

In Defense of Syntax: Informational, Intentional, and Rhetorical Structures in Discourse

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: 310-822-1511
fax: 310-823-6714
email: hovy@isi.edu

Introduction: The Point of this Paper

Much has been written on the nature and use of so-called rhetorical relations to govern the structure and coherence of discourse, and much has been written on the need for communicative intentions to govern discourse construction. In early research on automated text creation, the operationalized structural relations from Rhetorical Structure Theory (RST) [Mann & Thompson 88] were used both to carry the speaker's communicative intention (which enabled the use of a NOAH-like top-down expansion planner) and also to ensure coherence (by utilizing the constraints on Nucleus and Satellite from RST). This dual functionality is characteristic of the operators used in the various text structure planners that have been built to date, whether they were oriented more toward surface text structure (such as the relation/plans of the RST Structurer [Hovy 88, Hovy 90]) or more toward the communicative intentions underlying the text (as the text plans of the EES and PEA text planner in the later [Moore & Paris 89, Moore 89] work)¹.

However, the resulting perspective shift from surface (RST) oriented toward intentional did engender considerable discussion about the types of discourse structure relations and the form of discourse structure itself. In work collecting and classifying relations from many sources, [Maier & Hovy 91, Hovy & Maier 93] eventually created a taxonomy of three types of relations: ideational (semantic), interpersonal (intentional), and textual (rhetorical); for a more expanded taxonomy see [Hovy et al. 92]. Though the details were not explicitly spelled out, the idea was that a discourse can (and should) be described by at least three structures in parallel: the semantic, the interpersonal, and the rhetorical.

Recently, Moore and Pollack published a paper [Moore & Pollack 92] in which they show the need for two accounts of discourse structure, the semantic (what they call informational) and the interpersonal (what they call intentional). They also offer a convincing example to show that these two structures are not in general isomorphic. With these claims I have absolutely no problems — see the discussion of multiple parallel relations in [Hovy et al. 92] — except the fact that it doesn't go far enough. It doesn't recognize the need for the rhetorical relations. The current paper argues for the need for an additional, rhetorical, structure to describe discourse.

¹It is important to understand that merely using plan operators with more "intentional" names does not guarantee that either the text planner or the resulting text structure encapsulate "intentionality" in a real way (whatever "intentionality" may mean in this context). When a text plan library includes both "intentional plans" and RST-like operators side by side, and the planner uses them interchangeably, and they fulfill a similar function in the resulting discourse structure, then the difference between the two types of operator as implemented is one of nomenclature rather than of true functionality, and is thus open to McDermott's "Wishful Mnemonics" trap [McDermott 81]. The shift of nomenclature does however reflect a shift of perspective, namely the recognition that text planning should develop non-linguistic, intentionality-oriented terminology. The true import of this shift is only gradually becoming apparent.

Why is there Syntax?

I would like to describe a more complete model of discourse structure by analogy to single sentence structure. The typical model most of us have 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.

Now ask yourself: *why is syntactic structure necessary as a separate, autonomous construct?* If you have a well-formed semantic structure, possibly something like a case frame or a set of 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 (with pairs of preposition and filler) somewhat so; Japanese (with pairs of case marker and filler) more so. Under this view there's 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. That is to say, either the semantics is not derivable from the syntax, or the syntax contains more information than the semantic structure. Regardless of what theories of syntax and semantics one follows, the statement of non-isomorphism would be supported, I believe, by all experts.

Assuming that all required semantic structure is reflected in the syntactic structure, this non-isomorphism implies that the syntactic structure contains other information as well. Short of random inclusion of noise, there's no other explanation.

A moment's thought provides the answer: the syntactic structure of a sentence contains additional non-semantic information. 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 with a final rising intonational contour (as a question) is not semantic, it is Speech Act-related, hence interpersonal.

Briefly, then, 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. Necessarily, also, the syntactic structure is much closer to the surface form of the sentence than any of the other structures are.

The Model of Discourse Structure

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

1. Semantic information: In our normal computational approaches, semantic information consists of propositions in a knowledge base, possibly represented as case frames using terms defined in a taxonomic ontology. The interpropositional linkages are relations that have been called *semantic*, *ideational*, or *informational*. These include relations such as CAUSE, PART-WHOLE, IS-A, TEMPORAL-SEQUENCE, SPATIAL-SEQUENCE.

2. Interpersonal information: For computational discourse, interpersonal information takes the form of communicative goals. Properly, the goals should refer to the speaker, the hearer, and the desired

effect on the hearer' state of knowledge, state of happiness, state of belief, etc. (which includes portions of semantic information). The interpersonal goals include MOTIVATE (someone to do something), JUSTIFY (one's own actions), EXPLAIN (the operation or development of something or some events), CONCEDE (a point of contention), and so on. Since we are still sorting out the various terms, the precise nature of these goals in a discourse planning system is not yet clear; I believe for example that we are still confusing things a little when we talk about "interpersonal relations", since what's interpersonal are not relations but goals.

3. Control information: Despite the Attentional state in [Grosz & Sidner 86] and [Moore 89], the inclusion of control information in discourse planning systems has not received the attention it deserves, primarily I think because we do not understand clearly enough how discourse analysis works. 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 spend 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.

4. Style: As anyone knows after writing a text generator for a domain created by another person for another task, the mismatch between representations suited to the task and representations suited to language can be daunting. Usually, text generated directly from domain-oriented representations is stylistically awful, to say the least. As argued in [Rambow & Korelsky 92, Hovy 92], several so-called *sentence planning* tasks must be performed after text content selection and structuring and before actual surface realization. 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" instead of "Bush is sure to veto Bill 1711. He is also sure to veto Bill 2104.")
- pronominalization determination: this operation depends on discourse structure and on stylistic factors
- clause-internal structure: whether, for example, an attribute is realized as a relative clause ("the book, which is blue...") or as an adjective ("the blue book...")
- certain types of lexical choice: the determination of verbs can significantly affect the local structure of the discourse

An example of aggregation is given at the end of the paper.

The Rhetorical Structure

Pity the poor speaker. All this information, and more, must be packed into each clause! Small wonder that a distinct structure, one much closer to the surface form of the discourse, is useful. Just as in the case of single-sentence syntactic structures, 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. 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*, 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 all pattern with rhetorical ELABORATION. 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 intentional 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 text. RST provides one attempt at providing rhetorical structure. What has often been called a liability of RST, namely that its analyses mirror the text so closely, is in fact a virtue — it represents a minimal step of abstraction away from the surface form, and does not discard information that is not directly semantic or intentional. This is not to say that RST has no flaws; one of its principal problems is the conflation of intentional, semantic, and rhetorical relations. Not all the relations of RST are rhetorical — for example JUSTIFY and MOTIVATE are clearly intentional, and CAUSE and PART-WHOLE are clearly semantic.

What then does a rhetorical structure look like? It is in fact a structure very familiar to us, the one that first appears from considering text itself instead of the meaning or communicative intent thereof. To illustrate, I conclude with the Bush example and the problem of aggregation.

One of the most common mismatches between the representations constructed for a domain-oriented knowledge base and the representations needed for stylistically adequate generation occurs with multiple copies of very similar information, which the text planner has to “aggregate” in order to reduce redundancy. In most knowledge representations, it is overwhelmingly likely that the propositions underlying sentence (1):

- (1). George Bush supports big business.
- (2). He is sure to veto House bills 1711 and 2104.

will be stored separately, which under straightforward text production would give rise to

- (1). George Bush supports big business.
- (2). He is sure to veto House bill 1711.
- (3). (And) He is sure (also) to veto House bill 2104.

The problem of aggregating partially similar representations has been studied by several people 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 (see [Horacek 92]) and of style (see [Hovy 87]). 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). This kind of operation can only be performed on a fairly surface-level discourse structure; a semantic or interpersonal does not contain the appropriate information.

I think it is too early to try to define exactly what is and isn't part of the rhetorical structure; the most useful answers seem to crystallize from practical experience. I believe that as examples of rhetorical structure one should consider RST trees as given in [Mann & Thompson 88], though few of the specific RST relations used there are in my opinion turn out to be most productive. Instead, I believe that the relations identified in [Martin 92] and classified as Textual in [Hovy et al. 92] will be more useful. I also believe that the rhetorical structure will not necessarily contain clauses at its leaves, but may in fact contain information reaching “into” the clause itself; most likely as a set of attributes which the clause should exhibit, rather along the lines outlined in [Meteer 90]. The notions of Nucleus and Satellite seem to me very useful however.

Conclusion

Given, in any text generation system, the stylistic necessity of performing such sentence planning tasks as aggregation, pronominalization, etc., not to mention the reorganizations of material caused by lexical choice, and given the complexity of managing the disparate effects of these operations, a fairly surface-level structure that governs the realization of the text becomes a practical necessity. The interclausal relations employed in such a structure have to be fairly neutral in character, carrying as they do semantic, interpersonal, and “control” information simultaneously. Though the precise format of the rhetorical discourse structure and its rhetorical relations remains a topic of ongoing study, the need for and utility of such constructs cannot be denied by anyone who has actually tried to build a real system.

References

- [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.
- [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).
- [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 87] Hovy, E.H. 1987. Interpretation in Generation. *Proceedings of the 6th AAAI Conference*, Seattle (545-549). Also available as USC/Information Sciences Institute Research Report ISI/RS-88-186.
- [Hovy 88] Hovy, E.H. 1988. Planning Coherent Multisentential Text. *Proceedings of the 26th ACL Conference*, Buffalo (163-169).
- [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 93] Hovy, E.H. and Maier, E. 1993. Parsimonious or Profligate: How Many and which Discourse Structure Relations? *Discourse Processes* (to appear).
- [Kempen 91] Kempen, G. 1991. Conjunction Reduction and Gapping in Clause-Level Coordination: An Inheritance-Based Approach. *Computational Intelligence* 7(4).
- [Lavid & Hovy 93] Lavid, J.M. and Hovy, E.H. 1993. Focus, Theme, Given, and Other Dangerous Things. Working paper.
- [Maier & Hovy 91] Maier, E. and Hovy, E.H. 1991. Organizing Discourse Structure Relations using Metafunctions. In *New Concepts in Natural Language Generation: Planning, Realization, and Systems*, H. Horacek (ed), London: Pinter (to appear).
- [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.
- [McDermott 81] McDermott, D.V. 1981. Artificial Intelligence Meets Natural Stupidity. In *Mind Design*, J. Hauge-land (ed), Cambridge: MIT Press, (143-160).
- [Meteer 90] Meteer, M.W. 1990. *The "Generation Gap": The Problem of Expressibility in Text Planning*. Ph.D. dissertation, University of Massachusetts at Amherst. Available as BBN Report No. 7347, February 1990.
- [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 & Paris 89] Moore, J.D. and Paris, C.L. 1989. Planning Text for Advisory Dialogues. *Proceedings of the 27th ACL Conference*, Vancouver (67-75).
- [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.