

Segmentation of specialized texts: Frames, topics and moves

Proposed workshop at SLE 2025 (Bordeaux)

Keywords: segmentation; discourse annotation; discourse structure; specialized texts; corpora

This workshop aims to bring together linguists from various fields, including computational linguistics, to address the complexities of discourse segmentation in specialized texts. Through a combination of theoretical discussion and methodological innovation, we aim to synthesize differing views about the nature of segments and analyze specialized discourse more effectively. By examining both well-established and emerging approaches, this workshop seeks to foster new insights and collaborative research opportunities in discourse analysis.

Abstracts: Interested colleagues should submit a 300-word abstract (excl. references) for 20-minute talks by 15 November 2024 to marycatherine.lavissiere@univ-nantes.fr, johannes.dahm@univ-nantes.fr and warren.bonnard@univ-lorraine.fr. We will review the submissions and notify all presenters about the acceptance of their abstracts before submitting the workshop proposal to the SLE.

Workshop description:

Various fields in linguistics have recently called for the reintegration of the larger units of discourse into linguistic analysis. These units have long interested researchers in specialized discourse because this discourse is often studied in specific situations with precise functional goals. As such, researchers in specialized discourse have developed models that segment discourse into larger units (cf. Swales, 1990; 2004). This workshop will compare different linguistic models of segmentation of written or spoken specialized language, with the goal of identifying points of convergence and divergence among these models. Prospective contributions may be submitted in the areas of text linguistics and discourse analysis, and may address one of the research themes in the non-exhaustive list *infra*.

1. Issues in the theoretical status of discourse segments in specialized texts

The theoretical status of discourse segments in specialized texts presents a crucial question. Are segments in different genres or formats inherently different, or can they be compared across various forms of specialized discourse, such as argumentative prose and legal texts? For instance, do argumentative discourse segments in political speech operate as those in scientific articles, or is each type fundamentally shaped by the communicative goals and structures of their respective domains?

The workshop also focuses on how segments interact across multiple levels of discourse. In legal texts, recitals often serve as strategies for developing the conceptual framework that shape the whole document's structure. It is important to investigate whether this segmentation follows a truly topic-based inner structure or if it adheres to a more rigid, formalized framework specific to the legal domain. Contributions may thus assess whether more thematic-dependent or more domain-specific patterns can be used to describe the discourse structure of defined text types.

Contributions may also discuss how different linguistic theories conceptualize segments and segmentation within their frameworks. For example, how can theories developed for non-specialized texts inform the segmentation of specialized texts, and vice versa? Theories focusing on topics and information structure (Halliday, 1970; Lambrecht, 1994; Hajicová et al., 2013) may provide valuable insights when applied to specialized texts. The goal is to investigate whether discourse units and their boundaries can be more effectively identified by leveraging these theories, particularly when shifts in topics serve as segmentation markers.

Additionally, another objective of this analysis is to identify text structuring patterns at both a macro level (i.e., the prototypical sequences of large discourse units) and a micro level (the formulation within these units). It would be valuable to explore the extent to which specialized texts allow for variability in these patterns, with some genres displaying flexible move sequences, like judicial opinions (Lavissière and Bonnard, 2024) and others, like scientific texts, adhering to a more fixed structure.

2. Ontologies and segmentation in specialized texts

Ontology plays a central role in understanding the structure and organization of specialized texts. In this call, ontology means the conceptual framework that categorizes and defines the key elements, relationships, and hierarchies within a particular domain. In discourse analysis, ontologies serve as tools to build a structured representation of knowledge, guiding how information is segmented and interpreted. In specialized texts, should specific types of ontological relationships among segments be expected? How much do the ontologies used to organize these texts rely on specific extralinguistic knowledge, such as “knowledge frames” (Van Dijk, 2013), and does this have specific implications for text progression or information structure in specialized texts?

An additional focus may be on the continuum between the ontological structure of a specialized field and its textual surface patterns. Exploring how semantic frames (Fillmore, 2006) can serve as models to understand the mechanisms operating between these two poles offers a valuable perspective. This analysis may consider both the interaction between ontologies and semantic frames, and the relationship between framing elements and textual surface structures, such as phrasemes, grammatical constructions, collocations, and phraseological slot fillers with predefined lexical elements (see Dahm, forthcoming).

Contrastive approaches between languages are also welcome.

3. Issues on the selection of a unit of analysis and its influence on interpretation and annotation

Determining the appropriate unit of analysis for discourse segmentation is still under debate among linguists. Approaches have varied widely, considering units ranging from clauses (Longacre, 1983) and sentences (Polanyi, 1988), to turns of talk (Sacks, 1974) or intentionally defined discourse segments (Grosz and Sidner, 1986). Each approach brings its own set of assumptions about how discourse should be interpreted and annotated. More recently, Egbert et al. (2021) developed a new framework to segment discourse units in conversational speech,

moving from the Hymesian notion of a speech event to the term Discourse Unit, referring only to “functional segments of conversation” (p. 725).

How do different choices of analytical units influence the interpretation and segmentation of specialized texts? The papers may focus on assessing the impact of these units on the reader's perception of coherence and structure within specialized genres and across registers. The influence of each theory's focus—whether on clauses, sentences, or discourse moves—on segmentation and the subsequent understanding of information may also be examined.

4. Exploring linguistic phenomena that pose problems in terms of discourse segmentation

In corpus linguistics and computational linguistics, discourse segmentation is a key stage to corpus annotation. However, specialized texts, including legal or technical documents, frequently employ dense structures and hierarchical layers of information and thus often present unique challenges to conventional discourse segmentation techniques. In the workshop, linguistic phenomena that disrupt typical segmentation patterns, such as syntactic focusing devices, temporal expressions, relative clauses, and reported speech (Carlson & Marcu, 2001: 3), may be examined.

Additionally, the process of “incorporating interpretation into segmentation” (Hoek et al., 2018) would allow for the interpretive value of segments to be taken into account. Conversely, other recent approaches to specialized texts apply a segmentation based on formal units like the paragraph (see Rau, 2021) before interpreting their discursive function. Which segmentation approach maximizes the interpretative value for an annotator working with specialized texts and allows for analysis of segments that are either excessively small or large for meaningful interpretation? Should inferred coherence relations embedded within syntactic constructions be treated as independent segments or as integral parts of larger units?

5. Methodological approaches to discourse segmentation

The workshop will also include a focus on methodological advances in discourse segmentation and annotation. Emphasis may be placed on issues like annotator agreement on segment boundaries, a factor that significantly impacts the reliability of segmentation studies but is often overlooked in research.

Papers may also address strategies for handling segment overlaps and ambiguities, which are particularly common in specialized texts. How can annotation guidelines be developed and refined to ensure greater consistency and accuracy in the segmentation of specialized discourse?

References

Carlson, L., & Marcu, D. (2001). Discourse tagging reference manual. *ISI Technical Report ISI-TR-545*, 54(2001), 56.

- Dahm, J. (2024 – forthcoming). Frames im Fachdiskurs der Logistikbranche. In L. Gautier & S. Varga (Eds.), *Frames und Fachwissen*. Berlin: De Gruyter.
- Egbert, J., Wizner, S., Keller, D., Biber, D., McEnery, T., & Baker, P. (2021). Identifying and describing functional discourse units in the BNC Spoken 2014. *Text & Talk*, 41(5-6), 715-737.
- Fillmore, C. J. (2006). Frame semantics. *Cognitive linguistics: Basic readings*, 34, 373-400
- Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3), 175-204.
- Hajicová, E., Partee, B. B., & Sgall, P. (2013). *Topic-focus articulation, tripartite structures, and semantic content* (Vol. 71). Springer Science & Business Media.
- Halliday, M. A. K. (1970). Language structure and language function. *New horizons in Linguistics/Penguin*.
- Hirschberg, J., & Litman, D. (1993). Empirical studies on the disambiguation of cue phrases. *Computational Linguistics*, 19(3), 501-530.
- Hoek, J., Evers-Vermeul, J., & Sanders, T. J. (2018). Segmenting discourse: Incorporating interpretation into segmentation?. *Corpus Linguistics and Linguistic Theory*, 14(2), 357-386.
- Lavissière, M. C., & Bonnard, W. (2024). Who's really got the right moves? Analyzing recommendations for writing American judicial opinions. *Languages*, 9(4), 119.
- Longacre, R. E. (1983). *The Grammar of Discourse*. Plenum Press.
- Polanyi, L. (1988). *A formal model of the structure of discourse*. *Journal of Pragmatics*, 12(5-6), 601-638.
- Rau, G. (2021). Development of component analysis to support a research-based curriculum for writing engineering research articles. *English for Specific Purposes*, 62, 46-57.
- Sacks, H. (1974). An analysis of the course of a joke's telling in conversation. In R. Bauman & J. Sherzer (Eds.), *Explorations in the Ethnography of Speaking* (pp. 337-353). Cambridge University Press.
- Swales, J. M. (1990). *Genre analysis*. Cambridge university press.
- Swales, J. M. (2004). *Research genres: Explorations and applications*. Cambridge University Press.
- Van Dijk, T. A. (2013). Semantic macro-structures and knowledge frames in discourse comprehension. In *Cognitive processes in comprehension* (pp. 3-32). Psychology Press.