# Exploring Flexibility in Natural Language Generation through Discursive Analysis of New Textual Genres

Marta Vicente and Elena Lloret

University of Alicante
{mvicente,elloret}@dlsi.ua.es

**Abstract.** Since automatic language generation is a task able to enrich applications rooted in most of the language-related areas, from machine translation to interactive dialogue, it seems worthwhile to undertake a strategy focused on enhancing generation system's adaptability and flexibility. It is our first objective to understand the relation between the factors that contribute to discourse articulation in order to devise the techniques that will generate it. From that point, we want to determine the appropriate methods to automatically learn those factors. The role of genre on this approach remains essential as provider of the stable forms that are required in the discourse to meet certain communicative goals. The arising of new web-based genres and the accessibility of the data due to its digital nature, has prompted us to use reviews in our first attempt to learn the characteristics of their singular non-rigid structure. The process and the preliminary results are explained in the present paper.

## 1 Motivation

Working to develop language generation tools represents, nowadays, a challenging task that could substantially benefit a large number of applications. Natural language generation (NLG) is more than a stand-alone task, its progress impacts on and enriches multiple language-related areas (machine translation, question answering or summarization, for example). Furthermore, its outcomes can be applied in generation of image and video description (involving computer vision, artificial intelligence) along with construction of dialogues, recommendation systems or even instructions for interactive games. Multiple applications whose progress can be boosted by the explosive growth of visual and textual resources and the techniques developed to process them.

Nevertheless, for the time being, the developments on NLG are strongly determined by domain and application, they do not present enough awareness of their context. To introduce pragmatic concerns in order to obtain adaptable systems, it is imperative to understand how that elements affect the compounding of the outcome, how its structure or its contents are concerned. There are multiple factors that can modify how a text is articulated, from the targeted audience

to the intention or the communicative aim that is pursued. From a linguistic perspective, is in the figure of the genre that those features take shape.

We believe that we must pay special attention to this matter and looking into that relation poses the main goal of our current work. Additionally, the presence of Internet as a communication channel has fostered the emergence of new genres, as digital reviews or blogs, and the access to that data has become relatively straightforward. Under those circumstances, we have undertaken an approach whose initial steps are reported in the present paper.

The remainder of this document is organised as follows. Section 2 provides further insights on the task of text generation and, specifically, on the stage of Macroplanning. In section 3, we explain some remarks on discourse and genre. Next, on section 4, a deeper analysis of the relation between genre and structure is reported. Relevant research on common areas is summarised on section 5. In order to detect patterns, a clustering methodology has been selected, and is presented in section 6. The first experiments devised are described in section 7 and finally, section 8, outlines some general conclusions and future work.

## 2   Text Generation and Macroplanning

From a computational perspective, the generation of language is a highly complex process that must be tackled on the premise that the system's outcome will meet certain linguistic and communicative requirements (syntactic, semantic, pragmatic concerns, context related). Much of the work developed until now is focused on expressing some information, which can be given in the form of related concepts, raw data or even texts, through utterances whose elaboration can range from a single sentence to a much more complex structure (sets of paragraphs compounding discourses - reports, articles, stories, for instance). To achieve this, an NLG system must handle two main questions: *"**what** should be said?"* and *"**how** to realise it on the desired output?"*. Concerning the initial stage, this is known in literature as Macroplanning, and it refers both to the selection and planning of the content, processes that result in a structure called Document Plan [23]. Therefore, the Document Plan could be understood as the guideline for the rest of the process as well as the source of its meaning. Being able to encapsulate in such guidance the relevant content regarding the context implies the incorporation of a pragmatic facet of the linguistic realm which is the main objective of the present research on the long term. To understand and learn how this pragmatic issue affects the content and the structure of the discourse in order to reproduce it, have become our leading goal at the present time. We relate our investigation to the Discourse Theory, the relevance of the genre in the construction of the text and the relationship between genre and the intentionality that lies beneath the use of natural language as an act of communication.

## 3    Generating the Discourse

When NLG is concerned with the production of text, not as character sequences or single utterances but as sequence of paragraphs, it is fundamental to consider the relation between sentences in order to assure the cohesion and coherence of the text obtained [1]. Discourse theory establishes that to be called discourse, a text must be connected and structured. On the one hand, connection would be provided by linguistic cohesion or by implicatures. On the other hand, the structure would proceed from the contents represented, the intentional and the interactional structure that arise from the goals and conventions of the activity to which the discourse is related: its specific genre [11].

Some theories of the discourse [2, 12] have highlighted the fact that genres can be understood as social constructions, as providers of the connection between the discourse and the situation in which it is produced. They condition both the structure and the content of the text generated. According to Swavels [27]:

> "A genre comprises a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognised by the expert members of the parent discourse community, and thereby constitute the rationale for the genre. This rationale shapes the schematic structure of the discourse and influences and constraints choice of content and style."

There is yet another definition of genre proposed by Bhatia [5] that also contributes to the fundamentals of our approach:

> "Genre essentially refers to language use in a conventionalised communicative setting in order to give expression to a specific set of communicative goals of a disciplinary or social institution, which give rise to *stable structural forms* by imposing constraints on the use of lexico-grammatical as well as discoursal resources."

It is the detection of such patterns or stable forms the first step we are embracing in order to design the mechanisms able to produce natural language involving such discursive features.

Taking this into account, it can be observed that different genres accept different degrees of structural rigidity. In the case of news and research articles, patents, or Wikipedia articles, for instance, some structure can be detected even from a shallow analysis. However, there are other types of texts for which such structure is not as easily perceptible. It is usual in some of the genres intrinsically related to Internet and the new forms of communication arising with it. Reviews, blogs, tweets are examples of texts that can be included in the latter case.

## 4    Genre and Structure

In order to clarify the novelty of our research, it is necessary to introduce some assumptions regarding the notion of structure that is being pursued. In some

---

[1] Actually, coherence has been accepted as a quality indicator [29]

sense, text structure may be interpreted as the order of the ideas that should be expressed. However, it is not only the order what we have to determine, but the full characterisation of the parts of the discourse: their nature, their relevance for the outcome and, directly linked to our main primary goal, the way their constituents contribute to their purpose or communicative goal, and in turn, to the general aim of the text.

In this regard, the Systemic Functional Theory [12] should be brought up. It provides a notion of discourse and genre that connects situation types with semantic/lexico-grammatic patterns from a conception of language highly related to its socio-semiotic origin. Moreover, coming from this approach, their classification of communicative goals or *"fields of activity"* (*expounding, reporting, recreating, sharing, doing, enabling, recommending, exploring*) along with the related work developed on this line will serve as a cornerstone for our research [21].

Consequently, genres become valuable to our approach since they are meant to bear some communicative purpose: news function would be to inform; reviews, to persuade or fairy tales, to entertain. Nevertheless, it is possible a further analysis , detecting those functions in the sections of the discourse. This would be the case of the parts of any scientific paper or the classical narrative structure: exposition-climax-resolution, that we could relate to the fields of activity previously mentioned. In both cases, the order of such parts must remain the same, in the sense that it would be incoherent to present the conclusions of a research before the results, for example. And this stability provides a benefit when computing is involved.

However, for some genres, the structure is not perceived in a straightforward manner. We worked with reviews, since this genre illustrates such phenomenon. Besides, reviews represent a multi-purpose type of texts depending on several parameters as the object (e.g. books, hotels) or the intention of the writer (e.g. recommend, criticise). In respect to sections and their functions, they contain at the same time parts that expound the personal experience of the user, but also those that convey why a user went to that hotel or chose that book, the description of the room or the plot, or recommendations for anyone reading.

In table 1, manual analysis over hotel reviews shows both differences and similarities between two documents. In this case, there are three types of sections, relating respectively personal experience, description of the hotel (amenities, building, restaurants,...) and evaluation. Each review presents different number and order of the parts, but the language on each of them displays some visible regularities. For instance, the sentences in the descriptions share the use of the verb "to be" in a similar way (e.g. *the hotel is, the grounds were, The room was*); first person and past tense for sentences in personal experience fragments (e.g. *we visited this, we travelled in, I had stayed*) and so on.

The flexibility of such kind of structure leads to problems when automatic processes tackle the generation of this type of texts. Therefore, several questions arise: *how a text to be generated should be organised or planned in order to be coherent? in the case of reviews, are they written in a similar fashion, even*

| |
|---|
| **Review 1** |
| **Description (1 sentence):** |
| *A beautiful hotel, in a perfect location, with large rooms, and an architectural...* |
| **Personal Experience (8 sentences):** |
| *We visited this resort for our 25th wedding anniversary.* |
| *We traveled in March, which tends to be the high rain season...* |
| *It rained every day, as we were in the midst of a minor storm...* |
| **Description (19 sentences):** |
| *The hotel is a very open concept design that sits on 50 acres of the Poipu Beach...* |
| *The grounds were beautiful, and the interior of the hotel was open to the outside...* |
| *In the lobby were a beautiful collection of live, native birds, of magnificent...* |
| **Evaluation (2 sentences):** |
| *This hotel is a real winner, complete with it's own 18 hole golf course.* |
| *Although I would be cautious about the time of year I visited, I would say...* |
| |
| **Review 2** |
| **Personal Experience (3 sentences):** |
| *In New York for New Years 2003,I had chosen to stay at the Essex House because...* |
| *However, I became soon disappointed once I entered my standard room...* |
| *It appears to me that the renovation money did not go into the standard rooms...* |
| **Description (6 sentences):** |
| *My room was very small as it was difficult to maneuver within the room's limited space.* |
| *The room was dull with worn Old English furnishing.* |
| *There where handles missing off cabinet doors and the room lacked isolation...* |
| **Personal Experience (1 sentence):** |
| *I had stayed at Westin hotels before, and was never disappointed until now.* |
| **Description (7 sentences):** |
| *The standard rooms lack modern amenities.* |
| *They appear to be outdated of at least 10 to 20 years.* |
| *The corridors are decorated with non tasteful wallpaper, and the paint...* |
| **Evaluation (1 sentence):** |
| *Overall, I am very disappointed with the hotel, however I had a good time for...* |

**Table 1.** Hotel review ordering from a functional approach. Even at first sight some different features can be distinguished among parts (Personal Experience, Description, Evaluation): verb tenses, syntactic constructions, lexicon...

*not showing that rigid structure?* These questions would be worth investigating to address the Macroplanning stage, being that related to genres and their communicative objectives.

Regarding current NLG systems, it is normally assumed a pre-defined structure for a text in their developments [14, 16, 10]. Nonetheless, most of the cases are focused on very specific domains, such as weather reports [18, 17] or specific genres, as Wikipedia articles [26]. It should be noted, however, that some steps have been taken in order to increase flexibility in several areas. This is the case, for example, with storytelling, where the style of narration depends on the communicative goal it is pursued (to entertain, to motivate, to report facts) or on the character's perspective. The work in [19] reflects those pragmatic concerns.

Given this context, our present aim is to explore how to articulate document plans for texts that do not exhibit a structure at a first sight. To achieve this objective, a preliminary study within a set of hotel reviews is conducted, by which using clustering techniques applied to various lexical features, we would be able to determine whether heterogeneous reviews written by a wide range of users from different backgrounds, ages, and nationalities share similarities in their structure or not. Our analysis could be adapted to other genres, since we want to devise a methodology to analyse and determine the structure of a text in an automatic manner.

## 5    Existing approaches addressing Macroplanning and text structure

As indicated, Macroplanning consists of selecting and structuring the content that a generation system has to deliver as text. Back in 2003, Dubue and McKeown [8] attempted an approach by which content selection rules could be learnt pairing biographical texts with some semantic information, that would be the selected content. Trying to extend this approach, Barzilay and Lapata [4], included the relations coming from a database, which allowed them to reflect dependency between the involved items, and thus incorporate some structural information, but still very limited. More aware of the structural component, some approaches attempted to determine the distribution of the topics within the document to provide coherence to the result [3], or to extract content templates from an specific domain collection of documents that would incorporate that structure [26, 18, 17]. However, this became its main weakness, the domain restricting the performance of the systems.

Regarding the new digital genres previously mentioned, like the reviews, the research effort has been directed mainly to the area of sentiment analysis or polarity classification [24, 7], but few have been focused on their structure related to text genre. An approach that strongly relies on Systemic Functional Theory can be found in [28], outlining the relation between the genre itself and the structure of the document. An analysis over a corpus of movie reviews is carried out, where the results showed the presence of some parts concerning its purpose that were finally reduced to two main sections: evaluation and description.

There is interesting work on detecting the structure of text using patterns of lexical co-occurrence to identify paragraphs related to the same topic [13]. Taking into account term repetition, subtopics in explanatory texts were recognised, but results were not so good for other genres. Moreover, other features of the discourse were not considered (e.g. syntactic constructions, verb tenses, discourse markers) neither an approach on the functions of the parts of the text.

Bachand's research [1], from another perspective, is focused on the relations between text-type, discourse structures and rhetorical relations. Only considering this time rhetorical relations and markers, the good results obtained indicate that similar but enriched developments can lead to a better generation of accurate document plans.

Our research is novel in the sense that to the best of our knowledge, there is no previous work that attempts to determine and extract the document plan from a *flexible-structure* text genre, such as reviews. Our analysis will set the basis for generating a wide range of texts types, regardless the domain or text genre they belong to.

## 6 Preliminary Analysis: Clustering to reach Document Plans

Since the aim of our research is to come up with possible document plans after analysing not strictly organised texts, clustering techniques seem appropriate to detect some patterns across reviews. To accomplish our analysis, the Expectation Maximisation Algorithm (*EM*) implementation provided by Weka has been selected, due to the fact that it allows unsupervised learning over unlabelled data, which is our case. More information on *EM* can be found in [30].

***Feature Extraction and Analysis.*** For now, and regarding the features, our approach is preliminary because there is no set of them already consolidated. We will refine the selection and feed the experiments with several collections to determine the most appropriate for Macroplanning. About the labelling, part-of-speech tags are considered at different levels (a word can contribute as a verb, as a verb in past tense, etc). Named entities are also labelled. From a semantic perspective, verbs are classified as belonging to one of six categories: mental, material, relational, verbal, existential or modulation [9].

## 7 Experiments and Preliminary Results

This section describes with some detail the experiments devised to tease out some insights on the structure of reviews. First, the corpus and resources that were used to labelled it and, afterwards, the explanation of the different criteria applied and the analysis of results.

### 7.1   Corpus

The corpus used in our experiments is a collection of reviews gathered from Tripadvisor, regarding ten different hotels. It consists of roughly 1,400 reviews written in Spanish. The reviews were segmented in sentences (12,467 sentences). Around 200,000 words have been automatically labelled.

| | |
|---|---:|
| Number of reviews | 1400 |
| Sentences | 12,467 |
| Words labelled | approx. 200,000 |

**Table 2.** Corpus statistics.

### 7.2   Resources

The sentences were analysed and labelled with Freeling [22] using its lemmatiser, its PoS-Tagger and its named entity recogniser. Verb lemmas were confronted with the list of verb meanings from the *ADESSE* project [9] in order to establish their categories.

| **ADESSE verb senses** |
|---|
| Mental, material, relational, verbal, existential and modulation |
| **FREELING features** |
| PoS tagging: noun, adjective, pronoun, verb (tense, aspect, ...), etc. |

**Table 3.** Features annotated over the corpus of reviews.

### 7.3   Review Segmentation and Data Preparation

Firstly, a decision was made about gathering the information that could be computed from the labelled sentences. In this manner, clustering was applied to the corpus four times, each of them over the reviews divided from one to four sections. Our objective was to identify certain relations within those sections and between them. The length of each segment was the result of a equitable division of the number of sentences involved. This would allow us to discover whether reviews tend to condense specific information within a particular area, despite being different in their nature and content. In other words, although two

reviews can be written differently, they may contain a section (a set of sentences) expressing similar ideas or written to meet same purposes (e.g. informing about the advantages of the hotel). Even if the area is in different location, some patterns could be extracted.

It is when the division of the text has been set that the process starts computing the features of the sentence constituents and the aggregation of the results according to the number of sections. In order to properly compare the different reviews, we settle the following criterion: taking into account one review divided in sections and a feature *f*, we calculate for each section the percentage of *f* regarding the total occurrence of this feature in the review. As a result, two values were obtain: i) the maximum percentage calculated, and ii) the section in which this feature is prevalent.

Once this is computed, the Weka implementation of the *EM* algorithm is conducted considering the independence of the features and performing cross-validation.

### 7.4 Results and Discussion

At the moment, the results obtained are analysed with respect to the following issues: i) considering only the area in which the maximum percentage of the feature appears (*Area*), ii) considering only the maximum percentage (*%*) ; and iii) considering simultaneously both records (*Area+%*).

For each execution Weka calculates a number of clusters (Table 4) providing the number of instances included in them.

| Divisions | Area | % | Area+% |
|-----------|------|----|--------|
| 1 | 1 | 1 | 1 |
| 2 | 13 | 2 | 4 |
| 3 | 8 | 2 | 5 |
| 4 | 10 | 10 | 4 |

**Table 4.** Number of clusters obtained for each experiment, being:i) considering only the area in which the maximum percentage of the feature appears (*Area*), ii) considering only the maximum percentage (*%*) ; and iii) considering simultaneously both records (*Area+%*)

Regarding the amount of reviews enclosed on some clusters, some facts seem interesting. For example,in the case of clustering reviews divided in two parts and experiment *iii)*, the percentage of reviews for each cluster is 39% (0), 2% (1), 44% (2) and 15% (3). It is remarkable to find out how this distribution arises and what are its peculiarities. It is not an isolated case. On the other hand, the distribution of features when the number of clusters is high could be as interesting of this last to discard some relations or study some regularities.

A further examination may revel how could be possible from each cluster to expose correlation or co-occurrence between the features considered, and in a future step, to bound such relations with some communicative goals.

## 8    Conclusion and Future Work

Questions suggested earlier in this paper can now be addressed from a more informed perspective. How a NLG system would perform the production of a review and if so, could an inner structure be revealed by some automatic analysis? As a result of our experimentation and its analysis, we expect to find the definition of the review genre attending the characteristics of its structure and constituents. Such description would provide appropriate document plans to select and organise the information according to the inner traits of the text genre, and similar methodology could be applied to develop other text types.

Some of the future experiments involve changing the number of features, in order to obtain a comprehensive description, including the sentiment traits, specific semantic characteristics regarding the elements and its discursive relation, and pragmatic information related to the context of the production. Alternatively, the corpus will be increased at different stages to include firstly several types of reviews and afterwards, several types of genres.

Finally, we will take into account some of the resources rooted in Web Semantic technologies. Some research regarding genre have been done in this field. There is the *MARL Ontology Specification*[2] that could be appropriate for reviews, opinion and sentiment annotation. This data schema has been used in the *EuroSentiment Project* [6] and, from a Sentiment Analysis perspective, on reviews [25]. On News, it is relevant the effort invested by the *BBC*, in order to obtain more significant annotation of the documents. They provide a set of ontologies related to their contents and area. *DBPedia* has been already proved useful for Wikipedia articles researchers. *Drammar* [20] and *OntoMedia* [15] are ontology-based models for annotating features of media and cultural narratives. All of them resources that could make a difference in our clustering task and analysis.

## Acknowledgments

---

[2] http://www.gsi.dit.upm.es/ontologies/marl

# References

1. Bachand, F.H., Davoodi, E., Kosseim, L.: An investigation on the influence of genres and textual organisation on the use of discourse relations. In: International Conference on Intelligent Text Processing and Computational Linguistics. pp. 454–468. Springer (2014)
2. Bakhtin, M.M.: Speech genres and other late essays. University of Texas Press (2010)
3. Barzilay, R.: Probabilistic approaches for modeling text structure and their application to text-to-text generation. In: Empirical methods in natural language generation, pp. 1–12. Springer (2010)
4. Barzilay, R., Lapata, M.: Collective content selection for concept-to-text generation. In: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. pp. 331–338. Association for Computational Linguistics (2005)
5. Bhatia, V.: Worlds of written discourse: A genre-based view. A&C Black (2004)
6. Buitelaar, P., Arcan, M., Iglesias Fernandez, C.A., Sánchez Rada, J.F., Strapparava, C.: Linguistic linked data for sentiment analysis (2013)
7. Cambria, E., Schuller, B., Xia, Y., Havasi, C.: New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems 28(2), 15–21 (2013)
8. Duboue, P.A., McKeown, K.R.: Statistical acquisition of content selection rules for natural language generation. In: Proceedings of the 2003 conference on Empirical methods in natural language processing. pp. 121–128. Association for Computational Linguistics (2003)
9. García-Miguel, J.M., Vaamonde, G., Domínguez, F.G.: Adesse, a database with syntactic and semantic annotation of a corpus of spanish. In: LREC (2010), `http://dblp.uni-trier.de/db/conf/lrec/lrec2010.html#Garcia-MiguelVD10`
10. Ge, T., Pei, W., Ji, H., Li, S., Chang, B., Sui, Z.: Bring you to the past: Automatic generation of topically relevant event chronicles. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). pp. 575–585. Association for Computational Linguistics, Beijing, China (July 2015), `http://www.aclweb.org/anthology/P15-1056`
11. Gruber, H., Redeker, G.: The Pragmatics of Discourse Coherence: Theories and Applications, vol. 254. John Benjamins Publishing Company (2014)
12. Halliday, M., Matthiessen, C.M., Matthiessen, C.: An introduction to functional grammar. Routledge (2014)
13. Hearst, M.A.: Texttiling: Segmenting text into multi-paragraph subtopic passages. Computational linguistics 23(1), 33–64 (1997)
14. Hu, Y., Wan, X.: Automatic generation of related work sections in scientific papers: An optimization approach. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 1624–1633. Association for Computational Linguistics, Doha, Qatar (October 2014), `http://www.aclweb.org/anthology/D14-1170`
15. Jewell, M.O., Lawrence, K.F., Tuffield, M.M., Prugel-Bennett, A., Millard, D.E., Nixon, M.S., Shadbolt, N.R., et al.: Ontomedia: An ontology for the representation of heterogeneous media. In: In Proceeding of SIGIR workshop on Mutlimedia Information Retrieval. ACM SIGIR (2005)
16. Jha, R., Finegan-Dollak, C., King, B., Coke, R., Radev, D.: Content models for survey generation: A factoid-based evaluation. In: Proceedings of the 53rd Annual

Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). pp. 441–450. Association for Computational Linguistics, Beijing, China (July 2015), `http://www.aclweb.org/anthology/P15-1043`

17. Kondadadi, R., Howald, B., Schilder, F.: A statistical nlg framework for aggregated planning and realization. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 1406–1415. Association for Computational Linguistics, Sofia, Bulgaria (August 2013), `http://www.aclweb.org/anthology/P13-1138`

18. Konstas, I., Lapata, M.: A global model for concept-to-text generation. Journal of Artificial Intelligence Research 48, 305–346 (2013)

19. Li, B., Thakkar, M., Wang, Y., Riedl, M.O.: Storytelling with adjustable narrator styles and sentiments. In: Interactive Storytelling, pp. 1–12. Springer (2014)

20. Lombardo, V., Damiano, R.: Semantic annotation of narrative media objects. Multimedia Tools and Applications 59(2), 407–439 (2012)

21. Matthiessen, C.M.: Registerial cartography: context-based mapping of text types and their rhetorical-relational organization (2014)

22. Padró, L., Stanilovsky, E.: FreeLing 3.0: Towards Wider Multilinguality. In: Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). European Language Resources Association (ELRA) (2012)

23. Reiter, E., Dale, R., Feng, Z.: Building natural language generation systems, vol. 33. MIT Press (2000)

24. dos Santos, C.N., Gatti, M.: Deep convolutional neural networks for sentiment analysis of short texts. In: COLING. pp. 69–78 (2014)

25. Santosh, D.T., Vardhan, B.V.: Feature and sentiment based linked instance rdf data towards ontology based review categorization. In: Proceedings of the World Congress on Engineering. vol. 1 (2015)

26. Sauper, C., Barzilay, R.: Automatically generating wikipedia articles: A structure-aware approach. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1. pp. 208–216. Association for Computational Linguistics (2009)

27. Swavels, J.: Genre analysis: English in academic and research settings. Cambridge University Press (1990)

28. Taboada, M.: Stages in an online review genre. Text & Talk-An Interdisciplinary Journal of Language, Discourse & Communication Studies 31(2), 247–269 (2011)

29. Webber, B., Joshi, A.: Discourse structure and computation: past, present and future. In: Proceedings of the ACL-2012 Special Workshop on Rediscovering 50 Years of Discoveries. pp. 42–54. Association for Computational Linguistics (2012)

30. Witten, I.H., Frank, E., Hall, M.A.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3rd edn. (2011)