

THE USE OF GRAMMATICALLY ANNOTATED CORPORA FOR THE DISPLAY OF TEXTUAL PATTERNS

Aleksandra Bizjak Končar, Fran Ramovš Institute of the Slovenian Language, Scientific
Research Centre of the Slovenian Academy of Sciences and Arts

This paper looks at how meanings unfold across the genre of Contemporary Slovenian Sermons under the application of tools developed within the fields of systemic functional linguistics and corpus linguistics. The theoretical construct for the analysis is provided by Hasan's (1984) concept of genre and Cloran's (1994) concept of message semantics. To determine consistency of textual patterns for the genre of sermons, lexico-grammatical patterns in sermons are examined. It is argued that genre analysis, supported by visual presentation of lexico-grammatical patterns (as suggested by Biber et al. 1998) extracted from a grammatically annotated corpus of Slovenian sermons, provides a fuller picture of crucial properties of genre.

KEYWORDS: textual patterns, grammatical annotation, genre analysis visual presentation technique

1 INTRODUCTION

The last three decades have seen a growing research to find reliable criteria for classifying texts into types. Although a number of theoretical frameworks have been proposed within different fields of linguistics, they usually run out of explanatory power when applied to more complex discourses. One reason for this could be that, whereas the higher levels of context and discourse have been studied extensively, they have not been related to the levels of words and grammar in a systematic way. This question of how genres are realised lexico-grammatically needs to be explored further. To clarify that last point, the models need to be tested against a range of linguistic data before their general applicability can be claimed. It is only by observing real data that new facts can be discovered and language models modified accordingly. This paper is in line with the need for the extension and application of theoretical models to a wider range of discourse genres that exist in our cultures.

Our starting point for genre analysis is with data. Using a specialized corpus of Slovenian contemporary sermons, we will attempt to demonstrate how meanings unfold across a genre of texts. To be able to investigate the above question empirically, and not take sermons alone as examples for interpretation, one needs to start at the lexico-grammatical level and trace surface grammatical features that have to do with meanings. This is because it is only the appeal to meanings that tells us that texts in the corpus belong to the same genre-repertoire of sermon. Here the term MEANING

refers to “textual semantic components” (Hasan 1973:286) that are “relevant to structuring of the overall message form of the text” (Hasan 1989:103). The notion of relationship between meaning and wording is central to Hasan’s concept of genre (1978, 1984, 1989). How, exactly, the task of establishing relationship between meaning and wording is carried out, is a question that Hasan does not answer explicitly. Despite that we decided to apply her model, which is based on Halliday’s semantically driven model of language description, to sermonic discourse. Our reasons were: (i) It seemed to us that the gap between the lower level of lexico-grammar and the higher levels of discourse semantics can be filled by using the metalanguage developed by Hasan’s and Cloran’s work on message semantics. (ii) As Hasan’s model of identification of genre makes it possible to convey the rich complexities at play between the discourse, semantic and lexico-grammatical levels for the genre under consideration, it is especially suitable for the analysis of discourse meanings as well as for tracing lexico-grammatical features.

The above-addressed concerns are further elaborated in Section 2 and 3 which deal with the issue of combining relevant elements of genre and corpus-based analysis in order to provide a sound empirical base that allows “a clear link between theoretical claim and data” (Bateman et al. 2002:2). In some respects our paper follows the model of language and text proposed by Steiner (2003), trying to connect micro and macro levels of language description in order to provide grounds for empirically based discussion of language phenomena. How the automatic extraction of the lexico-grammatical patterns that are of generic significance is carried out in actual texts is exemplified in Section 4. Section 5 then interprets the results obtained from the corpus of Slovenian sermons and argues that by making textual patterns accessible to linguistic analysis we are able to reveal crucial properties of genre.

2 THEORETICAL FRAMEWORK AND THE NOTION OF TEXTUAL PATTERNS

As mentioned in Section 1, the model chosen for the purpose of corpus-based analysis of GENRE is Hasan’s theory of generic structure potential (1984). One of the most attractive features of her model is that it depends on the notion of generic structures that are identified first in semantic terms and then further specified in terms of lexico-grammatical patterns capable of realising specific semantic properties. Therefore, according to Hasan’s three levels of abstraction, one should be able to define genre: (i) at the level of discourse at which text structures are defined and generic structure is formulated; (ii) at the level of semantics, related to which is the statement about “crucial semantic attributes” (Hasan 1984:84) of the elements proposed at the discourse level; (iii) the lexico-grammatical level at which the description of lexico-grammatical patterns realising semantic properties of the elements in question is attempted. The semantic level is therefore the “key to the whole system” (Halliday 1974) as it is the semantic system that systematically relates the high-level concepts of meanings to their linguistic realisations at the level of lexico-grammar.

When it comes to applying Hasan’s theoretical model to complex types of discourses in our society, where a text’s semantic elements are realised by long segments of texts and not by particular utterance, one is faced with a problem. The lexico-grammatical analysis, which uses Halliday’s techniques, is essentially a device for

handling clausal phenomena. But text is a “structure of a higher order” (Butt 1988:78). The core of the problem lies in the fact that the relationship between semantics and grammar is not ‘one-to-one’. Using Halliday’s words “...one element in the semantic system is realised by more than one in the lexico-grammatical system” (Halliday 1974:91).

There is no need to enter here into a full discussion of the relationship between clause and text. We will simply note two issues that are relevant to semantics of text. Butt, in his article “Randomness, Order and the Latent Patterning of Text”, argues that patterns at the level of semantics, “which are realized across a text or across genre of texts” (1988:76) are not easily explicated. The first point that must be considered is the general problem of meta-language for the semantic level of a linguistic system. Butt argues that in order to discuss the level of semantics it is necessary to involve other levels of the linguistic system or refer to human experiences as such, and points out that “the terms of semantic description are revaluation of terms that do service in other environments” (1988:78). In order to deal with the first problem, we will design the framework on the basis of which meanings can be approached. The framework will be introduced in Section 2.1.

The second point that must be made here is the fact that textual patterns are established “through an ensemble” of invisible patterns (Butt 1988:78). These patterns are discussed also by Sinclair and Coulthard (1975:127-128) who acknowledge that textual patterns are not easily recognized at the lexico-grammar level because they carry discourse meanings. The above problem can only be exemplified by the analysis of textual patterns in texts, which is the central concern of this paper in Section 4.

In this section we have dealt with the notion of textual patterns, trying to give it some theoretical status within systemic functional linguistics. In what follows in Section 2.1 we shall be concerned with the framework on the basis of which meanings in text can be approached.

2.1 THE FRAMEWORK FOR DESCRIPTION OF TEXTUAL PATTERNS

In the search for a framework which would be able to provide some engineering beyond what grammar can handle, Hasan’s and Cloran’s writings on message semantics offer some solutions (Hasan 1991, Cloran 1994). Based on Hasan’s semantic networks (1991), Cloran proposed a new method of describing the meanings of a text (1994). In order to comprehend Cloran’s semantic analysis, one needs to be acquainted firstly with Hasan’s work on message semantics. For the purpose of this paper, we will highlight only the points of utmost importance for our analysis of sermons.

Hasan’s approach to description of meaning involves seeing MESSAGE as the smallest semantic unit which “serves as the environment for semantic choices” (Cloran 1994:148). Her way of approaching the analysis of meaning is by postulating that message is realised by the lexico-grammatical unit of CLAUSE. But, she contends, to make statements about the realisation patterns, the semantic networks for the experiential, interpersonal and textual domains need to be described in greater or lesser detail, depending on the purpose of analysis.

Cloran builds her theoretical framework on Hasan’s message semantics, but also develops further some semantic descriptions for the message components called

ENTITY and EVENT (1994: 175-248). Cloran's approach to description of meaning in terms of message components allows us to exemplify the relationship between the structure of entity and event on one hand, and Subject and Finite patterns on the other. By using semantic networks for Entity and Event developed by Cloran it is possible to state semantic properties of text in lexico-grammatical terms, which is a first condition when one wants to approach the investigation of semantic patterns from the empirical stand-point. For now we will leave the detailed description of Cloran's semantic networks aside, to return to it later in Section 4.1 when the micro and macro levels will be taken into account while displaying typical textual patterns in sermonic discourse. Before doing that, we must look a little more closely at some of the tools and concepts used in corpus linguistics by means of which we can provide grounds for the empirical investigation of sermons. Let us now proceed to Section 3 which introduces corpus collection of contemporary Slovenian sermons, linguistic annotation and visual presentation technique.

3 METHODOLOGICAL ISSUES

As was already pointed out our paper is concerned with the question of how textual patterns are realized across a genre of sermons. The stimulus for tracking lexico-grammatical features over the course of a text came from the work of Biber et al. (1998:166-130) although there are significant modifications when it comes to applying his technique to the analysis of sermons, as will be demonstrated in section 3.3. The two steps that enable us to investigate specific lexico-grammatical features for large quantities of data are: (i) the construction of a representative corpus; (ii) the linguistic annotation that supports the question to be investigated. In the following two sections (Section 3.1 and 3.2) the issues related to the corpus compilation and corpus annotation of Contemporary Slovenian sermons will be discussed in more detail.

3.1 A CORPUS OF CONTEMPORARY SLOVENIAN SERMONS

The language of sermons is a vast field so restriction of the scope of the study is necessary. Therefore, in the present paper only sermons belonging to the stream of Roman Catholicism, and, specifically, contemporary sermons, are considered.

In order to collect a representative sample of the genre under consideration it was originally planned that the material for the analysis would be obtained by recording sermons held during Sunday church services. But when collecting the sample data, a number of obstacles were encountered. To mention only two: (i) transcribing of sample sermons caused significant problems due to background noise; (ii) there were some complications involved in getting permission to do recording of sermons prepared by different preachers. In light of this, the decision was made to collect sermons broadcaste on our national television. Sermons were video-recorded, transcribed and computerized. It is important to realize, in connection with this, that the task of collecting corpus from scratch is extremely time-consuming. Our work on the construction of corpus of Contemporary Slovenian sermons lasted five years. Over this time we succeeded in collecting a corpus of 50 sermons. The corpus was considered to

be representative of the genre of sermons and allows for results to be applicable to the population of contemporary sermons in general.

The data yielded are introduced in more detail in the table 1. The sermons are arranged in columns according to the place (second column) and date of the recordings (third column).

Table 1: Sermons arranged according to place (second column) and time (third column) of recording/broadcasting

TEXT	PLACE	DATE
TEXT 1	Murska Sobota	1. 3. 1998
TEXT 2	Kidričevo	19. 4. 1998
TEXT 3	Primskovo pri Kranju	3. 5. 1998
TEXT 4	Piran	17. 5. 1998
TEXT 5	Dolsko pri Mengšu	7. 6. 1998
TEXT 6	Ljubljana, Koseze	21. 6. 1998
TEXT 7	Šempeter v Savinjski dolini	5. 7. 1998
TEXT 8	Kočevje	19. 7. 1998
TEXT 9	Mengeš	2. 8. 1998
TEXT 10	Petrovče	16. 8. 1998
TEXT 11	Beltinci	6. 9. 1998
TEXT 12	Kamenje	20. 9. 1998
TEXT 13	Višnja Gora	18. 10. 1998
TEXT 14	Slovenska Bistrica	15. 11. 1998
TEXT 15	Štanjel	6. 12. 1998
TEXT 16	Šmarje – Sap	20. 12. 1998
TEXT 17	Ljubljana – Vič	10. 1. 1999
TEXT 18	Otočec	24. 1. 1999
TEXT 19	Podlehnik	7. 2. 1999
TEXT 20	Ajdovščina	4. 4. 1999
TEXT 21	Dol pri Ljubljani	18. 4. 1999
TEXT 22	Žužemberk	9. 5. 1999
TEXT 23	Ribnica na Pohorju	23. 5. 1999
TEXT 24	Dobrna	18. 7. 1999
TEXT 25	Tržič	1. 8. 1999
TEXT 26	Brezje	15. 8. 1999
TEXT 27	Ponikve	5. 9. 1999
TEXT 28	Šmihel pri Pivki	3. 10. 1999
TEXT 29	Vojnik	21. 11. 1999
TEXT 30	Trebnje	5. 12. 1999
TEXT 31	Maribor	13. 2. 2000
TEXT 32	Tomaj	12. 3. 2000
TEXT 33	Bloke	9. 4. 2000
TEXT 34	Žiri	14. 5. 2000

TEXT	PLACE	DATE
TEXT 35	Teharje	4. 6. 2000
TEXT 36	Hrenovice pri Postojni	2. 7. 2000
TEXT 37	Železniki	1. 10. 2000
TEXT 38	Ljubljana	7. 1. 2001
TEXT 39	Maribor	18. 2. 2001
TEXT 40	Drežnica	4. 3. 2001
TEXT 41	Kresnice	1. 4. 2001
TEXT 42	Komenda	5. 8. 2001
TEXT 43	Šturje pri Ajdovščini	2. 9. 2001
TEXT 44	Ponikva	7. 10. 2001
TEXT 45	Korena	4. 11. 2001
TEXT 46	Cerknica	2. 12. 2001
TEXT 47	Ljubljana, Šentjakob	6. 1. 2002
TEXT 48	Maribor	3. 2. 2002
TEXT 49	Krkavče	3. 3. 2002
TEXT 50	Slovenj Gradec	7. 4. 2002

But corpus that does not contain additional linguistic information is not of much use when attempting to investigate more complex grammatical and semantic questions. In the next section (3.2) we will see how the corpus of Contemporary Slovenian sermons was annotated using a part-of-speech tagger developed specifically for Slovenian.

3.2 PART-OF-SPEECH TAGGING OF CORPUS

Much of the recent work in the field of corpus annotation falls within the scope of grammatical annotation or part-of-speech tagging (O'Donnell 2005). Over the past few years, part-of-speech tagging has also spread into the domain of those languages termed Central&Eastern European languages. During the same period of time, a Slovenian part-of-speech tagger has been developed as part of the project to build a Slovenian National Corpus at the Fran Ramovš Institute of the Slovenian language. The Slovenian tagger is principally the work of Jakopin et al. (1997). In what follows, some characteristics of the Slovenian tagger will be presented.

Let us first say a few words about the development of our annotation scheme. Besides Slovenian grammar some other sources were considered while assembling the tagset: the tagset of the Brown corpus, the Penn Treebank corpus and the tagset used in the frame of MULTTEXT/East project. The main criteria were the legibility of the tags and the minimal size required for disambiguation. The tags had to be short, derived from Slovenian wording of linguistic terms and self-explanatory to the extent that they would be not only machine-readable, but acceptable to human readers as well. The point is illuminated in the following sentence:

Seveda se lahko motijo. ('Certainly they may be wrong?')
 Č Gmp A Gcp

The tag Č stands for particle, Gmp for separate verbal morpheme, A for adjective, and Gcp for verb, third person, plural. We will not go further into details of the Slovenian tagset here, but it should be mentioned, taking into account current debates on the standardisation of annotation schemes, that our tagset can be easily converted to XML.

The software for the support of the tagging process has developed over time from a simple tool used as an aid for manual tagging to a two-step disambiguator in 1997. From the very beginning the tagger has been a part of text-editor EVA and therefore totally interactive. It is possible to tag just a single word or proceed from a selected place in text. The lexicon of words, their tags and tag frequencies is updated on the fly as another file, which can also be edited and adjusted when required. This file acts as a database which is used for production of different statistical tables that give the necessary feedback for fine-tuning of the tagger and the tagset.

The disambiguator has two steps. The first one is based on the history of tagged text: if a word form in the lexicon has one and only one tag, it is the right one; if not, the neighbourhood of the word has to be examined. If there exists a neighbourhood, from 2 to 5 words long (including the word in question), that has a match in the history database, with one and only one set of tags, the word is given its tag from this set; otherwise it is left untagged.

The second step is a probabilistic one, where the frequencies of all possible sets of tags for immediate tag neighbourhood (again 2 to 5 tags deep) are considered. If there is one and only one such set for any of the possible neighbourhoods, the corresponding tag is given to the observed word; otherwise the word is left untagged again.

In the field of POS tagging of texts in the Slovenian language, the first step has been made. A proven tagset, a disambiguator with an accuracy rate of 80 %, and the supporting software, both incorporated into an own text editor for effective and comfortable use, have been established. The wish list for the future includes a bigger database and a better disambiguator, with an accuracy rate of 90 % or better.

The corpus of Slovenian contemporary sermons was automatically tagged using the above described automatic POS tagger for Slovenian. We have already mentioned that our tagger is 80 percent accurate. Therefore, the final step was checking of tagged sermonic corpus and when necessary the tags were corrected. It is the grammatically annotated corpus of Slovenian sermons that enabled further automatic analysis of lexico-grammatical patterns in sermons as shown in Section 3.3.

3.3 VISUAL PRESENTATION TECHNIQUE

As described thus far, the Slovenian tagger has been from the very beginning a part of text-editor EVA, developed by Primož Jakopin. That was an advantage as it was possible to write a computer program that would extract combinations of lexico-grammatical features over the course of a text and display the results graphically as suggested by Biber et al. (1998: 126).

To get to the visual representation of textual patterns we followed the procedures described by Biber et al. (1998:126-130). That is, the computer program was written to cycle through sermons that were grammatically tagged and noted down linguistics features of verbs and nouns. With regard to the technique for visual

representation of data, our analysis takes as an example Biber's study (Biber et al. 1998). But when it comes to the concept of data our approach is more in line with Steiner's suggestion to "develop a linguistically richer and theoretically better motivated notion of data" (2005:10). We introduced the notion of textual patterns that are identified in semantic terms and further specified in terms of lexico-grammatical patterns. We were able to do that by incorporating the notion of textual patterns within the systemic functional framework and by relating it to the notion of genre and message semantics as pointed out in Section 2. Therefore, the extraction of data from the grammatically tagged corpus of contemporary Slovenian sermons has to do with the selection of grammatical features that are criterial to defining entities and events in semantic terms. How the textual patterns are realized across sermons will be exemplified in Section 4, which brings together two threads of the analysis: the one that has to do with encoding of meanings and the one that deals with the interpretation of meanings.

4 ANALYSIS

To this point our discussion of textual patterns in contemporary Slovenian sermons was, in a sense, abstract, focussing mainly on those elements of discourse semantics and corpus linguistics that are relevant to our empirically based investigation of the genre of sermons. This section is the culmination of this paper, bringing together the two strands of the analysis. It provides, however, only a sample analysis of one sermon that illustrates our approach (Section 4.1) and serves as a first introduction to discussion of textual patterns revealed by examining the corpus of contemporary Slovenian sermons (Section 4.2).

4.1 TEXTUAL PATTERNS OF SERMONS

As mentioned in Section 3, there are a number of steps to be taken to get to the visual representation of the results. The first step is the grammatical annotation of electronically stored texts. The second step is performed by a computer program that cycles through texts and records certain linguistic features of nouns and verbs that are critical to defining semantic features of entities and events. The final step is the graphic display of entity and event patterns. For the purpose of this section, which is concerned with graphical presentation of textual patterns in a text, there are just two further points to be made.

The first is that, in terms of the description of textual patterns, which are being expressed through a particular organisation and combination of entities and events, much of what counts as the realisation of semantic options on the semantic level would not necessarily be accounted for in our presentation of the results of the analysis. In order to see how the lexico-grammatical patterns are related to semantic options, we used the semantic networks for entity and event developed by Cloran. Her networks, which are schematised in Figure 1 and 2, specify the lexico-grammatical realisation of semantic options for entity and event.

Figure 1: Semantic options for Entity (from Cloran 1994: 242, Slovenian translation in brackets)

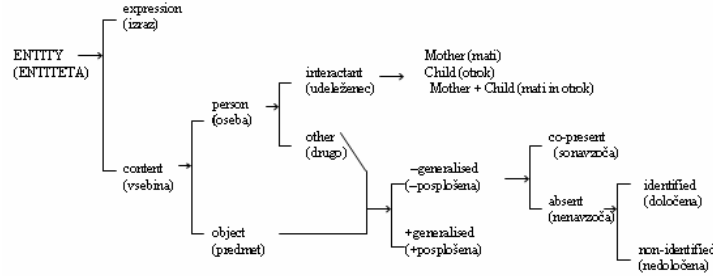
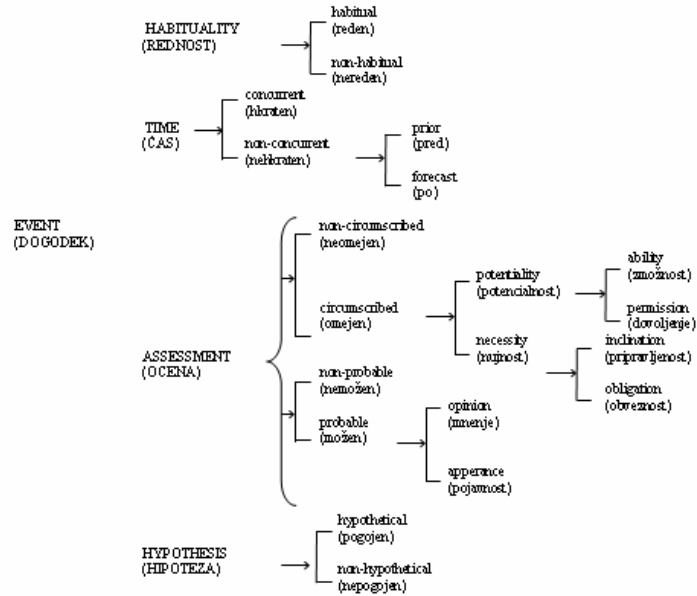


Figure 2: Semantic options for Event (from Cloran 1994: 221, Slovenian translation in brackets)



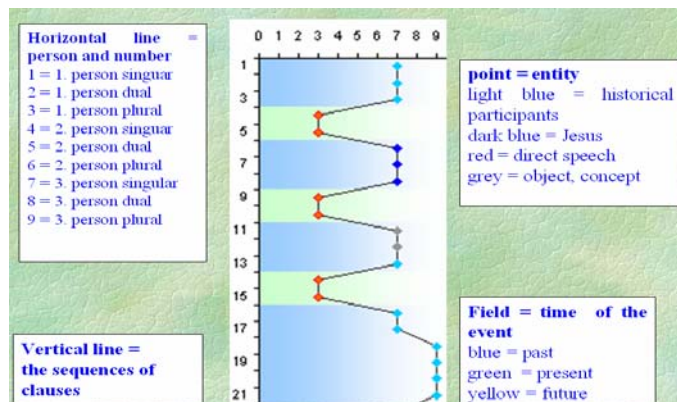
With the help of Cloran’s networks it is possible to extract lexico-grammatical features from grammatically annotated corpus automatically and turn them into semantic features. No attempt will be made here to describe all the options available in the network. However, we will use the selected parts of her network during our discussion of sermons in order to demonstrate how lexico-grammatical patterns are related to semantic options.

The second point that must be made here has to do with the text selected for illustration of our approach. For the reader to be able to follow the presentation of the analysis of the Slovenian text some additional time and space would be required for translation and explanation of particular features of Slovenian language. For example, *dual*, which is preserved in our language, and other features of inflectional languages, all

have bearing on the corpus analysis tools and the visual presentation of data. For someone not acquainted with the Slovenian language some further clarifications would be needed to enable them to follow the interpretation without too many unnecessary digressions from the main line of discussion. For the above reasons, the decision was made to select an English sermon in order to illustrate our approach.

Figure 3: Textual patterns for clauses 1-21

1. A leper came to him.
2. He fell on his knees in front of him.
3. He pleaded with his right hand over his mouth.
4. “If you want to,
5. you can heal me!”
6. Jesus looked at him, that creature of his, his brother,
7. and once again he was touched with pity.
8. He said:
9. “Of course I want to.
10. Be healed!”
11. Having learned
12. from what had happened to him before,
13. he added:
14. “Please don’t tell anyone except the priests
15. you need to tell to get your health certificate.”
16. The man not only went to those priests
17. he told the story left, right, and center.
18. And they came again, the hundreds and the thousands,
19. looking for the easy thing – his touch –
20. while remaining unchanged.
21. They wanted to be healed like that leper, only on the surface.



The sermon selected for the purpose of illustration is “*Don’t tell anyone*” by J. Donders (1988:96-97) (see Appendix for the full text). The results of the graphic-technique type

of analysis for the English sermon are shown in Figures 3 and 5. Figure 3 displays textual patterns for the sequence of clauses 1-21. Figure 5 shows patterning for the next sequence of clauses 22-51. For the convenience of the reader, the analysed paragraphs will be written down above the graphic presentation. The analysis begins with the explanation of Figure 3.

Let us first explain how to read the map by looking at Figure 3. The vertical axis represents the sequence of clauses, moving from the top to the bottom of the map. The horizontal axis represents the number and person features. The coloured points represent different classes of entities and coloured fields represent the time of event. Thus, in Figure 3, which displays the progression of entities and events for clauses 1-21, the light blue point of the first sentence corresponds to the entity leper. Light blue points always refer to entities denoting historical persons that belong to the book of Scriptures and can be called Scriptural characters.

Looking at the option of the event we can see that the first message construes the meaning *past*, which is represented by blue colour of the field. All the events that refer to previous occurring time are blue.

In order to see how the lexico-grammatical patterns are related to semantic options we used Cloran's semantic networks for entity and event (1994: 221, 242). The following options can be stated for the first paragraph.

Figure 4: Semantic options for entities and events in Figure 1

<p>ENTITY Content → person → other → (-)generalised → absent: historical characters (Jesus, leper ...)</p>
<p>EVENT time → non-concurrent → prior → past tense time → non-concurrent → prior → historical present, direct speech</p>

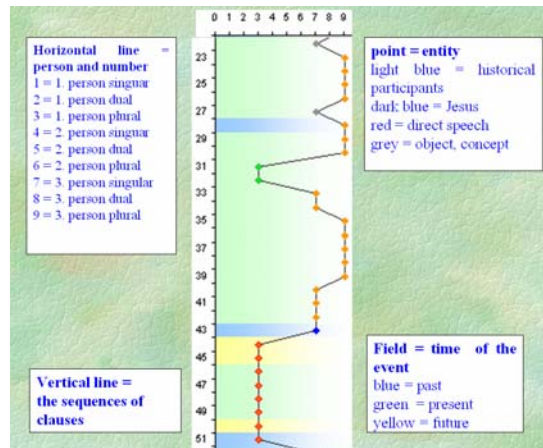
As Figure 4 shows, the Subject role of the first paragraph is focussed on historical characters that belong to the book of Scriptures and can thus be called Scriptural characters. Looking at the options of Finite we can see that present appears only in the "direct speech" or is used for specific stylistic purposes (historical present), while the rest of the messages construe the meaning "previous once-occurring historical time". When the Subject domain and time areas are brought together, it is significant that the focus is on historical characters and their placement in the previous historical time.

The discussion of the next paragraph of the sermon (clauses 22-51) will proceed in the same manner. The results of the analysis of the clauses 22-51 are displayed in Figure 5.

Figure 5: Textual patterns for clauses 22-51

- 22. It is a bit like those patients
- 23. doctors sometimes speak of.

24. Sick people
25. who are looking for an injection, an ointment, a powder, or a pill
26. but who are not willing to change the eating or drinking habits, the smoking, the life style,
27. that is really the cause of their sickness.
28. They were a bit like those sufferers from VD
29. who are quite ready to undergo all kinds of treatment
30. but are not prepared to stop their promiscuity.
31. Like all of us
32. who are willing to discuss all the ills of society
33. and how they should be overcome-
34. how the goods should be redivided,
35. how the poor should be helped,
36. how the drunkards should be healed,
37. how the prostitutes should be saved,
38. how the orphans should be fed,
39. how the wars should be stopped-
40. in meeting after meeting, minuting everything carefully in writing.
41. Yet nothing ever happens,
42. because nothing goes deeper than our skins.
43. Jesus said:
44. "I will heal your skin.
45. I will heal the surface,
46. but don't tell anyone.
47. Keep it a secret.
48. Because it is not all.
49. If you tell,
50. they will never understand
51. what I really came for."



Let us now examine the second paragraph, clauses 22-51, from the perspective of entities and events. In contrast to the first paragraph, where the focus is on the past once-occurring historical time, the messages of the second paragraph (Figure 5) construe the present time, which is identified in the map as a green field. As far as the Subject role for this paragraph is concerned we can distinguish the following categories according to Cloran's semantic network:

- (i) contemporary participants who are co-present are called interactants and are denoted by green points (see clauses 31, 32);
- (ii) contemporary participants who are absent, generalized entities are denoted by orange points (see clauses 23-30 and 34-42);
- (iii) the Jesus of faith who is regarded as co-present due to his divine nature is denoted by dark blue points (see clause 43).

In order to establish the distinctness of the second paragraph, Subject and Finite structures are summarised using Cloran's semantic network for entity and event (1994: 221, 242). For the second paragraph the following options can be stated:

Figure 6: Semantic options for entity and event for clauses 22-51

ENTITY
content → person → interactant: we
content → person → other → (+) generalised: contemporary participants (sick people, the poor, the prostitutes)
EVENT
time → concurrent → habitual

In contrast to the first paragraph, in the second paragraph the selection of Subjects includes participants that present the whole class of people or the typical behaviour of humans in general. The typicalness of behaviour is highlighted by selecting simple present, the tense that, according to Halliday (1985), when used with material and verbal processes is extended beyond the meaning 'here-and-now' to denote "habitual actions". When simple present is used with relational and existential processes, and the Subject represents a class of people, the meaning of the present becomes "habitual" as well.

We can conclude this short illustration of textual patterns, which are being expressed through a particular organisation and combination of entities and events, by arguing that the focus of the first paragraph is on the historical characters and their placement in the previous historical time, the time of Jesus. As the presentation of individual characters, their actions and behaviour in once-occurring events convey the meaning of individuation, we called this element Scriptural event.

In contrast to the first paragraph, in the second paragraph the selection of Subjects includes participants that present the whole class of people or the typical behaviour of humans in general. Therefore, typicalness and habitualness are two distinctive properties of the element called Contemporarisation.

We have tried in this section to show that graphical presentation of textual patterns, combined with a systemic functional model of language description, can help

us understand semantic movement of text. When the whole corpus of sermons is considered, patterns become far more elaborate. But what we would like to bring to your attention is the fact that every single sermon in the corpus displays the two distinctive elements: the Scriptural event and the Contemporarisation. What is there, then, about these two elements that brings them into such close interdependence? The answer will be sought in Section 4.2.

4.2 INTERPRETATION OF THE RESULTS

Based on the analysis of all texts in the corpus we tried to answer the question: what is the reason for the establishment of the elements of Scriptural event and Contemporarisation in sermonic discourse? The analysis reveals that in Scriptural event and Contemporarisation, the exposition of ethical principles takes place through the presentation of positive and negative ethical options. As the goal is not only to expose the believers to knowledge but also to persuade them to live according to Jesus's teaching, real-life situations are created in the positive and negative frames. On the basis of observation of the right and wrong behaviour of individual characters in historical situations, the believers are able to judge from their own experiences which actions call for imitation. It can, therefore, be claimed that the Scriptural event and the Contemporarisation construe an argument which is not based on scientific proof but on reasoning from experience (such as observation, imitation, testimony). The organisation of the Scriptural event and Contemporarisation into an argument provides the reason for the close interdependence of the two elements. It is found that Contemporarisation can be interspersed with the Scriptural event or realised as a discrete element. The comparison of texts where the Contemporarisation is discrete shows that the motivation for the division or non-division is related to the notion of how the historical event corresponds with the believer's experiences.

Reports of miracles or the presence of Jesus as divine and human in the historical event are happenings that run counter to believers' everyday experiences about the laws of nature. In this case, the pedagogical tool of analogy is used and a domain of similar contemporary situation is established by which the behaviour of historical participants can be understood in contemporary perspective, despite the cultural and temporal distance. The habitualness and typicalness of human behaviour is presented through the domain of contemporary situation.

When temporal distance is not a barrier for understanding the behaviour of historical participants in contemporary perspective, the semantics of individuation, typicalness and habitualness is interspersed, and the realisation of Contemporarisation non-discrete.

It can, therefore, be concluded that the close dependency of the domain of historical situation and contemporary situation reflects the fact that the main method of argumentation in sermonic discourse is based on everyday experiences and leads believers towards a Christian life-style and faith.

5 CONCLUSION

By applying the systemic functional model of language description and concepts of message semantics to the genre of sermons it was possible to show that sermons display an argumentative structure. Furthermore, by bringing together tools of systemic functional linguistics and corpus linguistics it was possible to reveal textual patterns, which are expressed through a particular organisation and combination of entities and events. Finally, the analysis, supported by detailed textual examples from corpus, sheds light on the complicated question of patterning at the level of semantics and makes the semantic movements – which are veiled in the language and perceived by the listener/reader only intuitively – evident and accessible to linguistic analysis.

This study suggests that further research along the lines of combining systemic functional theory and corpus linguistics can yield fruitful results as it makes the linguistic investigation more empirical on one side and on the other side helps to bring the lower levels of wording and higher levels of meaning closer together. Let me end this paper by quoting Butt (1988:97): “The clarification of lexico-grammatical patterning is necessary to the clarification of semantic design.”

REFERENCES

- Bateman, John, Judy Delin & Renate Henschel. 2002. *The motivation, design and implementation of a multilayered corpus for multimodal documents: the GeM model*. <http://www.fb10.uni-bremen.de/anglistik/langpro/projects/gem/newframe.html>
- Biber, Douglas, Susan Conrad & Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Butt, David. 1988. Randomness, order and the latent patterning of text. In D. Birch & M. O’Toole (eds.), *Functions of Style*, 74-97. London: Pinter Publishers.
- Cloran, Carmel. 1994. *Rhetorical Units and Decontextualization: An Enquiry into some Relations of Context, Meaning and Grammar*. (Monographs in Systemic Linguistics, 6) Nottingham: Department of English Studies, University of Nottingham.
- Donders, Joseph G. 1988. *Praying and Preaching the Sunday Gospel*. New York: Orbis Books.
- Halliday, Michael A. K. 1974. Interview. In Hermann Paret (ed.), *Discussing Language*, 81-120. The Hague-Paris: Mouton.
- Hasan, Ruqaiya. 1973. Code, Register and Social Dialect. In B. Bernstein (ed.), *Class, Codes and Control*, vol. 2: Applied studies towards a Sociology of Language, 253-292. London: Routledge and Kegan Paul.
- Hasan, Ruqaiya. 1978. Text in the Systemic-Functional Model. In W. U. Dressler (ed.), *Current Trends in Textlinguistics*, 228-240. Berlin: W. De Gruyter.
- Hasan, Ruqaiya. 1984. The Nursery Tale as a Genre. *Nottingham Linguistic Circular* 13. 71-102.

- Hasan, Ruqaiya. 1989. The identity of a text. In Michael A. K. Halliday & Ruqaiya Hasan (eds.), *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective*, 97-109. Geelong, Victoria: Deakin University.
- Hasan, Ruqaiya. 1991. Questions as mode of learning in everyday talk. In Thao Le & Mike McCausland (eds.), *Language Education: Interaction and Development*, 70-119. Launceston: University of Tasmania.
- Jakopin, Primož & Aleksandra Bizjak. 1997. O strojno podprtem oblikoslovnem označevanju slovenskega besedila. *Slavistična revija* 45/3-4. 513-532.
- O'Donnell, Matthew B. 2005. *Corpus Linguistics and the Greek of the New Testament*. (New Testament Monographs 6). Sheffield: Sheffield Phoenix Press.
- Sinclair, John & Malcolm Coulthard. 1975. *Towards an Analysis of Discourse: the English used by teachers and pupils*. London: Oxford University Press.
- Steiner, Erich. 2003. Macro- and micro-level approaches to translated texts – methodological contradictions or mutually enriching perspectives? In Nicole Baumgarten, Claudia Böttger, Markus Motz & Julia Probst (eds.), *Übersetzen, Interkulturelle Kommunikation, Spracherwerb und Sprachvermittlung - das Leben mit mehreren Sprachen*. Festschrift für Juliane House zum 60. Geburtstag. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 8(2-3). 1-5. <http://zif.spz.tu-darmstadt.de/jg-08-2-3/beitrag/Steiner1.htm>.

APPENDIX

Text: Don't tell anyone

A leper came to him. He fell on his knee in front of him. He pleaded with his right hand over his mouth. "If you want to, you can heal me!" Jesus looked at him, that creature of his, his brother, and once again he was touched with pity. He said: "Of course I want to. Be healed!" Having learned from what had happened to him before, he added: "Please don't tell anyone except the priests you need to tell to get your health certificate." The man not only went to those priests he told the story left, right, and center. And they came again, the hundreds and the thousands, looking for the easy thing - his touch - while remaining unchanged. They wanted to be healed like that leper, only on the surface. It is a bit like those patients doctors sometimes speak of. Sick people who are looking for an injection, an ointment, a powder, or a pill but who are not willing to change the eating or drinking habits, the smoking, the life style, that is really the cause of their sickness. They were a bit like those sufferers from VD who are quite ready to undergo all kinds of treatment but are not prepared to stop their promiscuity. Like all of us who are willing to discuss all the ills of society and how they should be overcome- how the goods should be redivided, how the poor should be helped, how the drunkards should be healed, how the prostitutes should be saved, how the orphans should be fed, how the wars should be stopped- in meeting after meeting, minuting everything carefully in writing. Yet nothing ever happens, because nothing goes deeper than our skins. Jesus said: "I will heal your skin. I will heal the surface, but don't tell anyone. Keep it a secret. Because it is not all. If you tell, they will never understand what I really came for." The ex-leper did tell. He was not really changed.

And Jesus was proved right. They came. We came, and we come, looking for the easy thing – his touch- while remaining unchanged. We are willing to pray for a healing, but less willing to listen to what bishops write about the changes to be made in this world in view of justice for all and peace in this world in his – Jesus’ – name.

Aleksandra Bizjak Končar
Fran Ramovš Institute of the Slovenian Language
Scientific Research Centre of the Slovenian Academy of Sciences and Arts
Novi trg 4
1000 Ljubljana

aleksandra.bizjak@zrc-sazu.si