**Anna Corazza, Simonetta Montemagni and Giovanni Semeraro (dir.)**

**Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016**
**5-6 December 2016, Napoli**

Accademia University Press

# Annotating Content Zones in News Articles

**Daniela Baiamonte, Tommaso Caselli and Irina Prodanof**

DOI: 10.4000/books.aaccademia.1695
Publisher: Accademia University Press
Place of publication: Torino
Year of publication: 2016
Published on OpenEdition Books: 26 July 2017
Serie: Collana dell'Associazione Italiana di Linguistica Computazionale
Electronic ISBN: 9788899982546

# Annotating Content Zones in News Articles

**Daniela Baiamonte**
University of Pavia
daniela.baiamonte01@universitadipavia.it

**Tommaso Caselli**
VU University Amsterdam
t.caselli@vu.nl

**Irina Prodanof**
University of Pavia
irina.prodanof@gmail.com

## Abstract

**English.** This paper presents a methodology for the annotation of the semantic and functional components of news articles (Content Zones, henceforth CZs). We distinguish between narrative and descriptive zones and, within them, among finer-grained units contributing to the overall communicative purpose of the text. Furthermore, we show that the segmentation in CZs could provide valuable cues for the recognition of time relations between events.

***Italiano.*** *In questo lavoro viene presentata una metodologia per l'annotazione delle componenti semantiche e funzionali del testo giornalistico (Zone di Contenuto). Distinguiamo tra zone narrative e descrittive e, al loro interno, tra ulteriori unità che contribuiscono al dispiegamento dello scopo comunicativo del testo. Inoltre, mostriamo che la segmentazione in Zone di Contenuto offre preziosi indizi per il riconoscimento delle relazioni temporali tra eventi.*

## 1 Introduction

The logical structure of a document, i.e. its hierarchical arrangement in sections, paragraphs, sentences and the like, reflects a functional organization of the information flow and creates expectations on where the desired information may be located. As it is often the case, however, breakups in sections and paragraphs are motivated by style or even arbitrary choices.

The segmentation of the text in Content Zones (CZs, henceforth), i.e. functional categories contributing to the overall message or purpose, as induced by the genre of the text[1], provides more reliable and fine-grained cues to access the structure of its *types of functional content*. Previous attempts to annotate CZs have mainly focused on highly standardized texts like scientific articles (Teufel et al., 2009; Liakata et al., 2012; Liakata and Teufel et al., 2010) and scheduling dialogues (Taboada and Lavid, 2003), or on semi-structured texts like film reviews (Bieler et al., 2007; Taboada et al., 2009). Other work (Palmer and Friedrich, 2014; Mavridou et al., 2015) adopts the theory of discourse modes (Smith, 2003) to distinguish between the different types of text passages in a text document.

To the best of our knowledge, no efforts have been undertaken to devise an annotation scheme targeting the functional structure of news articles in terms of their content: the inverted pyramid structure, i.e. the gathering of key details at the beginning, followed by supporting information in order of diminishing importance, is too coarse-grained to be effectively used for information extraction purposes. Our hypothesis is that modeling the document's content via CZs could yield benefits for high-level NLP tasks such as Temporal Processing, Summarization, Question-Answering, among others. In addition to this, CZs qualify as a higher-level analysis of a text/discourse which captures different information with respect to Discourse Relations. The remainder of the paper is structured as follows: in Section 2 the motivations of this work are presented, together with related studies. Section 3 reports on our inventory of CZs, used to annotate a corpus of English news articles. Details on the corpus are provided in Section 4. In Section 5, we describe a case-study on the correlation between CZs and temporal re-

[1] We adopt Systemic Functional Linguistics' view of genre as "a staged, goal oriented, purposeful activity in which speakers engage as members of our culture" (Martin, 1984:25).

lations to show that the segmentation in CZs can provide cues in recognizing temporal relations between events. Finally, Section 6 draws on conclusion and suggests directions for future work.

## 2 Motivations and related work

The bulk of the work on discourse structures has focused on low-level structures corresponding to Discourse Relations holding between textual segments pairs. CZs take a different view on texts, as they perform a function towards the text as a whole. As an instance of a particular genre, every text is meant to accomplish a culturally-established communicative purpose, e.g. a news article reports on events happening in the world. This goal is not accomplished all at once: separate functional *stages* (i.e. CZs) convey fragments of its overall meaning (Eggins and Martin, 1997). Therefore, the knowledge about the typical functional structure of genres can be exploited to predict the internal organization of a text. This kind of information can be of help to produce balanced summaries or to select the passages most likely to contain the answer to a question.

Teufel et al. (2009) and Liakata et al. (2010)'s works present two complementary perspectives on scientific papers: the former models their argumentative/rhetorical structure (following the knowledge claims made by the authors); the latter treats them as the humanly readable representations of scientific investigations. In the works of Bieler et al. (2007) and Taboada et al. (2009), two different kinds of zones are recognized in film reviews: formal zones (required by the genre, e.g. credits and cast) and functional zones (reflecting the abstract functions of describing and commenting).

In the elaboration of news articles' CZs, we were mostly inspired by Labov (2013)'s study of oral narratives of personal experiences and by Bell (1991)'s analysis of the structure of news stories.

## 3 Annotation Schema

The opposition between dynamicity and staticity, mainly realized by grammatical and lexical aspect, is adopted as the basic parameter for differentiating between two macro CZs: NARRATION and DESCRIPTION. The former is aimed at reporting temporally interrelated (dynamic) events, the latter is used to comment by focusing on selected entities, properties, and states of affairs. Each of these two macro CZs is further divided into more fine-grained categories.

The class NARRATION (NARR) includes the following zones:

- `Foreground` (FGR): text span containing the most salient events, i.e. those in the focus of attention (as intended by Boguraev and Kennedy, 1999). The information it conveys is both referentially and relationally new (Gundel and Fretheim, 2005), as it is usually mentioned at the beginning of the article.
- `Background` (BGR): ancillary, referentially and relationally old information performing an explanatory function (through causal and temporal precedence relations) towards FGRs.
- `Follow-up` (FUP): reactions and consequences to FGR events (to whom they're related through cause-effect and temporal succession relations), i.e. relationally new information moving the discourse forward.
- `Expectation` (EXP): assumptions and probable or possible outcomes, i.e. non factual information (e.g. conditionals, modality).

The class DESCRIPTION (DSCR) includes the following zones:

- `Description` (DES): characteristics of a person or an object, customary circumstances, or states of affairs.
- `Evaluation` (EVL): subjective descriptions, explicit judgements showing the author or some other agent's attitude towards a target.

In addition, a third macro-class is posited, OTHER (OTHR), containing categories performing auxiliary functions towards the other CZs:

- `Attribution` (ATT): text span containing the source and, if present, the cue of an attribution (as intended by Pareti and Prodanof, 2010) - while the content is assigned the relevant CZ(s).
- `Metatext` (MTX): text span guiding the reader's attention towards metatextual elements like figures or tables.
- `Interrogative` (INT): questions directly addressed to the reader, e.g. to introduce a new topic or to prompt a reaction.

Major approaches to functional discourse structuring adopt the sentence or the paragraph as unit of annotation. On the other hand, we have opted for a clause level annotation as this allows us to better deal with news articles' high level of information density. Although CZs are conceptually non-overlapping, empirical analysis indicates that an annotation unit may fit into more than one category, that is a clause may represent complex contents. Cases as such suggest that the more informative content should be preferred. In the example below, the tag ATT is assigned, even though a descriptive content may as well be recognized.

1. [On an office wall of the Senate intelligence committee hangs a quote from Chairman David Boren,]$_{ATT}$ {PDTB[2], wsj_0771}

The annotation of CZs is further complicated by the fact the distribution of the zones does not follow the linear order of the text. In most cases, CZs are discontinuous, that is either their contiguity may be "broken" by the presence of other CZs or the same CZ may appear in different sentences along the entire document (see example **??** for the FGR zone).

2. [South Korea registered a trade deficit of $101 million in October,]$_{FGR}$ [reflecting the country's economic sluggishness,]$_{EVL}$ [according to government figures released Wednesday.]$_{ATT}$ [Preliminary tallies by the Trade and Industry Ministry showed another trade deficit in October, the fifth monthly setback this year,]$_{FGR}$ [casting a cloud on South Korea's export-oriented economy.]$_{EVL}$ {PDTB, wsj_0011}

In other cases, due to the use of the clause as minimal annotation span of a CZ, nested CZs may occurr (see example **??**).

3. [South Korea's economic boom, [which began in 1986,]$_{BGR}$ stopped this year because of prolonged labor disputes, trade conflicts and sluggish exports.]$_{BGR}$ {PDTB, wsj_0011}

## 4 Description of the corpus

We used the CZs annotation schema and the annotation tool CAT (Bartalesi Lenzi et al., 2012) to construct a small corpus of 57 news articles

[2]Penn Discourse TreeBank (Prasad et al., 2008).

| **Tense** | **FGR** | **BGR** | **FUP** | **EXP** | **DES** | **EVL** | **ATT** | **MTX** | **INT** |
|---|---|---|---|---|---|---|---|---|---|
| PRESENT | 46 | 46 | 26 | 21 | 35 | 58 | 34 | 0 | 0 |
| PAST | 66 | 204 | 29 | 3 | 2 | 10 | 172 | 0 | 0 |
| FUTURE | 21 | 1 | 25 | 22 | 0 | 2 | 0 | 0 | 0 |
| INFINITIVE | 32 | 51 | 26 | 18 | 2 | 12 | 1 | 0 | 0 |
| PRESPART | 6 | 19 | 10 | 5 | 2 | 3 | 9 | 0 | 0 |
| PASTPART | 2 | 11 | 1 | 1 | 0 | 2 | 1 | 0 | 0 |
| NONE | 6 | 8 | 6 | 21 | 0 | 2 | 2 | 0 | 0 |

Table 1: Distribution of tenses among CZs.

(20 from the test section of TempEval-3 (UzZaman et al., 2013), 20 shared between the PDTB (Prasad et al., 2008) and the training section of the TimeBank (Pustejovsky et al., 2003), 17 from the PDTB). The corpus contains 2059 annotation units and it is dominated by narrative sections (57%). Within them, the most frequent CZ is the BGR (26.5%), followed by FGR (12.4%), EXP (9.6%) and FUP (8.4%). These figures show that news articles mostly consist of redundant information, only mentioned in order to help the reader to anchor the new data to the prior knowledge. Descriptive sections constitute the 25.5% of the corpus: EVLs are slightly more frequent than DESs (14.8% vs. 8.9%) — contradicting the alleged objectivity expected in news reports (note, however, that EVLs tend to occur in association with ATTs). As to the OTHER macro CZ, it makes up the 17.4% of the corpus: this percentage almost entirely refers to ATTs, since MTXs and INTs are only marginal zones (0.19% and 0.33%, respectively).

To test our hypotheses about some formal properties of CZs, we carried out a corpus study. The results are reported below.

**Position in the text**. 71.7% of FGRs are located in the opening sections of the articles and their occurrence decreases towards the central (18.4%) and closing sections (9.8%). BGRs show a fairly complementary distribution to FGRs, as they mostly occupy the central (31.6%) and closing sections (27.3%) of the articles. As expected, ATTs are quite evenly distributed among the three sections. The remaining CZs do not show any clear-cut tendencies.

**Verbal tenses**. Table **??** shows the distribution *target*

| *source* | FGR BEFORE | FGR INCLUDES | FGR DURING | FGR BEGINS | FGR ENDS | FGR SIMULTANEOUS | FGR IDENTITY | BGR BEFORE | BGR INCLUDES | BGR DURING | BGR BEGINS | BGR ENDS | BGR SIMULTANEOUS | BGR IDENTITY | FUP BEFORE | FUP INCLUDES | FUP DURING | FUP BEGINS | FUP ENDS | FUP SIMULTANEOUS | FUP IDENTITY |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FGR | 46 | 13 | 1 | - | 2 | 12 | 8 | 1 | 1 | - | - | - | 2 | 2 | 9 | - | - | - | - | - | 1 |
| BGR | 16 | 1 | - | 1 | - | 1 | 8 | 91 | 45 | 1 | 3 | 2 | 41 | 8 | 6 | - | - | - | - | - | - |
| FUP | 1 | 4 | - | - | - | 1 | 4 | - | - | - | - | - | - | - | 33 | 6 | - | 1 | 6 | 10 | - |
| EXP | - | 1 | - | - | - | - | 1 | - | - | - | - | - | - | 1 | - | - | - | - | - | - | - |
| DES | 2 | 1 | - | - | - | - | 1 | - | 1 | - | - | - | - | 2 | 2 | 1 | - | - | - | 1 | - |
| EVL | 1 | 1 | - | - | - | 1 | 3 | 2 | 2 | - | - | - | 1 | 1 | 2 | 2 | - | - | - | 1 | - |
| ATT | 16 | 1 | - | - | - | 1 | 1 | 18 | 2 | - | - | - | 4 | - | 18 | 1 | - | - | - | 1 | - |
| MTX | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| INT | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |

| *source* | EXP BEFORE | EXP INCLUDES | EXP DURING | EXP BEGINS | EXP ENDS | EXP SIMULTANEOUS | EXP IDENTITY | DES BEFORE | DES INCLUDES | DES DURING | DES BEGINS | DES ENDS | DES SIMULTANEOUS | DES IDENTITY | EVL BEFORE | EVL INCLUDES | EVL DURING | EVL BEGINS | EVL ENDS | EVL SIMULTANEOUS | EVL IDENTITY |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FGR | 2 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 1 | 1 | 3 |
| BGR | 7 | - | - | - | - | - | - | 3 | - | - | - | - | - | - | 5 | - | - | - | - | - | - |
| FUP | 1 | - | - | - | - | - | - | - | - | - | - | - | 1 | 1 | 2 | 1 | 1 | - | - | - | - |
| EXP | 24 | 10 | - | 1 | 1 | 7 | - | - | - | - | - | - | - | - | 2 | 2 | - | - | - | - | - |
| DES | 5 | - | - | - | - | - | 1 | 4 | - | 2 | - | - | 2 | - | 1 | 1 | - | - | - | 1 | - |
| EVL | 4 | - | - | - | - | - | - | - | - | - | - | - | 1 | - | 15 | - | - | - | - | 5 | - |
| ATT | 21 | 1 | - | - | - | - | - | 3 | - | - | - | - | - | - | 2 | - | - | - | - | 1 | - |
| MTX | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| INT | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |

| *source* | ATT BEFORE | ATT INCLUDES | ATT DURING | ATT BEGINS | ATT ENDS | ATT SIMULTANEOUS | ATT IDENTITY | MTX BEFORE | MTX INCLUDES | MTX DURING | MTX BEGINS | MTX ENDS | MTX SIMULTANEOUS | MTX IDENTITY | INT BEFORE | INT INCLUDES | INT DURING | INT BEGINS | INT ENDS | INT SIMULTANEOUS | INT IDENTITY |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FGR | 38 | 7 | - | - | - | 1 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| BGR | 81 | 5 | - | - | - | 5 | 2 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| FUP | - | - | - | - | - | 3 | 1 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| EXP | 2 | 1 | - | - | - | 1 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| DES | 5 | 3 | - | - | - | 1 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| EVL | 7 | 3 | 2 | - | - | 5 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| ATT | 14 | 5 | - | - | - | 44 | 21 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| MTX | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| INT | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |

BEFORE, INCLUDES, DURING, BEGINS, ENDS, SIMULTANEOUS, IDENTITY

Table 2: Distribution of time relations among CZs.

of verbal tenses, as annotated in the TimeBank corpus, among CZs. BGRs and ATTs are dominated by the past tense, this is in accordance with our expectations as the former is characterized by temporal precedence relations to FGR events and the second mostly contains events of saying. CZs belonging to the DSCR class are significantly dominated by the present tense, usually associated with imperfective aspect and staticity. The high frequency of present tenses in FGRs and BGRs doesn't necessarily defy our expectations, since FGRs contain both dynamic and static events and the tag PRESENT is also used to refer to instances of present perfect.

**Modality markers**. The majority of modality markers is located in EXPs and, more broadly, in the narrative CZs, as shown in Figure **??**. In the TimeBank corpus, the MODALITY tag is mostly assigned to modal auxiliaries, we believe that the annotation of modal adverbs would further raise the percentages observed in EXPs and in the NARR class.

**Pronouns**. Looking at Figure **??** we can see that almost 50% of all pronouns is located in BGRs. The percentages are consistent with our expectations as BGRs convey referentially old information and, although FUPs and EXPs elaborate on FGR events, they often introduce new referents. Note that the distribution of pronouns is not, alone, a sufficient indicator of referential oldness since also lexical and zero anaphoras should be taken into account.

## 5 Interactions between CZs and time relations

In news articles events are not iconically presented in the linear order of their real succession, this poses a challenge to systems aimed at uncovering their temporal event structure. Therefore, we used the annotations available for the TimeBank section of the corpus to check whether some connections between CZs and temporal relations between event pairs exist. The full set of temporal relations specified in TimeML contains 14 types of relations: BEFORE, AFTER, IBEFORE, IAFTER, BEGINS, BEGUN_BY, ENDS, ENDED_BY, DURING, DURING_INV, INCLUDES, IS_INCLUDED, SIMULTANEOUS and IDENTITY. We simplified the set as follows: the relation types that invert each other were collapsed into a single type; given the low frequency of the relation type IBEFORE, it was mapped to the corresponding more coarse-grained type BEFORE.

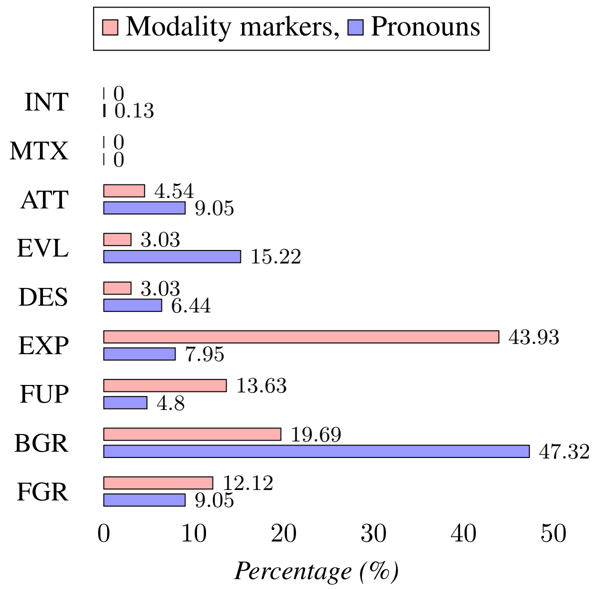


Figure 1: Distribution of modality markers and pronouns among CZs.

Given the narrative shape of news articles, the corpus is considerably dominated by precedence

| **source - target** | **BFR** | **INCL** | **DUR** | **BEG** | **ENDS** | **SMLT** | **IDNT** |
|---|---|---|---|---|---|---|---|
| NARR - NARR | 218 | 77 | 1 | 6 | 11 | 71 | 26 |
| NARR - DSCR | 17 | 1 | 1 | 0 | 1 | 2 | 7 |
| NARR - OTHR | 119 | 9 | 0 | 0 | 0 | 10 | 3 |
| DSCR - DSCR | 30 | 5 | 3 | 0 | 0 | 12 | 6 |
| DSCR - NARR | 20 | 8 | 0 | 0 | 0 | 4 | 6 |
| DSCR - OTHR | 16 | 10 | 2 | 0 | 0 | 6 | 0 |
| OTHR - OTHR | 14 | 5 | 0 | 0 | 0 | 44 | 21 |
| OTHR - NARR | 72 | 5 | 0 | 0 | 0 | 6 | 0 |
| OTHR - DSCR | 6 | 0 | 0 | 0 | 0 | 1 | 1 |

Table 3: Distribution of time relations among the macro-classes.

(BEFORE) and succession (AFTER) relations. Table **??** shows that the majority of temporal relations holds between events belonging to the same CZ types: events tend to precede, include, occur during, begin, end, be simultaneous to and anaphorically evoke (through TimeML IDENTITY temporal relations) other events belonging to the same zone.

FGR events precede rather than follow ATT, FUP and EXP events. BGR events, the most involved in BEFORE relations, tend to precede other events, especially if located in ATTs and FGRs. Unexpected outcomes mostly occur in cases like the following, where the FGR event precedes the BGR one. This is because conflicting contents may be expressed in the same unit (in this case a reaction to the FGR event and the list of its premises):

4. [Delta Air Lines earnings soared to 33% to a record in the fiscal first quarter,]$_{FGR}$ [bucking the industry trend toward declining profits.]$_{FUP}$ [The Atlanta-based airline, the third largest in the U.S., attributed the increase to higher passenger traffic, new international routes and reduced service by Rival Eastern Airlines...]$_{BGR}$ {PDTB, wsj_1011}

As highlighted in Table **??**, NARR events begin or end other NARR or DSCR events (more specifically, these relations hold between events belonging to instances of the same CZ) and DSCR events include (rather than being included in) other events.

IDENTITY relations mostly involve FGRs: as a result of their textual salience, FGR events can be mentioned in other FGRs or further clarified in narrative or descriptive sections.

## 6 Conclusions and future work

We have developed an inventory of zone labels for the genre *news article* and shown that the so-generated *content structure* could help narrowing down the range of time relations connecting events.

Future work would involve testing the stability and reproducibility of the annotation scheme through the measurement of inter-annotator agreement and elaborating a separate annotation scheme for editorials, whose argumentative style reflects different structuring principles than those acting in news reports. Finally, we would like to automatize the process of annotation and test the effectiveness of the approach in texts belonging to different genres, e.g. novels (Ouyang and McKeown, 2014) and historical essays. Even the basic distinction between narrative and descriptive zones could facilitate the performance of more complex NLP tasks by targeting the relevant informational zones. The corpus and the annotation guidelines are publicly available[3].

## Acknowledgment

This has been partially supported by the Erasmus + Traineeship Program 2015/2016 from University of Pavia and the NWO Spinoza Prize project Understanding Language by Machines (sub-track 3).

## References

Baiamonte, D. 2016. *Annotazione di Zone di Contenuto: una strutturazione funzionale del testo giornalistico*. Thesis of the Master in Theoretical and Applied Linguistics. University of Pavia, Pavia.

Bärenfänger, M., Hilbert, M., Lobin, H., Lüngen, H., Puskás, C. 2006. Cues and constraints for the relational discourse analysis of complex text types - the role of logical and generic document structure. Sidner C.L., Harpur J. Benz A., Kühnlein P. (eds.), *Proceedings of the Workshop on Constraints in Discourse*. Maynooth, Ireland. 27-34.

Bartalesi Lenzi, V., Moretti, G., Sprugnoli, R. 2012. CAT: the CELCT Annotation Tool. *Proceedings of LREC 2012*. Istanbul.

[3] `https://github.com/cltl/ContentZones.git`

Bell, A. 1991. *The Language of News Media.* Blackwell Publishers, Oxford.

Bieler, H., Dipper, S., Stede, M. 2007. Identifying Formal and Functional Zones in Film Reviews. Keizer S., Bunt H., Paek T. (eds.), *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*. Antwerp, Belgium. 75-78.

Boguraev, B. and Kennedy, C. 1999. Salience-based content characterisation of text documents. Inderjeet M. and Maybury M. T. (eds.), *Advances in Automatic Text Summarization*. MIT Press, Cambridge, MA.

Eggins, S. and Martin, J. R. 1997. Genres and registers of discourse. van Dijk T. (ed.), *Discourse Studies. Discourse as structure and process*, volume 1. Sage, London (UK) and Thousand Oaks (CA). 230-257.

Gundel, J. K. and Fretheim, T. 2005. Topic and Focus. Horn L. and Ward G. (eds.), *The Handbook of Pragmatics*. Blackwell Publishing, 175-196.

Labov, W. 2013. *The Language of Life and Death*. Cambridge University Press, Cambridge, UK.

Liakata, M., Saha, S., Dobnik, S., Batchelor, C., Rebholz-Schuhmann, D. 2012. Automatic recognition of conceptualization zones in scientific articles and two life science applications. *Bioinformatics 2012*, volume 28. 991-1000.

Liakata, M., Teufel, S., Siddharthan, A., Batchelor, C. 2010. Corpora for conceptualisation and zoning of scientific papers. *Proceedings of the 7th conference on Internation Language Resource and Evaluation (LREC10).*

Martin, J. R. 1984. Language, register and genre. Christie F. (ed.), *Language studies: Children writing. Reader.* Deakin University Press, Geelong, Australia. 21-30.

Mavridou, K., Friedrich, A., Peate Sørensen, M., Palmer, A., and Pinkal, M. 2015. Linking discourse modes and situation entity types in a cross-linguistic corpus study. September 2015. In Proceedings of Linking Models of Lexical. *Sentential and Discourse-level Semantics (LSDSem).* , Lisbon, Portugal.

Ouyang, J. and McKeown, K. 2014. Towards automatic detection of narrative structure. *Proceedings of LREC14*, Reykjavik, Iceland.

Palmer, A. and Friedrich, A. 2014. Genre distinctions and discourse modes: Text types differ in their situation type distributions. *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing. Forlí-Cesena, Italy.*

Pareti, S. and Prodanof, I. 2010. Annotating Attribution Relations: Towards an Italian Discourse Treebank. *Proceedings of LREC10.*

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., Webber, B. 2008. The Penn Discourse TreeBank 2.0. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC).* Marrakech, Morocco.

Pustejovsky, J., Hanks, P., Sauri, R., See, A., Day, D., Ferro, L., Gaizauskas, R., Lazo, M., Setzerr, A., Sundheim, B. 2003. The TimeBank Corpus. *Corpus Linguistics.* 647-56.

Smith, S. Carlota 2003. *Modes of discourse: The local structure of texts.* Cambridge University Press.

Stede, M. 2011. *Discourse Processing.* Morgan & Claypool Publishers. 7-38.

Taboada, M., Brooke, J., Stede, M. 2009. Genre based paragraph classification for sentiment analysis. *Proceedings of SIGDIAL 2009: the 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue.* Queen Mary University of London. 62-70.

Taboada, M. and Lavid, J. 2003. Rhetorical and Thematic Patterns in Scheduling Dialogues: A Generic Characterization. *Functions of Language, 10(2).* 147-179.

Teufel, S., Siddharthan, A., Batchelor, C. 2009. Towards discipline-independent argumentative zoning: Evidence from chemistry and computational linguistics. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, EMNLP 2009.* Singapore.

UzZaman, N., Llorens, H., Derczynski, L., Allen, J. Verhagen, M., Pustejovsky, J. 2013. SemEval-2013 Task 1: TempEval-3: Evaluating Time Expressions, Events, and Temporal Relations. *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*