

ŠÁRKA ZIKÁNOVÁ ET AL.,
DISCOURSE AND COHERENCE.
FROM SENTENCE STRUCTURE TO RELATIONS IN TEXT

Ústav formální a aplikované lingvistiky, Praha 2015.

In their monograph the group of authors affiliated with the Institute of Formal and Applied Linguistics (Charles University in Prague, Faculty of Mathematics and Physics) present the results of their research on the interplay of intrasentential and intersentential relations which contribute to textual coherence. Their primary focus is on discourse relations, coreference, bridging relations and topic-focus articulation.

The organization of the monograph resembles that of a research paper. The initial chapters (2–5) establish the theoretical background of the research; chapters 6–8 introduce the data, and the subsequent chapters (9–14) present case studies which bring new insights into the relations between discourse and syntactic analysis, Czech connectives, coherence and coreference, and the activation of elements of the stock of shared knowledge in a text.

Drawing on the Functional Generative Description, the authors have developed an original theoretical approach which allows them to give a detailed complex description of discourse relations at several mutually interwoven layers of text. The network of discourse relations is captured by a unique system of annotation in the Prague Dependency Treebank. The theoretical part of the monograph addresses such fundamental issues as the nature, delimitation and size of the basic units of description (viz. discourse arguments), or the manifestation of coherence at various layers of language (coreference and bridging relations, or information structure).

In Chapter 2, discourse relations are defined as “semantic relations that connect two discourse units” (p. 13). The smallest discourse unit is represented by a simple clause with one predication; “the two discourse units building a discourse relation are referred to as discourse arguments” (p. 17). The most apparent indicators of discourse relations are discourse connectives, which not only make it possible to detect these relations (intra- and inter-sentential) but also facilitate the identification of a particular semantic type of relation that obtains between the discourse units.

While discourse relations obtain between predicative units, the relations between non-predicative units (especially between noun phrases) can be described as coreference phenomena (Chapter 3). The approach adopted in the monograph distinguishes between grammatical and textual coreference. The scope of grammatical coreference does not normally exceed the sentence, “it is grammar-driven by the use of pronouns and, in most cases, it is possible to identify the antecedent on the basis of grammatical rules” (p. 33). Textual coherence, on the other hand, does not rely solely on grammatical means. It often connects entities across sentence boundaries. Coreference relations are distinguished from bridging relations (including, e.g. meronymical relations, the relations between a set and its subset(s) or between an entity and its singular function, see Chapter 4), which contribute to text cohesion by linking non-coreferential noun phrases.

Chapter 5 presents another aspect of coherence, viz. topic-focus articulation. The description and annotation of the information structure of the sentence rests on

“three main features: (i) contextual boundness; (ii) communicative dynamism and (iii) sentence division into topic and focus” (p. 67). Topic and focus are detected on the basis of contextual boundness and/or by using the criteria of the so-called question test and test by negation (p. 73 ff). In the Prague Dependency Treebank, topic-focus articulation is annotated at the deep syntactic layer from the perspectives of contextual boundness and communicative dynamism.

The theoretical approach described in the initial chapters of the monograph has served as a basis for the annotation of various types of discourse relations in a corpus of modern Czech journalistic texts, the Prague Dependency Treebank (PDT). What makes the PDT really outstanding is the three-layer annotation of its almost 50 thousand sentences. The corpus has been annotated (mostly manually) at the morphological, analytical (i.e. surface syntactic structure) and tectogrammatical (deep syntactic structure including topic-focus articulation) levels. “All the discourse coherence phenomena have been annotated on the same data, i.e. on top of the whole tectogrammatical layer.” (p. 88) The annotation encompasses pronominal and grammatical coreference, textual coreference, relations of association and discourse relations. The annotation of discourse phenomena on syntactic trees makes it possible “to preserve the connection with and information from the analyses of previous levels” (p. 23). The benefits of this unique approach are numerous, including the retrievability of intra-sentential discourse relations and their connectives. The annotation comprises several other discourse-related phenomena: list structures, headings, captions, meta-texts, and genre classification.

The case studies presented in Chapters 9–14 were carried out on the data of the PDT. All the studies link up the issues of annotation and analysis of discourse phenomena. Chapter 9 is typical in this respect: on the one hand, it shows “which syntactic characteristics were exploited during the discourse analysis of the text” (i.e. annotation of discourse relations), and on the other hand it proceeds from the discourse relations, overtly signalled by connectives, to the description of “their characteristics from the syntactic perspective” (p. 146). The authors demonstrate a high level of correspondence between intra-sentential discourse relations and syntactic and semantic annotation, which allows the researcher to take advantage of the syntactic annotation when extracting discourse structure information automatically. The syntactic and discourse analyses, however, were shown to “differ especially with regard to semantics — while syntactic analysis considers form and meaning as equally important features of structures, discourse analysis tracks meaning systematically disregarding the form” (p. 235). Besides, syntactic analysis cannot provide information on inter-sentential relations.

Chapters 10 and 11 deal with primary and secondary (multi-word) discourse connectives respectively. Again, the authors present the morpho-syntactic properties of connectives “both from a theoretical perspective and from the perspective of manually annotated data” (p. 163), contributing substantially to the description of the under-researched area of these discourse-structuring devices. Both types of connectives have been annotated in the corpus, which makes it possible to explore their syntactic and semantic characteristics. The scalar approach to connectives also allows the authors to describe in detail “the possible boundaries between explicitness and





implicitness of discourse relations, as well as ... the diversity of discourse connectives” (p. 177).

Starting from places in the PDT texts unmarked for coherence relations, Chapter 12 focusses on weak coherence and coherence disruptions. The annotation of no relation between discourse arguments has led to the discovery of new types of coherence relations (text-organizing devices, attribution, and the relation of question and answer) as well as to the formulation of a unifying principle which accounts for the integration of the unattached segments into the text, namely “the readers’ expectation of a coherent text and on their willingness to search for an appropriate coherent interpretation of a seemingly incoherent sequence of sentences” (p. 195).

While Chapter 12 looks at cases of weak coherence, the following one explores interdependencies of topic-focus articulation, coreference and bridging relations to shed light on contextually bound elements of the sentence with no coreference links. Surprisingly, such nodes constitute 30 per cent of the data. The authors show that the lack of anaphoric links can be due to three groups of reasons: “(i) contextually bound nominal groups semantically or pragmatically related to previous textual or extralinguistic context but not specified as bridging relations within the PDT; (ii) secondary circumstances (temporal, local, etc.) and (iii) nominal groups with low referential potential” (p. 213).

Chapter 13 concludes the discussion of various discourse-related phenomena presented in the previous chapters with a study of “activation of objects in readers’ minds while reading the text” (p. 215). The authors describe an algorithm which makes it possible to trace salience of elements of the stock of shared knowledge in a text. Drawing on the PDT, machine learning methods were applied to model the rule-based salience algorithm and “explore the possibility to predict the salience degree automatically” (p. 237).

When looking at the monograph for the first time, the reader’s attention is attracted immediately by two bookmarks, orange and black. As we are reading the book, however, two bookmarks prove quite insufficient since every chapter brings new insights into the topic. We are guided through the complex system of intertwined phenomena which contribute to coherence at various levels: discourse relations, connectors, coreference, bridging relations or information structure. At the same time, we are invited to observe the connections between the theoretical framework and annotation in the PDT, forming the basis of — and at the same time being enhanced by — the original findings of the individual case studies. The monograph is a truly remarkable contribution to our understanding of the interplay of various aspects of discourse structure which results in the constitution of a coherent text.

Markéta Malá (Prague)

Department of English Language and ELT Methodology,
Faculty of Arts, Charles University in Prague
nám. J. Palacha 2, 116 38 Praha 1, Czech Republic
marketa.mala@ff.cuni.cz