Creating a system of annotation for FSP

Jana Chamonikolasová, Martin Adam, Irena Headlandová Kalischová, Martin Drápela, Lenka Stehlíková (Brno)

ABSTRACT
The article presents the recent initiative of the authors of the article to prepare the ground for setting up a corpus of texts annotated from the viewpoint of Functional Sentence Perspective (FSP). The authors are followers of Jan Firbas’s approach to information structure, who have carried out a parallel analysis of a text of fiction in search of concepts within the FSP theory that need elaboration. The article outlines the discrepancies between different interpretations of selected phenomena within the text and suggests a refinement of some FSP concepts. It presents a simple FSP tagging system, which allows the annotation of FSP functions and degrees of communicative dynamism carried by communicative units.

KEYWORDS
information structure, corpora, FSP annotation, communicative units, parallel analysis

1. INTRODUCTION
The present paper aims at preparing the ground for the annotation of language corpora according to the criteria of one of the current theories of information structure — the theory of Functional Sentence Perspective (FSP) as developed by Jan Firbas and his followers. The level of FSP, seen as one of the core levels in the organization of language (Daneš, 1964), has always been given a great deal of attention in the history of Czech and Slovak linguistics, as indicated in Tyl (1970), Firbas and Golková (1975), Dušková (2005, 2015), and Drápela (2015). In spite of the extensive research in the field, “we know little at present about the general patterns of discourse organization across a large representative sample of texts from a genre”, as was pointed out by Biber et al. (2007: 10–11).

Due to the complexity of the FSP analysis, there are so far no large-scale FSP-tagged corpora available for FSP-oriented studies; however, important advances in the annotation of information structure have been made by Prague scholars developing the theory of Topic-Focus Articulation (TFA; cf. Hajičová, 2012, Sgall et al., 1986). Scholars at the Institute of Formal and Applied Linguistics of Charles University have created a TFA tagging system, which was in the past applied mainly for the annotation of Czech texts. At present, TFA annotation is also being applied to parallel texts within the Prague Czech-English Dependency Treebank (cf. https://ufal.mff.cuni.cz/~polakova/2015/docs/Anotace_aktualniho_cleneni_v_anglictine_o.pdf). Although the theories of TFA and FSP share many features, they differ in the approach to certain theoretical concepts and have to apply different annotation systems reflecting their specific interpretation of information structure phenomena.

In order to prepare the FSP annotation of language corpora, the authors of this article, all followers of the Brno theory of FSP established by Jan Firbas, have decided...
to reevaluate and adjust the existing methodology of the FSP analysis and to develop a unified FSP tagging system. The theoretical framework of this project is based on the following sources: Firbas (1990, 1992, 2010, 2013, 2014); Svoboda (1968, 1981, 1989); Chamonikolasová (2005, 2007, 2009, 2010); Adam (2009, 2010, 2013); Headlandová Kalischová (2010a, b, c); Drápela (2011a, b; 2013; 2014a, b, c); and Stehlíková (2013).

The present paper offers a discussion of some of the problems that, in their attempt at a unification of the existing FSP methodology, the team have encountered in their parallel FSP analysis of a passage from Muriel Spark’s *Memento Mori* (1961, pp. 127–129). The problems that the authors have identified cover a number of theoretical and practical issues connected with the FSP analysis.

2. FSP TAGGING SYSTEM

One of the problems that the team encountered before they started analyzing the selected passage was the relative diversity of the FSP tags used today and in the past in various monographs and articles describing FSP. For their analysis, the team adopted a modified version of a tagging system for FSP described originally in Drápela (2011a, 2011b), with the most recent appendices to it available in Drápela (2014a, 2014b, 2014c). In contrast with the various FSP nomenclatures used to this day by different FSP researchers, the proposed tagging system, *FSPML 1.0 Strict*, allows for the analysis to be stored and retrieved in a less ambiguous form, mainly with regard to the identification and tagging of communicative fields, communicative units and the degrees of communicative dynamism. In its core, the tagging system assigns each communicative unit an alphabetical and a numerical value depending on the functional assessment of the unit in the flow of communication. The team have used the following core FSP tags for the purpose of the parallel analysis: t (theme), T (diatheme), ® (transition proper), r (transition non-proper), and R (rheme). These tags are complemented with a number reflecting the degree of communicative dynamism of the unit within either the thematic or the non-thematic part of the respective communicative field — the higher the number, the higher the degree of communicative dynamism. The adopted tagging system will be further elaborated by modifications stemming from the present parallel FSP analysis and additional analyses to be conducted by the research team in the future.

Figure 1 below represents a sample of Firbas’s original analysis of the text under examination. This analysis, containing a prosodic transcription taken over by Firbas from O’Connor (1971: 2–4), was the basis of Firbas’s 1990 paper “Degrees of communicative dynamism and degrees of prosodic prominence”.

Figure 2 represents a sample of the six parallel interpretations (one by Firbas and five by the five authors of this paper), which have provided the basis for the present discussion. In the sentences in Figure 2, the label following the slash indicates the number of the sentence in the corpus. The line marked with JF represents the transformation of Firbas’s original interpretation (illustrated in Figure 1) into the new tagging system. The transformation of Firbas’s interpretation of the entire text was provided by Jiří Lukl.
1 Charmian (B, DTh) made (+; Q, Tr) her way (Sp, Rh) to the library (FSp; RhPr) 2 and cautiously (Sp; Rh) built (+; Q, Tr) up (FSpl; Rh) the fire / [ which (d, B, DTh) had (+) burnt (+; Q, Tr) low (Sp, RhPr) ] (FSp2, RhPr). 3 The effort of stooping (B, DTh) tired (+; Q, RhPr) her (d, Set, ThPr) and she (d, B, ThPr) sat (+; Q, Tr) for a moment (Set, DTh) in the big chair (Sp, RhPr). 5 After a while (Set, DTh) it was (+; AQ, Tr) tea-time (Q, RhPr). 6 She (d, B, ThPr) thought (+; Q, Tr), / for a space (Set, DTh), / about tea (Sp, RhPr). 7 Then (Set, DTh) she (d, B, ThPr) made (+; Q, Tr) her way (Sp, Rh) to the kitchen / [ where (d, Set, DTho) the tray (B, DTh) had been (+) set (+; Q, Tr) by Mrs Anthony (Sp, Rh) / in readiness [[ for Mrs Pettigrew (Ph, RhPr) to make (Pr, DTh) the tea (d, Set, ThPr) ]] (FSp, RhPr) ] (FSp, RhPr). 8 But (+o) Mrs Pettigrew (d, B, DTh) / had (+) gone

**Figure 1:** Example of Firbas's analysis (Firbas 1990)
3. CONTEXT

Although the presented tagging system does not contain specific tags indicating context dependence/independence of communicative units, the contextual status was naturally one of the criteria for the interpretation of the text. In comparing their analyses of the text, the researchers reexamined the role of context in the evaluation of the degrees of communicative dynamism of communicative units. Context is an extremely complex phenomenon and its analysis is not a straightforward task. Differences in the contextual interpretation of some elements in the text indicate areas that deserve elaboration.

Firbas introduced the concept of the immediately relevant context, which is restricted to ideas that have been explicitly mentioned in the immediately preceding written or spoken discourse and to ideas that are firmly integrated in the immediate situational context. He works with a binary opposition between retrievability and irretrievability of elements from the immediately relevant context. In Firbas’s view, a communicative unit is always either retrievable (entirely or predominantly), or irretrievable (entirely or predominantly). While fully appreciating and applying Firbas’s concept of the immediately relevant context, the analysts have also tested the possibility of extending Firbas’s binary system into a system containing the medial category of partial retrievability or accessibility applied by other scholars. It seems that the consideration of partial retrievability might facilitate the interpretation of the functions of some types of elements, for example the dynamic semantic functions of subjects. retrievable subjects are easy to identify: they can only function as Quality Bearers. Subjects denoted as irretrievable by Firbas, however, are more difficult to interpret: they can function as Phenomena, as well as Quality Bearers. It seems that Phenomena and irretrievable Quality Bearers could be distinguished on the basis of the distinction between entirely irretrievable subjects (Phenomena) and subjects that are in some way accessible, although they have not been explicitly introduced into the context of communication (Quality Bearers).
Within their considerations of the concept of retrievability, the authors have also examined the contextual status of communicative units functioning as Themes Proper and as Diathemes. Themes Proper are by definition retrievable (and deeply integrated in the context of communication). Diathemes, however, may be retrievable as well as irretrievable. Although Svoboda and Firbas have described the prominent features of Themes Proper and Diathemes, some retrievable Diathemes are not easily recognized. It seems that the distinction between Themes Proper and Diathemes has to be described more precisely, taking into consideration retrievability/irretrievability, lexical form (pronoun/noun phrase), repetition patterns, presence of contrast, and perhaps other factors. Since the diathematic function is often related to the function of Quality Bearer, a more detailed description of the Diatheme would further facilitate the separation of diathematic Quality Bearers (often irretrievable) from rhematic Phenomena (always irretrievable).

4. SEGMENTATION OF DISTRIBUTIONAL FIELDS INTO COMMUNICATIVE UNITS

The parallel analysis carried out by the individual analysts contains discrepancies in the identification of individual communicative units. Problems in segmentation have been observed especially in distributional fields containing complex verbal structures, especially multi-word and idiomatic verb phrases (VPs) of different kinds, such as made her way (S1, S7), was set in readiness (S7), or was half-filled with water (S13). While some analysts interpreted such complex structures as one communicative unit performing one FSP function, others ascribed to them two different functions (in addition to the function of the Transition Proper (®)), as illustrated below.

(S1)
<Charmian|T1> <made her way|®1r2> <to the library|R3> (...).
<Charmian|T1> <made|®1r2> <her way|r3> <to the library|R4> (...).
<Charmian|T1> <made|®1r2> <her way|R3> <to the library|R4> (...).

The strict syntactic approach dividing the phrase made her way into two communicative units was applied by Firbas and one other researcher, who interpreted the verb made as Transition and the noun phrase her way as Rheme (i.e. Rheme non-proper, followed by another element functioning as Rheme Proper). The remaining four researchers, however, interpreted it as one unit functioning as Transition (followed by the Rheme Proper). This decision was based on the static semantic interpretation of the verb phrase: made her way = went. Syntax operates at the surface level; only when seen at the interface with static semantics, a syntactic analysis may reveal essential processes in the language.

To illustrate, let us consider other analogous examples of complex VPs (V_trans+N; or V_prep+A) functioning as single communicative units taken from other corpora (Adam 2013): inch somebody’s way, coil somebody’s way, step into the spotlight. The static semantic load of these VPs in fact expresses analytically, i.e. in a complex, itemized way,
what can be conveyed by a single-word verb on other occasions, such as *come, go, appear* etc. Such interpretation may be readily corroborated by Quirk et al. (1985: 1150), who treat the multi-word verbs as units that behave “lexically or syntactically as [single syntactic constituent].”

Another type of complex VPs that triggers questions in terms of FSP interpretation are phrasal verbs such as *go out* in sentence (S8) and *built up* in sentence (S2) below. (See Quirk et al. 1985: 1150–1168 for the distinction between different subtypes of verbs: phrasal; prepositional; and phrasal-prepositional). Even though Firbas interprets the two components of the phrasal verbs as two separate units (the verbal component functioning as Transition, and the prepositional component as Rheme), the research team — in harmony with Quirk et al. (1985: 1150) — feel inclined to consider the phrasal verb cluster as a single unit performing one function — either that of Transition (if it is followed by another non-thematic unit as in S2), or that of Rheme (if it completes the message as in S8). In both conceptions, in addition to the transitional or the rhematic function, the verb performs the function of Transition Proper ([®]). Below are the two interpretations of the phrasal verb *to go out*.

(S8)

But $<Mrs\ Pettigrew|T1> <had@1> <gone@1r2> <out|R3>$,
But $<Mrs\ Pettigrew|T1> <had\ gone\ out@1R2>.$

Another example of a discrepancy between the different interpretations is in S13:

(S13)

... $<it|t1> <was@1r2> <half-filled\ with\ water|R3>.$

The complex structure *was half-filled with water* was interpreted in three different ways:

$<was@1r2> <half-filled\ with\ water|R3>$
$<was\ half-filled@1r2> <with\ water|R3>$
$<was@1r2> <half-filled|R3> <with\ water|R4>$

Two researchers interpreted the copula as one unit (Ascription of Quality; AofQ), and the rest of the phrase as another unit functioning as Quality; one researcher separated the verbal elements from the nominal part of the predicate; finally, three researchers including Firbas differentiated separate dynamic semantic functions of AofQ (copular *was*), Quality (*was filled*) and Specification (*with water*) respectively.

To sum up, it seems that the solution to the segmentation questions lies in the very syntactic character of the complex constructions. It is both the degree of idiomaticity (static semantics) and the syntactic structure that seem to lie at the root of the interpretative questions of complex constructions in FSP analysis. It follows that we deal with a tentative scale, a functional gradient of multi-word phrases (deriving from Quirk et al.’s classification) rather than a black-and-white solution to the questions under discussion. The array of segmentation issues examined above can be summarized as follows:
Type of complex phrase | Example | Tentative solution
--- | --- | ---
multiple sentence element | <strong and fearless> | 1 unit
phrasal verb | <go out> | 1 unit
prepositional verb | <look> <at me> | 2 units
copular verb + complement + prepositional phrase | <was> <half-filled> <with water> | ???

Figure 3: FSP Segmentation Issues

Finally, it would be fair to add that the segmentation controversies presumably do not represent the most painful issue in the area of FSP analysis. Important though the problematic segmentation may seem for a unified analysis carried out by the team, it definitely is not crucial. Language is a living, natural phenomenon in the end and cannot be squeezed into analytic slots with mathematical precision and natural science definiteness. This note is in harmony with the well known observations of Mathesius and Firbas concerning the potentiality of language. Incidentally, language does not work practically in communicative units; its operation is by far more complex and refined.

5. INTERPRETATION OF ADVERBIALS

Another area which proved challenging for the team was the analysis of adverbials, in particular temporal adverbials and adverbials of manner. Assigning the possible communicative value to these elements brought forth more questions than expected. Unanimous consensus of all the analysts was reached in the following two cases:

(i) all temporal adverbials in a sentence-initial position were marked as thematic units (diathemic elements to be precise), e.g.

(S7)
<Then|T2> <she|t1> <made her way|®1r2> <to the kitchen ...|R3>.

(S5)
<After a while|T2> <it|t1> <was|®1r2> <tea-time|R3>.

It is not surprising that in all sentences from this category the word order complied with the systemic ordering of items for English as suggested by the scholars of the Praguian functional generative approach, i.e. Temporal — Actor — Objective — Origin — Effect — Manner — Means — From where — Which way — Where to — Locative (Hajičová, 1998: 366).

(ii) the adverb of manner safely was unanimously assessed as a rhematic unit in sentence (S16):

(S16)
<At last|T3> <she|t1> <had lifted|®1r2> <the kettle|T2> <safely|R3> <on the gas ring|R4>.
In all other cases, however, the analyses differed as to what degree of communicative dynamism should be assigned to individual adverbials. For instance, in the opening line of S2 (see below) the adverb cautiously was viewed as a thematic unit by some and as a rhematic unit by others; the same treatment occurred later in sentence S27, compare:

(Charmian made her way to the library)  
(S2)  
and <cautiously|T1 vs. R3> <built up|@1r2> <the fire which had burnt low|R3 or R4>.

(S27)  
<Then|T3>, <cautiously|T4 vs. R3>, <she|t1> <bore|@1r2> <the little quivering flame|T2> <to the kitchen|R3 or R4>, ...

In terms of dynamic semantics, the question is whether cautiously functions as a setting to the following action or whether it specifies the action, thus functioning as a specification, which is communicatively more dynamic than the verb. The guideline provided by Firbas is not as straightforward as one might wish for, which only asserts the complexity of the matter: “For instance, in Casually, Leslie greeted the stranger, the relationship of casually to the verb is loosened and its relationship to the subject strengthened […]. This weakens casually as a specification and shifts it towards a setting […]. The insertion of a comma after Casually is in harmony with this interpretation. But as its relation to the verb has not been entirely severed, Casually continues to serve as a specification, though not in a pure and typical way.” (Firbas 1992: 53) Chamonikolasová (1987) interprets similar loosened adverbials as ‘loose diathemes’, i.e. as settings rather than specifications. The difficulty seems to lie in the employment of one’s intuition and feeling as opposed to a meticulous application of certain rules. Consequently, the judgment becomes more subjective than is perhaps desirable. It is evident that a thorough investigation into the matter remains pending.

6. CONCLUSION

The outline of areas of difficulties encountered by a group of analysts, authors of this article, in the interpretation of a text from the viewpoint of the Brno approach to information structure indicates the necessity of the refinement of certain concepts within the theoretical framework of the Brno FSP theory. Among others, the members of the research team have to set clear criteria for the identification of communicative units represented by idiomatic and complex verbo-nominal structures, which allow several segmentation patterns; and for the disambiguation of thematic and rhematic adverbials of manner, as well as thematic and rhematic subjects and objects. These steps will allow the team to apply the newly developed FSP tagging system to corpora of different text genres. An FSP annotated corpus will provide a powerful tool for the analysis of the information structure of discourse across genres.
REFERENCES


