maier, E., Mittal, V., and Paris, C.L. 1992. Employing Knowledge Resources in a New Text Planner Architecture. In *Aspects of Automated Natural Language Generation*, R. Dale, E. Hovy, D. Rösner, O. Stock (eds). Heidelberg: Springer Verlag Lecture Notes in AI number 587 (57–72).
- [Hovy & Maier 92] Hovy, E.H. and Maier, E. 1993. Parsimonious or Profligate: How Many and which Discourse Structure Relations? *Discourse Processes* (to appear).
- [Kamp 81] Kamp, H. 1981. A Theory of Truth and Semantic Representation. In *Formal Methods in the Study of Language*, Groenendijk, J.A.G., Janssen, T.M.V. and Stokhof, M.B.J. (eds), Mathematical Centre Tracts (vol 136), Amsterdam, (277–322).
- [Kempen 91] Kempen, G. 1991. Conjunction Reduction and Gapping in Clause-Level Coordination: An Inheritance-Based Approach. *Computational Intelligence* 7(4).
- [Knott & Dale 94] Knott, A. and Dale, R. 1994. Defining a Taxonomy of Coherence Relations Using a Set of Cue Phrases.
- [Lavid & Hovy 93] Lavid, J.M. and Hovy, E.H. 1993. Focus, Theme, Given, and Other Dangerous Things. Working paper.
- [Mann & Moore 80] Mann, W.C. and Moore, J.A. 1980. Computer as Author: Results and Prospects. USC/Information Sciences Institute Research Report ISI/RR-79-82.
- [Mann & Thompson 88] Mann, W.C. and Thompson, S.A. 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text* 8(3) (243–281). Also available as USC/Information Sciences Institute Research Report RR-87-190.
- [Martin 92] Martin, J.R. 1992. *English Text: System and Structure*. Amsterdam: Benjamins Press.
- [Maybury 90] Maybury, M.T. 1990. *Planning Multisentential English Text Using Communicative Acts*. Ph.D. dissertation, Cambridge University. Also available as RADC Technical Report 90-411.
- [Moore 89] Moore, J.D. 1989. *A Reactive Approach to Explanation in Expert and Advice-Giving Systems*. Ph.D. dissertation, University of California in Los Angeles.
- [Moore & Pollack 92] Moore, J.D. and Pollack, M.E. 1992. A Problem for RST: The Need for Multi-Level Discourse Analysis. Squib in *Computational Linguistics* 18(4).
- [Rambow & Korelsky 92] Rambow, O. and Korelsky, T. 1992. Applied Text Generation. *Proceedings of the Applied Natural Language Processing Conference*, Trento, Italy.

- [Redeker 91] Redeker, G. 1991. Lexical Markers of Discourse Coherence. *Linguistics* 29(6) (1139–1171).
- [Sanders et al. 92] Sanders, T.J.M., Spooren, W.P.M.S., and Noordman, L.G.M. 1992. Towards a Taxonomy of Coherence Relations. *Discourse Processes* 15 (1–35).

Appendix

Semantic and rhetorical relations and interpersonal intentions, adapted from [Hovy et al. 92].